Knocking on heaven's door: The gap between health institutions and academies in generating knowledge utilizing real-world data

Giovanni Corrao[1,2], Matteo Franchi[1,2,3]* and Giuseppe Mancia[4]

[1]National Centre for Healthcare Research and Pharmacoepidemiology, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy, [2]Unit of Biostatistics, Epidemiology and Public Health, Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy, [3]Centro Cardiologico Monzino IRCCS, Milan, Italy, [4]University of Milano-Bicocca, Milan, Italy

Albeit with an inevitable delay from the eruption of the first wave of the Coronavirus disease 2019 (COVID-19) pandemic, medical research has offered populations and patients monumental achievements that should not be forgotten. The virus has been isolated and its variants can be promptly isolated as they appear. Rapid and increasingly more-reliable diagnostic tests have been made available at progressively lower prices. The diverse local and systemic manifestations of COVID-19 viral infections have been described in detail together with the multiple risk factors responsible for their evolution to severe and lethal forms. Direct (antiviral drugs and monoclonal antibodies) and indirect treatment strategies have been discovered or refined with a measurable benefit on infection-related outcomes. The long-term consequences of COVID-19 remain mechanistically unclear, but their multiple clinical phenotypes have been characterized and are now better understood.

Finally, and most importantly, efficacious and safe vaccines against COVID-19 have been developed at a speed never achieved in the past, thanks to the huge efforts of the pharmaceutical industry backed by government support (1). Well-designed randomized clinical trials have provided robust evidence of the immunogenicity, efficacy, and safety of most available vaccines (2–8), and vaccination campaigns have successfully involved large population strata in numerous countries, despite multiple critical issues and difficulties. Somewhat unexpectedly, vaccination has turned out to have limited ability to prevent variant-related COVID-19 infections, but the documented efficacy of vaccination against severe and lethal diseases has now prevented hospitals from becoming overwhelmed (9–12). In summary, it is undeniable that the world is much better equipped to fight future pandemics.

Despite the above-mentioned contributions, research has not provided unequivocal answers to several important basic as well as diagnostic and therapeutic questions. For example, concerning vaccination, unresolved questions abound. First, how should individuals most prone to severe infection be identified and prioritized for vaccination? (13). Second, what are the reasons for inequality in the distribution of vaccination within populations? (14). Third, to what extent do vaccination programs prevent COVID-19 infections, hospitalizations, or death, and how does this change in relation to emerging variants? (15, 16). Fourth, what harms are associated with COVID-19 vaccines and how can rare, serious adverse effects be explained, predicted, and avoided? (17). Lastly, what is the actual duration of vaccine-dependent protection and how can we measure the various components and overall efficacy of vaccination? (18). These issues have a direct impact on the population's health worldwide.

These outstanding issues justify the efforts to continuously monitor and assess vaccination campaigns beyond the classical pharmacovigilance approach that has been implemented in several countries, including Israel, the United States, the United Kingdom, and Italy (15, 16, 19–21). In some countries, efforts have also been made to do research into COVID-19 by using a coordinated or integrated approach. In these integrated research pathways, basic science helps in the identification of preventive and curative agents, experimental trials are carried out to prove efficacy and safety, and real-world observations complement trial evidence. This approach provides integrated information with which to guide the use of new COVID-19 treatments in medical practices and at the level of public health. This implies the recognition of real-life-based research as a fundamental component of the knowledge obtained by the research community in the COVID-19 pandemic, with a critical role to play in the future.

The above-mentioned integrated research pathway, and the inclusion of real-life data as a necessary (although insufficient) research step, is not a new approach. What has recently happened as a result of the COVID-19 pandemic has only accelerated a process that had begun at least a decade ago to tackle other medical issues. For example, in precision medicine (22), healthcare strategies are personalized and, therefore, patient subgroups who will benefit most from new drugs need to be identified. Precision medicine requires that the population is characterized by a wide range of individual biomolecular, clinical, demographic, and socio-economic features that can only be made available by real-life information. Real-life data is also required for the adoption of treatment guidelines. Even when these guidelines are evidence-based, their recommendations need to be verified in the context of real-life application to measure levels of acceptability and adherence by physicians and patients, as well as the appropriateness of their use (23).

Even today, many medical actions are dictated by tradition or mechanistic inferences, rather than by trial findings or other types of evidence; therefore, real-life data from clinical practice may be necessary to verify the actual clinical impact and cost implications of these actions (24). Service implementation, availability, accessibility, and integration can be optimized in so-called "value-based healthcare", which maximizes clinical benefits at lower costs (25), and this can only be provided by real-world evidence. As conceptualized by Tanahashi in a pivotal publication (26), we need to measure potential coverage (service availability, accessibility, and acceptability), contact coverage (use of evidence-based healthcare recommendations), and effective coverage (use of healthcare data which translates into health benefits) in people who are stratified according to health needs.

However, two main problems may slow down the use and development of observational studies based on routinely collected data. The first problem is that accessibility to (and interconnection of) the now widely available, real-world big data in a large number of countries is questioned by individuals who place the secrecy of health-related information above any other consideration. Of course, an individual's health status should remain strictly confidential, but tools and techniques capable of protecting this information entered research several years ago. Privacy by design (27) is now an integral part of any study protocol that involves data of this type, and universal application of this rule should become mandatory for any future real-world-based research. This being the case, the ethical principle according to which the privacy of each individual citizen should be protected does not justify a rigid (and regretfully widespread) attitude that prevents access to valuable healthcare data, which can be of fundamental importance for medical research and knowledge, even more during pandemics. In countries such as Sweden, this problem has been avoided by asking citizens to authorize the use of their anonymized healthcare data through the no-reply assent approach (28), which has enabled this country to collect widespread clinical registry data of inestimable research value. Hopefully, other countries will soon follow this example.

The second problem is that even when real-life big data is available, the generation of credible evidence is not guaranteed, even if the data source is of good quality. This problem occurs because if some basic rules are not respected, even good-quality data can give a distorted image of reality. One basic rule, for example, is to avoid "a fishing expedition" (29). A researcher throwing the hook into a very fishy sea of data will almost always catch something. However, the goal of meaningful research is not to look for whatever can be caught in the large sea of data (for example, statistical correlations) but to identify causal links that aid the progression of knowledge and improve health system quality. Another basic rule is to avoid what statisticians call data torture [30], which is the re-analysis of the same data again and again until results plausible for the original hypothesis of

the investigator are obtained. Obviously, this has little to do with scientific research.

In conclusion, the COVID-19 pandemic has made it clear that it is time to stop considering observational research based on widely available big data, as at best "hypothesis generating" and realize that these data represent a final fundamental research step without which knowledge obtained by classical research approaches lose much of their potential value. Anyone who is able to design a study that uses current data and their interconnections to improve knowledge and support decision-making, who has documented skills that guarantee rigorous data analysis, who respects privacy, and who makes the results of the study available to health systems should be enabled to access the data. In other words, observational research by means of big data should be available to anyone who guarantees an ethical approach to respect the rules of privacy protection and follows good research practices.

Author contributions

GC, MF, and GM were responsible for the design of this work. GC drafted the work and is accountable for all aspects of the work to ensure its accuracy and integrity. All authors gave the final approval of the version to be published.

Conflict of interest

Publisher's note

References

1. Conte C, Sogni F, Affanni P, Veronesi L, Argentiero A, Esposito S. Vaccines against Coronaviruses: the state of the art. *Vaccines (Basel).* (2020) 8:309. doi: 10.3390/vaccines8020309

2. Forman R, Shah S, Jeurissen P, Jit M, Mossialos E. COVID-19 vaccine challenges: what have we learned so far and what remains to be done? *Health Policy.* (2021) 125:553–67. doi: 10.1016/j.healthpol.2021.03.013

3. Polack FP, Thomas SJ, Kitchin N, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New Engl J Med.* (2021) 383:2603–15. doi: 10.1056/NEJMoa2034577

4. Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. COVE Study Group. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N Engl J Med.* (2021) 384:403–16. doi: 10.1056/NEJMoa2035389

5. Ramasamy MN, Minassian AM, Ewer KJ, et al. Oxford COVID Vaccine Trial Group. Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial. *Lancet.* (2021) 396:1979–93.

6. Sadoff J, Gray G, Vandebosch A, et al. ENSEMBLE Study Group. Safety and efficacy of single-dose Ad26COV2S vaccine against Covid-19. *N Engl J Med.* (2021) 384:2187–201. doi: 10.1056/NEJMoa2101544

7. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. on behalf of the Oxford COVID Vaccine Trial Group. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil S. and the UK. *Lancet.* (2021) 397:99–111.

8. Voysey M, Costa Clemens SA, Madhi SA, et al. Oxford COVID Vaccine Trial Group. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. *Lancet.* (2021) 397:881–91.

9. World Health Organization. *WHO Coronavirus Disease (COVID-19) Dashboard.* Available online at: https://covid19.who.int/ (accessed July 12, 2022).

10. Johns Hopkins University of Medicine. Coronavirus Research Center. *Mortality Analyses.* Available online at: https://coronavirus.jhu.edu/data/mortality (accessed July 12, 2022).

11. European Centre for Disease Prevention and Control. *COVID-19 Situation Dashboard.* Available online at: https://qap.ecdc.europa.eu/public/extensions/COVID-19/COVID-19.html#global-overview-tab (accessed July 12, 2022).

12. Our World in Data. *Statistics and Research. Coronavirus Pandemic (COVID-19).* Available online at: https://ourworldindata.org/coronavirus (accessed July 12, 2022).

13. Corrao G, Franchi M, Cereda D, Bortolan F, Leoni O, Jara J, et al. Factors associated to severe or fatal clinical manifestations of SARS-CoV-2 infection after receiving the third dose of vaccine against COVID-19. *J Intern Med.* (2022) in press. doi: 10.1111/joim.13551

14. Etienne CF. COVID-19 has revealed a pandemic of inequality. *Nat Med.* (2022) 28:17. doi: 10.1038/s41591-021-01596-z

15. Jones M, Khader K, Branch-Elliman W. Estimated impact of the US COVID-19 vaccination campaign-getting to 94% of deaths prevented. *JAMA Netw Open.* (2022) 5:e2220391. doi: 10.1001/jamanetworkopen.2022.20391

16. Steele MK, Couture A, Reed C, Iuliano D, Whitaker M, Fast H, et al. Estimated number of COVID-19 infections, hospitalizations, and deaths prevented among vaccinated persons in the US, December 2020 to September 2021. *JAMA Netw Open.* (2022) 5:e2220385. doi: 10.1001/jamanetworkopen.2022.20385

17. Corrao G, Rea F, Franchi M, Cereda D, Barone A, Borriello CR, et al. Balancing benefits and harms of COVID-19 vaccines: lessons from the ongoing mass vaccination campaign in Lombardy, Italy. *Vaccines (Basel).* (2022) 10:623. doi: 10.3390/vaccines10040623

18. Corrao G, Franchi M, Cereda D, Bortolan F, Zoli A, Leoni O, et al. Persistence of protection against SARS-CoV-2 clinical outcomes up to 9 months since vaccine completion: a retrospective observational analysis in Lombardy, Italy. *Lancet Infect Dis.* (2022) 22:649–56. doi: 10.1016/S1473-3099(21)00813-6

19. Dagan N, Barda N, Kepten E, Miron O, Perchik S, Katz MA, et al. BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N Engl J Med.* (2021) 384:1412–23. doi: 10.1056/NEJMoa2101765

20. Antonelli M, Penfold RS, Merino J, Sudre CH, Molteni E, Berry S, et al. Risk factors and disease profile of post-vaccination SARS-CoV-2 infection in UK users of the COVID Symptom Study app: a prospective, community-based, nested, case-control study. *Lancet Infect Dis.* (2022) 22:43–55. doi: 10.1016/S1473-3099(21)00460-6

21. Corrao G, Bertolaso G, Pavesi G, Moratti L. Eight good reasons for careful monitoring and evaluation of the vaccine campaign against COVID-19: lessons learned through the Lombardy experience for dealing with next challenges. *Int J Environ Res Public Health.* (2022) 19:1073. doi: 10.3390/ijerph19031073

22. National Research Council (US) Committee on a Framework for Developing a New Taxonomy of Disease. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease.* Washington (DC): National Academies Press (US) (2011).

23. Cohn JN. Monitoring adherence to guidelines. *Circ Heart Fail.* (2008) 1:87–8. doi: 10.1161/CIRCHEARTFAILURE.108.795294

24. Corrao G, Mancia G. Research strategies in treatment of hypertension: value of retrospective real-life data. *Eur Heart J.* (2022) 43:3312–22. doi: 10.1093/eurheartj/ehab899

25. Cossio-Gil Y, Omara M, Watson C, Casey J, Chakhunashvili A, Gutiérrez-San Miguel M, et al. The roadmap for implementing value-based healthcare in European University Hospitals-consensus report and recommendations. *Value Health.* (2022) 25:1148–56. doi: 10.1016/j.jval.2021.11.1355

26. Tanahashi T. Health service coverage and its evaluation. *Bull World Health Organ.* (1978) 56:295–303.

27. Kum HC, Ahalt S. Privacy-by-design: understanding data access models for secondary data. *AMIA Jt Summits Transl Sci Proc.* (2013) 2013:126–30.

28. Roberts P. *Avoid the "Fishing Expedition" Approach to Analytics Projects.* Available online at: https://www.datasciencecentral.com/avoid-the-fishing-expedition-approach-to-analytics-projects/ (accessed July 12, 2022).

29. Mills JL. Data torturing. *N Engl J Med.* (1993) 329:1196–9. doi: 10.1056/NEJM199310143291613