



Maria van Kerkhove, COVID-19 Technical Lead at the WHO, gives one of many press conferences. Credit: WHO

Rise of the preprint: how rapid data sharing during COVID-19 has changed science forever

Medical discoveries have been shared at an unprecedented pace during the COVID-19 pandemic, but so have fraudulent studies, which has led to worries about scientific integrity.

Clare Watson

Within two weeks of learning about a troubling cluster of pneumonia cases in Wuhan, China, the World Health Organization (WHO) published its first guidance on the novel coronavirus that would become known as SARS-CoV-2. Days later, the first preprint on COVID-19 appeared on the open-access preprint server *bioRxiv*: a study attempting to model the transmissibility of the virus on the basis of what scant information was available.

Peter Horby, the British physician and infectious-disease epidemiologist who

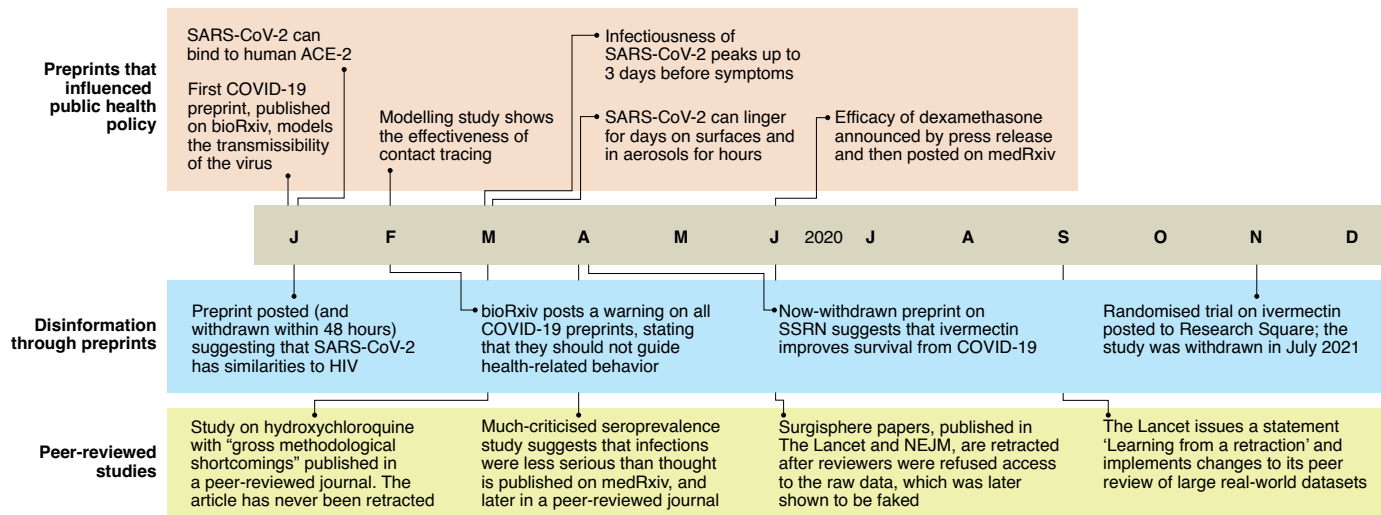
would go on to co-lead one of the most informative clinical trials of COVID-19 therapies, was hastily collating first-hand data on cases as they emerged, to estimate how deadly the new virus might be. Not long after, epidemiologists Maria Van Kerkhove and Gabriel Leung hit the ground in China, on a mission to gather data from local health officials to understand where and how the disease was spreading. Researchers had already confirmed that the virus could bind to ACE2 receptors on human cells via its spike protein. A tsunami of preprints and papers soon followed.

“[When] there’s so little information on a novel pathogen, any information that you can get your hands on is absolutely critical,” says Van Kerkhove, who sprang into action as technical lead of the WHO’s COVID-19 response the same way she had done before for major outbreaks years earlier. A global network of virologists and infectious-disease experts was activated, journal editors were contacted, and unpublished data and preliminary findings were shared.

But as with many public-health officials, what Van Kerkhove did not count on was the outsized influence of preprints,

Preprints, press releases, and policy

Preprints influenced public health policy, but others spread disinformation, with some peer-reviewed articles also criticized and retracted.



which she says have been both a blessing and a curse during the pandemic. "In the beginning it was manageable because there were very few [preprints] and it was really critical pieces of information, but it did quickly become overwhelming."

A plethora of preprints

A staggering 19,389 articles about COVID-19 were shared in the first four months of the pandemic, a third of which were preprints, unvetted and unfiltered for all to see. That number would steadily grow, as scientists raced to find drugs to treat COVID-19, develop vaccines and wrangle with viral variants. The stakes had never been higher, swift action was vital, and pre-printing results aided rapid data sharing, which expedited research. But it also exposed the inner workings of the scientific process to a new audience and laid bare the best and worst of pandemic research.

Despite the drawbacks and deadly consequences, there is little doubt that preprint publishing is here to stay. The question is how science will handle it. "We are down a pathway of open science, and that pathway is going to accelerate," says Kyle Sheldrick, a medical doctor-cum-data sleuth at the University of New South Wales, Australia. "Our choice is not whether it occurs or not; our choice is whether it occurs responsibly."

Among the first to make waves, in late January 2020, was a preprint suggesting that the new coronavirus had 'uncanny' similarities to human immunodeficiency virus. It was swiftly criticized and was withdrawn from *bioRxiv* within 48 hours, although that did little to quash the conspiracy theories it spawned. And it was

soon **usurped** by another questionable study of antibody seroprevalence that insinuated SARS-CoV-2 infections were more common and less serious than feared.

Still, other early preprints were indispensable, quickening data insights in ways **not seen in previous outbreaks**. One preprint, which was **posted to medRxiv** weeks ahead of peer review and was subsequently cited in national policy guidance from Europe to Africa, modeled early epidemiological data to show that most outbreaks of the new pathogen could be contained if upwards of 70% of close contacts were traced and isolated. But it was still unclear if and for how long people were transmitting the virus before showing symptoms.

Those data soon arrived, in **another preprint** that showed that the infectiousness of SARS-CoV-2 peaked up to three days before symptom onset, eliminating the one-week window contact tracers had with SARS. The data, which were shared in advance with the WHO and appeared in *Nature Medicine* a month later, "completely changed quarantine and contact tracing policies" for many countries, because they demonstrated that pre-symptomatic transmission was occurring, says Leung, who is based at the University of Hong Kong.

In the meantime, another preprint investigating whether SARS-CoV-2 lingers **on surfaces or in the air** had caused a stir, and many more papers were piling up. To find the signal among the noise was the task for public-health agencies. Van Kerkhove says that she and her teams would weigh up "every shred of evidence," forming a position after critiquing the fine print of each study on a topic, and never looking at one paper alone. As the pandemic picked up pace, Van

Kerkhove moved to using preprints as clues to anticipate attention-grabbing findings that would invariably provoke questions in public briefings.

Van Kerkhove maintains that overall, preprints have been a positive in this pandemic, accelerating the pace of research and directly informing public-health policies. "But for many [people] I think the jury is still out on how helpful [preprints] are because they can be quite damaging," she says. "They can misdirect a policy or they can lead you astray if you don't stay rooted in the totality of the science."

Life-saving discoveries

Nevertheless, the benefits of preprints have shone through in dark times, says cell biologist and data analyst Jonathon Coates at Queen Mary University of London. One clear standout is the first result from the UK RECOVERY trial, which would swell to become the world's largest clinical trial testing COVID-19 therapies. Dexamethasone, a cheap, common steroid that could be plucked off pharmacy shelves, reduced death by up to one third among critically ill patients on respiratory support.

Seeing the result late one Friday evening, Horby says he felt a mixture of elation and anxiety: elation at having found a life-saving treatment for COVID-19, and anxiety about whether the result was right. "When we saw the [dexamethasone] result, we had to try and break it," the University of Oxford epidemiologist recalls, "because what we didn't want to do was put out a result that was either wrong or misleading." Statisticians worked through the weekend to triple-check the analyses and look for any holes or imbalances in their data. Only after

Horby and co-investigator Martin Landray had confidence the result was solid did they share it with the world, in a [press release](#) on Tuesday 16 June 2020.

Horby says, “We announced it at lunch time, and by tea time, [dexamethasone] was being used across the UK,” endorsed by the UK National Health Service. Six days later, Horby posted the results to *medRxiv*. Within a week, the drug was being used to treat critically ill patients in intensive care units around the globe, including in Australia, much to the relief of critical-care specialist Andrew Udy at The Alfred Hospital in Melbourne, who documented the “almost immediate dramatic change in corticosteroid use.” All in all, the world knew that dexamethasone could save lives a month before the trial results were published in *The New England Journal of Medicine*, and by the year’s end, the drug had saved an estimated one million lives globally.

Despite dexamethasone’s massive global impact, Horby says that the speed of preprint publishing is a double-edged sword. It enables faster data sharing in a crisis and allows researchers to improve their work with feedback. But preprints also open the door to alluring results from slapdash science being able to find a public audience before critical review. “It speaks to the need for science to maintain a very high bar in terms of the quality,” Horby says.

Ineffective treatments

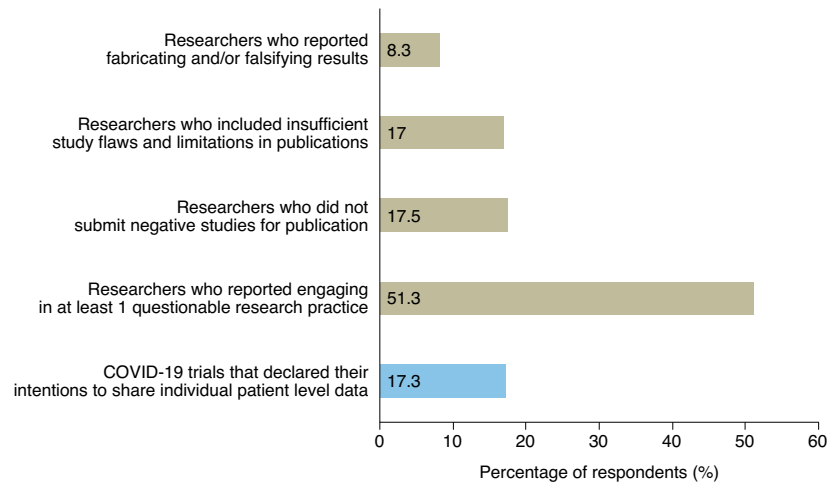
Preprints servers such as *medRxiv* have taken steps to combat the irresponsible use of preprints, introducing [additional screening measures](#) to block manuscripts that endanger people or threaten public health, and brandishing disclaimers. Research also [shows](#) that most preprints do not differ substantially, in their abstracts, figures or conclusions, from peer-reviewed versions.

But a number of questionable preprints that turned out to be fraudulent, yet continued to reverberate long after they were either withdrawn or refuted illustrates the danger of preprints that never pass peer review.

Among them is a preprint that no longer exists online save for one researcher’s blog, that of Carlos Chaccour, a malaria researcher at the Barcelona Institute of Global Health in Spain who critiqued the data. The observational study, posted to the SSRN server in early April 2020, suggested that the anti-parasitic drug ivermectin improved survival. The data — from the now-discredited Surgisphere database — included more African patients than there were cases on the African continent at the time, which [alarmed Chaccour](#). But before

Questionable research

A survey of 6813 researchers at Dutch institutes found that many admitted to engaging in questionable research practices, as did data sharing statements from 924 registered COVID-19 trials.



Sources: G. Gopalakrishna et al. *MetaArXiv*, 6 July 2021; R. Li et al. *Trials* 22, 153 (2021)

the study disappeared sometime in May, the preprint was cited in a white paper to the Peruvian government that recommended the use of ivermectin to treat COVID-19 in a country engulfed by the disease. The next week, it was national policy (although this was later retracted). Bolivia, Venezuela, India, South Africa and Slovakia followed suit.

The ramifications were huge. Ivermectin’s popularity [soared](#) before the drug could be properly tested. The hype also led to immense harm. People ingested dangerous amounts of ivermectin, and calls to poison-control centers in the United States [quintupled](#) after clinicians pushed unsubstantiated claims in the US Congress. A South African study also discovered ivermectin formulations that [contained undeclared substances](#), including benzodiazepines, while Papua New Guinea and Togo rolled out mass drug-administration campaigns, [siphoning ivermectin supplies](#) away from neglected tropical-disease programs that are tackling diseases such as river blindness.

“I could follow this story by the journalists that were calling me from different parts of the world,” says Chaccour, who from the start urged his peers to uphold scientific rigor, even in pandemic times. “There is a need, paradoxically, to slow down and commit with very firm steps, because even the smallest drop can cause a huge wave,” he says now.

Peer-review failures

The pace of pandemic publishing magnified shortcomings in peer review, too. Many researchers, including Coates, point out that journal publications are potentially much

more dangerous than unvetted preprints if readers assume peer review equates to certified quality science.

In the case of hydroxychloroquine, a French study with “[gross methodological shortcomings](#)” accepted for publication in March 2020 less than a day after submission fueled global demand for a drug the authors claimed quashed viral load. Prescriptions for the anti-malaria drug skyrocketed, mostly among clinicians who had never prescribed it before, as presidents and pundits peddled the unproven treatment. Nine months later, hydroxychloroquine was still being [prescribed above normal levels](#), despite convincing evidence that it was useless for treating COVID-19. The paper was never retracted.

But it was a pair of papers published in, and [retracted](#) by, two of the world’s most prestigious medical journals, *The Lancet* and *The New England Journal of Medicine*, that sent [shock waves](#) through the scientific community after [investigations](#) found that the large, real-world datasets were faked. The [Surgisphere scandal](#), as it became known, would leave academics questioning the state of science and peer review itself.

“We have been led to believe that peer review as it currently stands is the stamp of approval of quality of research, and it is not always the case,” says Gowri Gopalakrishna, an epidemiologist at Amsterdam University Medical Center who has turned her attention to research integrity.

Of course, both social media and the mainstream media have roles in spreading disinformation and sowing distrust. “It’s very difficult to look at the impact of preprints alone without considering how

they have been utilized in the media,” says Gopalakrishna. “Unfortunately,” adds Seth Trueger, an emergency physician at Northwestern University, Illinois, and a digital media editor at *JAMA Network Open*, “there are a lot of bad-faith actors who jump on complex science or shoddy preprints to advance their narratives, and this can truly impact public health behaviors like masking and vaccination.”

But the pitfalls of pandemic publishing have raised some tough questions for academia itself — about peer-review processes that [lack transparency](#), open-science practices meant to foster accountability, failures of scientific integrity, the reliability of meta-analyses, and the true extent of data fraud.

“The speed and intensity with which this research has come out has really put a magnifying glass on the cracks in the wall, so to speak,” says Gopalakrishna, whose own research, an anonymous survey of nearly 7,000 scientists at Dutch universities (available on *MetaArXiv* and undergoing peer review), found that mid-pandemic, [one in two respondents](#) admitted to engaging in questionable research practices, such as downplaying a study’s flaws and limitations.

Unreliable meta-analyses

According to Kyle Sheldrick, who has spent countless hours investigating fake, fraudulent and mistaken pandemic science, the real danger lies in meta-analyses, which have the potential to amplify flawed trial data. A prime example is an Egyptian trial by Elgazzar and colleagues posted to *Research Square* in November 2020. The largest single randomized trial on ivermectin at the time, it purported to show ivermectin reduced COVID-19 deaths by 90%, an effect so large it [swung highly cited meta-analyses in ivermectin’s favor](#).

But the preprint, which led Sheldrick and four other data detectives to uncover [a handful more of flawed or potentially fraudulent studies](#), contained some impossible numbers and duplicate data. The preprint was withdrawn in July 2021 after Sheldrick and colleagues raised concerns. Yet meta-analyses that included the now-withdrawn study are still used to promote ivermectin as a wonder drug. As Sheldrick’s collaborator, epidemiologist Gideon Meyerowitz-Katz, [puts it](#): “No one noticed until it was far, far too late.”

The Elgazzar preprint — which has never been published in a peer-reviewed journal — also exposes the limitations on what can be done once a preprint turns out to be bad, says Sheldrick. Preprint servers have shown great agility, withdrawing

suspect manuscripts days after being alerted to serious ethical concern, but he says that because they lack the authority to formally retract fraudulent research the way journals do, discredited preprints can continue to wield influence online.

Through it all, Sheldrick has been shocked by the brazenness of some fraudulent operators and their sense of impunity. He also wonders how science will deal with dodgy research practices when the culture of medical research equates professionalism with blindly trusting other academics. “These are not datasets that people thought would pass serious scrutiny. These are datasets people never expected to face serious scrutiny,” Sheldrick remarks.

Requesting that clinical trial investigators [release raw patient data](#) for meta-analysts to review and scrutinize — while excluding any studies that do not comply — could help change that, or at least prevent the amplification of flawed data by meta-analyses that have the power to change clinical practice and public policy. “Meta-analyses with the wrong conclusions are the single most dangerous papers that any journal can publish,” says Sheldrick.

Questions about research quality

Some scientists have [argued](#) that subpar research is the unfortunate but inevitable fallout of the pandemic, which called for nothing less than speedy science. Others say preprints have served their purpose, aiding public-health decisions and accelerating research. One concern is that attractive or dangerous ideas from preprints take hold long before more-robust research can be done, and science is notoriously slow to self-correct. “It’s just so hard to unring the bell,” says Trueger.

Discussing preprints also leads some researchers to suggestions of how to improve the scientific process. From a clinical perspective, Udy, an intensivist, says, living systematic reviews that synthesize emerging evidence can act as a robust filter to remove inaccurate data and stop illegitimate results from being translated into clinical practice. But ultimately, clinicians have a responsibility to scrutinize the data underpinning results and use proven therapies. “The onus is on them,” Udy says. “If clinicians use information that is disingenuous, or is in fact inaccurate or wrong, that can lead to patient harm.”

Gopalakrishna, an advocate of preprints, says promoting open science must go hand in hand with deeper efforts to improve the quality of research. This includes full data sharing, publishing registered reports of study protocols before

trials commence, and making modeling used in policy decisions public, for greater transparency and accountability. “These are all steps that will improve the overall quality of research, be it preprints or journal publications,” she says.

However, a recent study assessing COVID-19 clinical trials shows that a considerable proportion of investigators are [unwilling to freely share data](#), which matches Sheldrick’s experience: “Far and away the hardest step has actually been to get [the] data,” to check their validity, he says, “because if these people choose not to respond there are absolutely no consequences for them.”

Gopalakrishna also [worries](#) that researchers, universities and institutes are reluctant to even discuss poor research practices, which means academia is only reacting to the symptoms of sloppy research — piles of preprints, scores of dubious papers and some monumental retractions — rather than investigating its root cause. “We’re being rewarded on the number of publications, and not on the quality and rigor of the science,” nor for contributions to peer review, she says.

Chaccour agrees that academia’s ‘publish or perish’ mentality creates perverse incentives for researchers to publish hyped-up or half-baked studies, fast. “There needs to be a revamping of academia and how we value work,” he says. Post-publication reviews can help flag erroneous research and widen the community of reviewers beyond those whom journals select, but voluntary data sleuthing is not sustainable, nor is it fail-safe. Furthermore, counting on scientists to combat the disinformation sowed by preprints amplifies the pressure on researchers with already intense workloads, says Chaccour, who, like many outspoken experts, has received abuse for their efforts. Some even faced [death threats](#).

Horby, looking back at the [hundreds of small or duplicative trials](#) that amounted to wasteful research, says that some responsibility also lies with funders, hospitals and universities in not funding or approving poorly designed trials that are bound to deliver weak or meaningless results. “There needs to be some culpability there, for the institutions that have allowed that kind of work to be done within their walls,” he says. □

Clare Watson

Freelance science journalist, Wollongong, Australia.

Published online: 14 January 2022
<https://doi.org/10.1038/s41591-021-01654-6>