

Supplementary written evidence submitted by Dr Alina Chan (RRE0098)

I would like to submit the following recommendations for how publishers can practically incentivize research integrity and reproducibility. During the meeting on 15 December 2021, we heard that there is no incentive for publishers to prioritize research integrity and that the process of correction can take extended periods of time. We also heard that publishers are receiving so many submissions that they take statements and declarations of conflicts of interests on trust. During times of crisis, these shortcomings in the research publication system can be exploited by bad actors or unintentionally lead to the perpetuation of unsound narratives.

In the Covid-19 pandemic, we have seen how withholding or delaying the sharing of information, e.g., relating to human-to-human transmission, pre-symptomatic transmission, the SARS-CoV-2 virus genome, can have a cost measured in human lives. The following policies should be enforced specifically when manuscripts describe outbreaks or pathogen research:

1. Make preprints mandatory prior to submission.

- a. This is important so that readers can see what the original data, analysis, and assertions are prior to editorial and peer review. This enables quality control of the peer review process, e.g., was a paper significantly revised between preprinting and publication or did it get waved through peer review?
- b. Preprinting enables the public to observe which papers are being accepted or rejected at journals, and if there is gatekeeping or censorship at play.
- c. If there are rare papers that cannot be preprinted prior to submission, the journal should clearly publish the reasons for the exception.

2. Publish (anonymized or not) peer reviews, even for rejected manuscripts.

- a. As discussed during the STC meeting, these reviews are unpaid highly expert work that mostly never get published even though they would greatly benefit the rest of the research community.
- b. Open peer review enables the public to watch for gatekeeping and biases.
- c. Access to expert reviews accelerates the identification of flaws in each paper.

3. Journals should check that, by time of submission, all data described in the paper have been deposited on international databases where alterations are tracked.

- a. In the case of exceptions, the journal should publish the reasons. If a particular government will not allow its scientists to deposit data onto internationally curated databases, journals should decline to publish these papers.
- b. It should be mandatory that post-publication data additions be associated with a correction or erratum. For example, new data/sequences have repeatedly been added to the Zhou et al. Wuhan Institute of Virology *Nature* paper post-publication with barely any explanation as to how the new data or sequences were generated (Zhou et al. 2020).

- c. An example of why this is important can be found in Bloom et al. *Molecular Biology and Evolution* (Bloom 2021), which describes how early SARS-CoV-2 sequences were deleted from the international NCBI database by request of the data depositors. The recovery of these sequences one year later by Dr. Jesse Bloom led to a stronger assessment that earlier variants of SARS-CoV-2 had preceded the cluster of Covid-19 cases at the Huanan seafood market.
- d. Another example is Latinne et al. *Nature Communications* (Latinne et al. 2020) which deposited its data prior to publication and prior to the pandemic. After the paper and its dataset of virus sequences collected up to 2015 was published in 2020, independent analysts were able to identify 8 close relatives to SARS-CoV-2 in this collection being studied at the Wuhan Institute of Virology.

4. Incentivize replication studies by quickly publishing submissions that correctly call out errors in papers

- a. Corrections and retractions should be executed as quickly as a paper was peer reviewed. The peer review in these processes should be open.
- b. If journals make the review process harrowing and unrewarding for scientists who bring forward constructive criticism and clarifications of existing publications, this deters efforts to reproduce published work. When there is little effort to reproduce published work, negligence or research misconduct will rarely be detected and publicly confronted. This degrades research integrity and harms the entire research community.
- c. Examples include the late corrections and addendum applied to papers describing close relatives to SARS-CoV-2: The pangolin papers, Xiao et al. *Nature* (Xiao et al. 2021) and Liu et al. *PLoS Pathogens* (Liu et al. 2021) were only corrected in mid/late 2021 after their release in February 2020 had stimulated speculation in the research and media communities for much of 2020. *Nature* also declined to publish the letter pointing out the missing data and incorrect sample histories in the Xiao et al. *Nature* paper (Chan and Zhan 2020) (I am one of the authors of this letter), opting instead to only acknowledge the Chan and Zhan preprint in their extensive correction that encompassed one figure, one table, one new metagenomic dataset, one new supplementary table, one new PCR sequencing dataset, and six new sequences. Separately, *Nature* did not publish letters pointing out the sample history of RaTG13, the closest relative to SARS-CoV-2 at the time, in Zhou et al. *Nature*; an addendum by the authors from the Wuhan Institute of Virology was only posted nine months later (Zhou et al. 2020).

5. Refuse to publish papers where novel pathogen samples/sequences have not been shared in international databases for more than a year post-discovery.

- a. Many research groups tend to hoard pathogen sequences for a long period of time before publishing them in batches for each manuscript. This has resulted in a situation where we have no insight into the pathogen sequences most recently collected and being studied in laboratories around the world. For instance,

although Latinne et al. *Nature Communications* was published in 2020, it only described samples collected up to 2015. We have some but very little insight into the samples and sequences that have been collected by the Wuhan Institute of Virology after 2016.

- b. Journals should enforce a policy where sequences described in submitted manuscripts must have been deposited in international databases at maximum one-year post-discovery, even if the data remains embargoed. This incentivizes data sharing, transparency and accountability relating to pathogen research.

If leading publishers such as Springer Nature, Elsevier, AAAS, and the United States National Academy of Sciences agree to implement these new policies, other publishers will follow and scientists will be incentivized to publish pathogen or outbreak-related research of the highest integrity and hold each other accountable when research is found not to be reproducible.

In addition to the five recommendations above, I would like to specifically request that the STC conduct an inquiry into the origin of Covid-19. The UK is one of the leading countries in the world in biomedical research, has a particular strength in global health issues through its academic institutions, its academies and the Wellcome Trust. The Covid-19 pandemic represents an extraordinary situation (hopefully once-in-a-century) for which exceptions to the rule should be made. For instance, in response to the September 11 attacks, investigators were able to access individual information about passengers and ticket purchases, which are normally kept confidential. The pandemic is a public health catastrophe that has spanned at least 351 million infected persons and 5.6 million deaths. Yet, there is no access to key relevant information held at journals due to confidentiality policies surrounding the peer review process. It makes sense for the UK parliament to advise the UK government to initiate an international investigation into the origin of Covid-19.

Making available all relevant submitted manuscripts at journals in 2019 and 2020 to investigators is vital. The STC can compel research journals, especially those based in the United Kingdom, to share the original manuscripts (regardless of published, rejected, or withdrawn by authors), editorial correspondences, and peer reviews relating to SARS-CoV-2 or related viruses. Independent investigators will be able to assess the full breadth of information and data describing the beginnings of the pandemic, determine if mistakes or deliberate errors were made, if gatekeeping and biases were at play, and what steps to take to prevent similar errors in future crises.

January 2022

Citations

- Bloom, Jesse D. 2021. "Recovery of Deleted Deep Sequencing Data Sheds More Light on the Early Wuhan SARS-CoV-2 Epidemic." *Molecular Biology and Evolution* 38 (12): 5211–24.
- Chan, Yujia Alina, and Shing Hei Zhan. 2020. "Single Source of Pangolin CoVs with a near Identical Spike RBD to SARS-CoV-2." *bioRxiv*. <https://doi.org/10.1101/2020.07.07.184374>.
- Latinne, Alice, Ben Hu, Kevin J. Olival, Guangjian Zhu, Libiao Zhang, Hongying Li, Aleksei A. Chmura, et al. 2020. "Origin and Cross-Species Transmission of Bat Coronaviruses in China." *Nature Communications* 11 (1): 4235.
- Liu, Ping, Jing-Zhe Jiang, Xiu-Feng Wan, Yan Hua, Linmiao Li, Jiabin Zhou, Xiaohu Wang, et al. 2021. "Correction: Are Pangolins the Intermediate Host of the 2019 Novel Coronavirus (SARS-CoV-2)?" *PLoS Pathogens* 17 (6): e1009664.
- Xiao, Kangpeng, Junqiong Zhai, Yaoyu Feng, Niu Zhou, Xu Zhang, Jie-Jian Zou, Na Li, et al. 2021. "Author Correction: Isolation of SARS-CoV-2-Related Coronavirus from Malayan Pangolins." *Nature* 600 (7887): E8–10.
- Zhou, Peng, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, et al. 2020. "Addendum: A Pneumonia Outbreak Associated with a New Coronavirus of Probable Bat Origin." *Nature* 588 (7836): E6.