Trust, but thoroughly verify

Two recent high-profile retractions of COVID-19 papers exemplify that trust cannot be taken for granted. To strengthen it, scientific review will have to become more transparent.

A global health emergency; expected yet influential findings; impressive real-world datasets. These were strongly motivating conditions for the publication of two notorious observational studies — later retracted[1,2] — about the absence of benefit and the potential for harm, in patients with COVID-19, of two malaria drugs (chloroquine and hydroxychloroquine) and of routinely taking common medication for lowering blood pressure (angiotensin-converting enzyme inhibitors). The studies analysed data from electronic health records supposedly obtained from hundreds of hospitals around the world.

Experts scrutinizing the published analyses soon found puzzling inconsistencies[3] in the data (such as implausible demographics and dosing, and inadequate adjustment for confounders) and hard-to-believe claims[4] about the provenance of the source datasets (a score of hospitals whose assistance would have been indispensable have denied providing any data). An investigation[5] of publicly available information by the British newspaper *The Guardian* revealed that Surgisphere, the analytics company that provided the data for both studies and that claimed to run a large multinational database of medical records, appeared to have few employees — most of them without an apparent scientific background — and a meagre online presence (since mid-June it appears that the company has ceased operations). The firm's founder, chief executive and co-author of both studies boasted that "with data like these, do we even need a randomized control trial?". In fact, in view of the claimed potential harm and lack of therapeutic benefit of hydroxychloroquine, recruitment for clinical trials of the drug were paused shortly after publication of the study[1] (and later resumed[6]). In Latin America, a different study also using Surgisphere data (published only as a preprint and later withdrawn) claiming that the antiparasitic drug ivermectin reduced the mortality of patients with COVID-19 prompted the rapid authorization and use of the drug[7] in these patients.

What went wrong before the publication of the studies? Nothing, and nearly everything.

IN THE
SCIENTIFIC
METHOD
WE TRUST
EVERYONE
VERIFY
THE DATA

In the context of how science publishing customarily functions, the process worked as it should. Timely results with the potential to change medical practice were submitted to influential journals. Journal editors logically considered the studies for expert vetting. Peer review did not catch every fault and inconsistency with the data and claims in the paper. Post-publication examination found holes in the analyses. The journals quickly issued expressions of concern for the studies, asked for an independent investigation into the veracity and completeness of the source data, and, when Surgisphere declined to co-operate, retracted the studies at the authors' request.

But study co-authors, editors and reviewers trusted the veracity of the data. They seemingly did not investigate how an unknown small company could have access to large international datasets of medical records that are difficult to de-identify. The editorial and peer-review processes could have been more stringent, even if at the expense of speed. Before publication of the studies, the journals could have solicited independent replication of the analyses and proof of the origination of the datasets.

Yet foresighted processes that catch falsities of all sorts (the unintended and, to some extent, unavoidable; and the deliberate

kind, which are rarer yet news-making) before it is too late cannot be easily planned and managed. Trust is essential to both the doing and the vetting of science. When it fails, reputations take a hit, and public trust in the scientific process is undermined. Pre-publication peer review is one imperfect yet essential tool for the verification and refinement of data, methods and claims. Careful study design, execution and evaluation among co-authors; protocol approvals and compliance checks by the researchers' institutions; vigilant checks by journals and stringent editorial oversight of peer review; support for preprints and for data and code deposition when feasible; and wider post-publication peer review[8]; all are underused tools that should be refined. Replication and reproducibility, although typically costly, slow and burdensome endeavours (and hence necessarily targeted) should be incentivized.

As with the balance of benefits and harms associated with most drug approvals and rejections, an optimal vetting process for scientific results involves the juggling of multiple imperfect choices. Should the findings first be published as a preprint, as most COVID-19 papers are? Should the study funders, the authors' institutions or journals mandate the full or partial disclosure of the raw and analysed data? When should journals run more stringent and transparent peer-review processes? How can post-publication scrutiny and reproducibility studies be encouraged, systematized, resourced and recognized? Blanket answers (and mandates) for these and many other such considerations are unfeasible. The expertise of the stakeholders of these processes and services, and their experience in setting them up and in running them matters. Discipline-specific needs, constraints and traditions, funder goals, institutional aims and journal strategies also do.

Reassuringly, there are encouraging trends and plenty of room for improvement. The number of preprints has been steadily growing, and the circumstances around COVID-19 are accelerating their use and acceptance. Initiatives for the curation or independent scientific review of preprints (such as the launch, announced by MIT Press, of an open-access journal that will publish reviews of COVID-19 preprints) are

increasingly popping up. With transparent peer review[9], reviewer recognition[10], granular peer-review timelines (public or private), incentives for thorough methodological reporting and for making data and code available, and the peer review of code[11], new scientific knowledge will be easier to reproduce and build on. The Nature research journals, including *Nature Biomedical Engineering*, have pioneered or are adopting most of these progressive practices.

Failure-proof vetting is a fallacy. As with science, it is under constant refinement, and the best we can aspire to is that what comes out after thorough verification can be largely trusted. ❐

References

1. Mehra, M. R., Ruschitzka, F. & Patel, A. N. *Lancet* **395**, 1820 (2020).
2. Mehra, M. R., Desai, S. S., Kuy, S., Henry, T. D. & Patel, A. N. *N. Eng. J. Med.* **382**, 2582 (2020).
3. Watson, J. et al. An open letter to Mehra et al. and *The Lancet*. *Zenodo* https://go.nature.com/3eRhiuo (2020).
4. A health data firm making extraordinary claims about its data. *Free Range Statistics* https://go.nature.com/2YPEhAz (2020).
5. Davey, M., Kirchgaessner, S. & Boseley, S. Surgisphere: governments and WHO changed Covid-19 policy based on suspect data from tiny US company. *The Guardian* https://go.nature.com/3dTu9uE (3 June 2020).
6. Joseph, A. WHO resumes hydroxychloroquine study for Covid-19, after reviewing safety concerns. *STAT* https://go.nature.com/2D4jlxl (2020).
7. Offord, C. Surgisphere sows confusion about another unproven COVID-19 drug. *The Scientist* https://go.nature.com/38kkcp7 (2020).
8. Time for the data to speak. *Nat. Biotech.* **35**, 689 (2017).
9. Opening up peer review. *Nature* **560**, 527 (2018).
10. Three-year trial shows support for recognizing peer reviewers. *Nature* **568**, 275 (2019).
11. Pastrana, E & Swaminathan, S. What we have learnt testing container-platforms for peer review and publication of code. *Of Schemes and Memes* https://go.nature.com/2YRBabn (2020).