

Uses and abuses of real-world data in generating evidence during a pandemic

Kamlesh Khunti¹ , Francesco Zaccardi¹, Nazrul Islam²  and Tom Yates¹

¹Diabetes Research Centre, University of Leicester, Leicester LE5 4PW, UK

²Medical Statistics, University of Oxford, Oxford OX3 7LF, UK

Corresponding author: Kamlesh Khunti. Email: kk22@le.ac.uk

On 11 March 2020, the World Health Organization declared COVID-19 a pandemic and called for immediate collaborative initiatives for faster access to available data, with a view to generating robust research evidence informing global and local public health policy.¹ This urgency has helped a number of national bodies to secure data and their linkages and to provide safe analytical environment for researchers to ask important questions, including pseudonymised data linkages, high-throughput computing environment, and access and authentication processes with clear information governance.^{2,3} Linkages of multiple sources of clinical data within a trusted environment are being granted at a rapid pace and there is a greater provision of access to COVID-19 studies, improved collaboration, expedited governance and ethical approval of studies.⁴ Some organisations have also been proactive in getting groups together to work collaboratively on relevant research questions which will rapidly benefit clinical care and public health alike.

In view of the pandemic unfolding in the big data era and the urgency of actionable information, the potential consequences and responses to this pandemic have been mainly using ‘real-world data’. Although disputed, real-world data may be defined as data collected primarily for purpose other than research (‘secondary data’) and the evidence generated from real-world data as ‘real-world evidence’. The fast growth of real-world data translating into real-world evidence has been made possible due to the rapid availability of semi-structured and unstructured data from a network of resources. For example, the Johns Hopkins Coronavirus Resource Center and the Worldometer are updating global real-world data on a daily basis in real time providing live statistics on the number of cases and fatalities from COVID-19.^{5,6} Along with summary statistics, real-world data have provided valuable information regarding patient

characteristics, treatment pattern, and clinical outcomes and risk factors for hospitalised patients and mortality (for example, elderly, males, obesity or individuals of minority ethnic groups).^{2,7} Real-world data have also been used to develop risk prediction models for the diagnosis and prognosis of COVID-19 and pragmatic trials of therapies using routine data to capture outcomes.^{7,8}

The rapid deployment of and the exponential rise in real-world data and real-world evidence precipitated by the COVID-19 pandemic has been, however, criticised for conducting and reporting studies of poor quality, which have been amplified during COVID-19.^{4,5} The necessity of a rapid while proper peer-review process for the submitted articles to be quickly published in journals and gain citations has been problematic, and likely contributed to the 400% increase in preprint publications compared to mainstream journals in relation to COVID-19.^{4,9} This urge to expedite the peer-review process may also have contributed to the publication of poor-quality research as well as fraud, including the recent alarming high-profile retractions of COVID-19-related research papers.¹⁰ Most of COVID-19 publications have not been registered (i.e. National Library of Medicine) as, in most cases, there are no requirements to register observational studies while only a few randomised controlled trials, for which registration is mandatory, have been registered.^{4,11} While it is recognised that the rapid access to data and new findings are crucial in the context of an emerging epidemic, it can also result in poorly designed studies gaining media attention, influencing policy, wasting resources, and overall loss of trust by the public and policy makers. There have also been duplicated publications often finding different answers using the same databases, with implications for quantitative evidence synthesis (i.e. meta-analysis) where duplicate publications have included the same cohort of patients.

Research during a pandemic obviously needs to be conducted and published at pace to inform global policy and help save lives. There is, however, potentially a massive waste that has been exacerbated by the pandemic. If we are to continue having trust in real-world data and the quality of real-world evidence publications, the international scientific community needs to give immediate attention to retain credibility of real-world data and real-world evidence. It is fundamentally important that in the rush to publish, scientific quality and high standards for research are retained. For example, greater effort is required to ensure that all published real-world evidence conform to the basic established requirements for observational research. The process of pre-print and journal submission and peer review may also need to be considered to provide greater clarity and transparency. Indeed, some funders have already invested in newer models of publishing that could be used to better effect during a pandemic. For example, the Wellcome Open Research uses the F1000Research platform to provide a transparent hybrid open access pathway between preprint archive, open peer review and author responses and subsequent manuscript revision history.

This is a call for the academic community to learn from the COVID-19 pandemic and develop guidelines and processes for real-world data and real-world evidence that work to provide an appropriate balance between rapid access to new data and findings during public health emergencies, while also retaining quality and reducing the risk of fraud.

Declarations

Competing Interests: None declared.

Funding: KK, FZ and TY are supported by the National Institute for Health Research (NIHR) Applied Research Collaboration East Midlands (ARC EM) and the NIHR Leicester Biomedical Research Centre (BRC).

Ethics approval: Not applicable.

Guarantor: KK.

Acknowledgements: None.

Provenance: Not commissioned; editorial review.

ORCID iDs: Kamlesh Khunti  <https://orcid.org/0000-0003-2343-7099>

Nazrul Islam  <https://orcid.org/0000-0003-3982-4325>

References

1. Naci H, Kesselheim AS, Rottingen JA, Salanti G, Vandvik PO and Cipriani A. Producing and using timely comparative evidence on drugs: lessons from clinical trials for covid-19. *BMJ* 2020; 371: m3869.
2. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* 2020; 584: 430–436.
3. Stephen B. Thacker CDC Library. *COVID-19 Secondary Data and Statistics*. See <https://www.cdc.gov/library/researchguides/2019novelcoronavirus/datastatistics.html> (last checked July 27 2020).
4. Glasziou PP, Sanders S and Hoffmann T. Waste in covid-19 research. *BMJ* 2020; 369: m1847.
5. Johns Hopkins Coronavirus Resource Center. See <https://coronavirus.jhu.edu/map.html> (last checked July 27 2020).
6. Worldometer. *Worldometer COVID-19 Data*. See <https://www.worldometers.info/coronavirus/about> (last checked July 27 2020).
7. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020; 369: m1328.
8. ClinicalTrials.gov. *Dapagliflozin in Respiratory Failure in Patients with COVID-19 (DARE-19)*, April 17 2020. See <https://clinicaltrials.gov/ct2/show/NCT04350593> (last checked July 13 2020).
9. Kousha K and Thelwall M. COVID-19 publications: database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quantit Sci Stud* 2020; 1: 1068–1091.
10. Piller C and Travis J. Authors, elite journals under fire after major retractions. *Science* 2020; 368: 1167–1168.
11. Singh AK, Gillies CL, Singh R, Singh A, Chudasama Y, Coles B, et al. Prevalence of co-morbidities and their association with mortality in patients with COVID-19: a systematic review and meta-analysis. *Diabetes Obes Metabol* 2020; 22: 1915–1924.