Contents lists available at ScienceDirect

Clinical Microbiology and Infection

journal homepage: www.clinicalmicrobiologyandinfection.com

Commentary

The role of observational studies based on secondary data in studying SARS-CoV-2 vaccines

Noam Barda [1, 2, *], Noa Dagan [1, 2, 3, 4]

1) Software and Information Systems Engineering, Ben Gurion University, Be'er Sheva, Israel
2) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA
3) Clalit Research Institute, Innovation Division, Clalit Health Services, Tel Aviv, Israel
4) The Ivan and Francesca Berkowitz Family Living Laboratory Collaboration at Harvard Medical School and Clalit Research Institute, Boston, MA, USA

ARTICLE INFO

The Covid-19 pandemic is an ongoing public health crisis of enormous proportions. Of the many public health interventions taken to mitigate and contain the pandemic's effects, SARS-COV-2 vaccines constitute a critical measure. As new vaccines are rapidly developed and the pandemic continues to evolve with new variants appearing and receding, many important scientific questions naturally rise. These questions demand valid and timely answers to inform policy, and randomized controlled trials (RCTs) can provide only some of them. Observational studies based on secondary data—registry and clinical data originally collected for other purposes—are being used to fill these gaps.

In this issue of the journal, Vokó et al. [1] report a study which makes use of Hungarian nationwide centralized vaccine and outcome registries to estimate and compare the effectiveness of five different SARS-CoV-2 vaccines against SARS-CoV-2 infection and Covid-19-related death, using regression to adjust for differences between the study populations. The study was performed during a period when the alpha (B.1.1.7) variant was dominant in Hungary.

This interesting study has several strengths. First, the reality in Hungary, in which several vaccines were used concurrently, allows

the authors to study these different interventions in a single setting. This is particularly interesting as Hungary deployed, and this study includes, SARS-CoV-2 vaccines that have yet to be approved by the European Medicines Agency and have not been as extensively studied in real-world settings. Second, the use of nationwide linked registries, which include exposure and outcome information, leads to a large sample size with little to no selection of individuals; this allows for precise estimation (i.e., with narrow confidence intervals) that should also generalize well to other locales. Last, the authors perform multiple sensitivity analyses to explore different modelling options and time period definitions, finding their estimates robust to these choices.

This study also has certain limitations, which the authors candidly acknowledge. First, without access to data on a patient's baseline health status and health behaviours, the adjustment performed is minimal, likely resulting in residual confounding. This is particularly concerning because, as the authors state, "some vaccines were specifically indicated for use in elderly and chronically ill patients". Second, the authors opt to model all the follow-up time available for each patient at once, implicitly assuming a constant effect throughout the study period. With the growing evidence of waning immunity, we know this not to be true. Last, as has now been discussed extensively in other studies [2], it is likely that not all infections are identified, and that this misclassification occurs differentially between treatment groups. While these limitations are important, the effects observed, which are congruent with previous studies, are informative and provide a valuable addition to existing evidence.

RCTs are the reference standard for medical scientific evidence. Owing to the benefits of randomization and adherence to strict protocols, the internal validity of the evidence generated by such trials is high. This validity underscores their crucial role in directing public health policy and regulatory approval of therapeutics. However, due to logistical and ethical considerations, RCTs cannot answer all scientific questions of interest, necessitating observational studies, today mostly based on secondary data sources. This was never more evident as in research on Covid-19 vaccines, where invariably RCTs answered initial questions regarding vaccine

efficacy and safety, and observational studies proceeded to address a wide range of resulting issues, including real-world effectiveness, safety in regard to rare adverse events, waning immunity, effectiveness against different variants, effectiveness in pregnant women and more.

The two types of study are complementary. For example, safety signals originally generated from RCTs [3] were further explored using observational studies with larger sample sizes [4]. In a more methodologically interesting example, RCTs established the early period following vaccination as a negative control outcome (in which no effect of the vaccine is expected) [3], which was then used by observational studies to detect bias [5].

The main advantages of observational studies based on secondary data are the large sample size, which allows exploration of rare outcomes relating to vaccine effectiveness (e.g., severe disease and death) and vaccine safety, and exploration of outcomes within subgroups; the fact that they include less selected populations, such as individuals with unstable chronic conditions and pregnant women; their reflection of real-life conditions in which adherence to predetermined protocols may be less strict; the integration with different sources of data, which allows studying varying outcomes and adjusting for many confounders; and the immediate availability of the data with little additional costs, which allows rapid answers to emerging questions (e.g., waning immunity [6]).

Observational studies that are based on secondary data sources also have important disadvantages for vaccine studies, as they do for other questions. The first potential disadvantage concerns the quality of the data, which are not collected for research purposes, and for which quality assurance measures vary between locales and times. To address this, the researcher must be intimately familiar with the data collection and curation mechanisms and to know which data are trustworthy. A second major disadvantage is that secondary data sources may amplify the usual threats to validity of observational studies. Specific variables that were not documented (e.g., behavioural factors) allow the possibility of residual confounding; measurement error is a common challenge as, e.g., individuals select whether to be tested [7]; selection bias is a possibility as when including only individuals infected, tested or admitted to the hospital [8]; and missing data are a constant threat. There are no easy solutions to any of these problems. Negative controls can be particularly helpful in these circumstances, and

often complex methodology and many bias analyses are required to ensure valid conclusions.

Despite these disadvantages, the crucial role played by observational studies based on secondary data during the Covid-19 pandemic cannot be ignored. As more high-quality data infrastructures are created, integrating data on background clinical and sociodemographic characteristics with real-time data on relevant exposures (e.g., vaccination) and outcomes (e.g., infections, hospitalizations, deaths), the role of such studies is projected to grow, both within the context of infectious disease epidemiology and beyond. This emphasizes a goal that healthcare organizations must strive for: creating integrated and high-quality clinical databases that can allow for reliable research.

Transparency declaration

References

[1] Vokó Zoltán, Kiss Zoltán, Surján György, Surján Orsolya, Barcza Zsófia, et al. Nationwide effectiveness of five SARS-CoV-2 vaccines in Hungary—the HUN-VE study. Clin Microbiol Infect 2022;28:398–404.

[2] Bar-On YM, Goldberg Y, Mandel M, Bodenheimer O, Freedman L, Kalkstein N, et al. Protection of BNT162b2 vaccine booster against Covid-19 in Israel. N Engl J Med 2021;385:1393–400.

[3] Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N Engl J Med 2020;383: 2603–15.

[4] Barda N, Dagan N, Ben-Shlomo Y, Kepten E, Waxman J, Ohana R, et al. Safety of the BNT162b2 mRNA Covid-19 vaccine in a nationwide setting. N Engl J Med 2021;385:1078–90.

[5] 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.

[6] Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman L, Haas EJ, et al. Waning immunity after the BNT162b2 vaccine in Israel. N Engl J Med 2021;385:e85.

[7] Barda N, Dagan N, Cohen C, Hernán MA, Lipsitch M, Kohane IS, et al. Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: an observational study. Lancet 2021;398: 2093–100.

[8] Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Comm 2020;11:5749.