



Using observational epidemiology to evaluate COVID-19 vaccines: integrating traditional methods with new data sources and tools

Catharine Chambers^{1,2}

Received: 6 January 2021 / Accepted: 10 June 2021 / Published online: 23 June 2021
© The Canadian Public Health Association 2021

Abstract

Although clinical trials are necessary for vaccine approval, observational epidemiology will be required to evaluate the long-term effectiveness, safety, and population impacts of newly approved COVID-19 vaccines under real-world field conditions. In this commentary, I argue that a hybrid approach that combines new data sources and tools, including COVID-19 vaccine registries, with traditional epidemiological methods will be needed to evaluate COVID-19 vaccines using observational epidemiology. Wherever possible, primary data collection, active surveillance, and linkage with existing population-based cohorts should be leveraged to supplement secondary data sources and passive surveillance systems. Evidence-informed public health decision making around provincial COVID-19 immunization programs will need to account for potential biases, incomplete or conflicting information, and heterogeneity across subpopulations.

Résumé

Des essais cliniques sont nécessaires à l'approbation des vaccins, mais il faudra recourir à l'épidémiologie d'observation pour évaluer en conditions réelles de terrain l'efficacité à long terme, l'innocuité et les effets sur les populations des vaccins contre la COVID-19 nouvellement approuvés. Dans ce commentaire, je fais valoir qu'il faudra adopter une démarche hybride combinant de nouvelles sources de données et de nouveaux outils, dont les registres de vaccins anti-COVID-19, et des méthodes épidémiologiques classiques pour évaluer les vaccins anti-COVID-19 à l'aide de l'épidémiologie d'observation. Dans la mesure du possible, il faudra utiliser la collecte de données primaires, la surveillance active et les maillages avec les cohortes populationnelles existantes pour compléter les sources de données secondaires et les systèmes de surveillance passive. Les décisions de santé publique éclairées par les données probantes sur les programmes d'immunisation provinciaux contre la COVID-19 devront tenir compte des biais possibles, des informations incomplètes ou contradictoires et de l'hétérogénéité des sous-populations.

Keywords COVID-19 · Vaccine · Epidemiology · Public health · Observational study

Mots-clés COVID-19 · vaccin · épidémiologie · santé publique · étude observationnelle

Introduction

Development of vaccines against SARS-CoV-2, the virus that causes coronavirus disease 2019 (COVID-19), has occurred at

an unprecedented pace. As of May 2021, Health Canada has authorized four COVID-19 vaccines under its interim order, with two other vaccines under review at the time of writing (Health Canada 2021). Based on early trial results, these vaccines are up to 95% efficacious against symptomatic infection (Baden et al. 2020; Polack et al. 2020; Sadoff et al. 2021; Voysey et al. 2021). Nevertheless, as with all newly approved vaccines, the real-world effectiveness and long-term safety of these vaccines will be unknown, even as they are deployed on a population scale. Large-scale, high-quality clinical trials that measure safety, immunogenicity, and efficacy are necessary for vaccine approval. However, these trials are typically restricted to young, healthy volunteers and may not reflect real-

✉ Catharine Chambers
catharine.chambers@mail.utoronto.ca

¹ Department of Public Health Sciences, Dalla Lana School of Public Health, University of Toronto, 155 College St, Toronto, ON M5T 3M7, Canada

² MAP Centre for Urban Health Solutions, St. Michael's Hospital, Unity Health Toronto, 30 Bond St, Toronto, ON M5B 1W8, Canada

world, field conditions (Hanquet et al. 2013). Although promising, early COVID-19 vaccine trials were underpowered to assess efficacy in certain subgroups or against asymptomatic transmission or severe outcomes such as hospitalization or death. Furthermore, as these trials were rapidly implemented in response to an emerging public health crisis, they were not designed to assess long-term efficacy. Finally, given widespread vaccine eligibility for adult age groups in many jurisdictions, including Canada, it has become unethical for many trials to continue without unblinding participants or to initiate new placebo-controlled trials (Polack et al. 2020; Singh and Upshur 2020).

Post-licensure evaluation will be required to understand how these vaccines work in specific populations (e.g., pregnant women, immunocompromised hosts, children), under certain scenarios (e.g., single-dose regimens, extended dose intervals), and in different settings (e.g., long-term care facilities). These types of evaluations will be critical as COVID-19 vaccines are being immediately rolled out in the population, and also to assess vaccine effectiveness over the long term, including the potential need for booster doses as new SARS-CoV-2 variants emerge. From a public health perspective, understanding the pragmatic effectiveness of COVID-19 vaccines and their real-world impacts—along with monitoring for rare adverse events on a population scale, such as the vaccine-induced thrombotic thrombocytopenia associated with viral vector-based vaccines—will be equally as important as knowing their clinical efficacy and safety in time-limited, controlled settings (Hanquet et al. 2013). In this commentary, I discuss the strengths and limitations of existing public health information systems and how emerging data sources and tools will be needed to evaluate COVID-19 immunization programs in Canada.

Using observational epidemiology for vaccine field evaluation

While randomized clinical trials measure vaccine efficacy against well-defined endpoints under ideal settings, observational studies measure vaccine effectiveness under real-life field conditions (Crowcroft and Klein 2018). Orenstein et al. (1988) outline four aspects of post-licensure vaccine evaluation using observational studies: (1) standardized case definitions, preferably using laboratory-confirmed outcomes; (2) equal case finding; (3) accurate vaccination status ascertainment; and (4) comparable exposure groups. Traditional observational study designs include cohort and case-control studies, and also newer approaches such as the test-negative design, a modified case-control study whereby vaccination status is compared between test-positive cases and test-negative controls that meet a prespecified clinical presentation (Crowcroft and Klein 2018; Hanquet et al. 2013). This design has been

used extensively in the evaluation of seasonal influenza vaccine (Chua et al. 2020), but may pose methodological challenges for severe COVID-19 infections (Patel et al. 2020).

As with all observational studies, post-licensure evaluations must strive to minimize potential sources of bias associated with selection of participants, misclassification of exposure or outcome, and confounding. Because vaccination is not randomly allocated by design, there is greater potential for confounding due to imbalance in measured covariates between exposure groups. Multivariable regression adjustment, propensity score models, and instrumental variable analysis are potential techniques that can be used to address confounding, but these require access to a sufficient set of covariates that are measured without error and/or availability of appropriate instruments.

Limitations of existing data sources for COVID-19 vaccine evaluation

The COVID-19 pandemic has exposed multiple weaknesses in our public health data infrastructure resulting from decades of underfunding. These include gaps in surveillance of high-risk settings such as long-term care facilities, absence of data on racialized and low socio-economic communities, and reliance on manual data entry and outdated ways to transmit personal health information. Despite notifiable disease regulations, passive surveillance systems will invariably miss COVID-19 cases. Indeed, serological evidence suggests that only about one in every eight or nine COVID-19 cases are captured in official case counts in Europe and the Americas (Chen et al. 2021). It is not always feasible or cost-efficient to test all probable cases, even during a global pandemic, and national and provincial case definitions include provision for clinical or epidemiologically linked cases. For those that are laboratory confirmed, current testing algorithms only capture individuals who can access COVID-19 assessment centres and will systematically underreport mild or asymptomatic cases that do not seek testing. As a result, surveillance data are typically skewed toward those with a higher pre-test likelihood of positivity, particularly in the early stages of the pandemic when testing was more limited. At the same time, they will overrepresent cases in certain demographic groups, such as essential workers or those living in congregate settings who are undergoing routine testing.

Similarly, administrative healthcare data are limited in their ability to evaluate immunization programs. In recent years, public health has moved away from traditional epidemiological methods (e.g., primary data collection, active case finding) toward using existing administrative data sources and electronic medical records. These secondary data were originally collected for insurance billing, healthcare, and health system management, not research purposes. They are often

missing data on key variables, such as symptom history or onset dates, that are required to standardize testing indications and minimize outcome misclassification. More often, these analyses rely on proxy or composite measures that cannot always be validated against external sources for completeness and accuracy. Although new ICD-10-CA diagnostic codes for COVID-19-related healthcare encounters have been recently added, these may not capture the full clinical spectrum, for example, “long COVID-19” or multisystem inflammatory syndrome in children, or those who do not access the healthcare system. There are often delays in accessing data, which may limit their potential to inform real-time decision making; however, some datasets have been made available more frequently since the COVID-19 pandemic.

One of the biggest challenges facing COVID-19 vaccine program evaluation in Canada is the absence of a national vaccine registry for adult immunizations (Crowcroft and Levy-Bruhl 2017). These registries must be population-wide in scale and large enough in scope to capture sufficient detail (e.g., vaccine type, lot number, number of doses, dates of immunization). They should also be routinely linked to notifiable diseases and other public health information systems to enable vaccine effectiveness evaluation and adverse event monitoring. Although some provinces have recently introduced surveillance systems to collect data on routine childhood immunizations (Wilson et al. 2017a), these systems will require major investment and expansion for COVID-19. Administrative billing data for certain vaccines exist in some provinces (and are anticipated for COVID-19 vaccines), but in others may only capture immunizations administered in certain settings (e.g., physicians’ offices). There is some progress being made where more complete immunization registries exist, but data quality issues persist, and methods for assessing immunization coverage are not standardized across provinces (Wilson et al. 2017a; Wilson et al. 2017b).

Leveraging existing data and the need for new data sources and analytic tools

Canadian researchers now have unparalleled access to population-based administrative data that combine various sources of health and health services data. The main advantage is their size and scope, allowing for the potential evaluation of rare outcomes or comparisons of specific exposures over the entire population. For example, administrative data from Alberta were used to compare live-attenuated and inactivated influenza vaccines in children following reports of reduced effectiveness against the A(H1N1)pdm09 component (Buchan et al. 2018). Similarly, public health information systems matched to administrative data held at ICES in Ontario were used to evaluate waning immunity from the acellular pertussis vaccine (Crowcroft et al. 2019).

Given the expected scale of COVID-19 vaccine rollout and multiple vaccine products, the creation of new data systems or reconfiguration of existing ones to serve as comprehensive, linked public health information systems will be critical. Wherever possible, researchers should leverage existing population-based cohorts that are already collecting detailed demographic, socio-behavioural, and/or clinical variables, as these data are often not routinely available from administrative records. Active surveillance of cohort members for potential exposures and COVID-19-related symptoms along with biological specimen collection to detect incident or past infections should also be considered. Through data linkage, these studies can be used to fill in the gaps of our existing data platforms and passive surveillance systems. These linkages will enable identification of risk groups and prioritization of targeted interventions like immunizations, along with subgroup analyses and confounder adjustment to minimize biases associated with observational study designs.

Alongside these new data sources are novel analytical methods, such as regression discontinuity and difference-in-difference approaches, prominent in economics and other fields. These quasi-experimental methods have been used, for example, to evaluate the implementation of school-based programs for human papillomavirus vaccine (Smith et al. 2015) or the impact of pneumococcal vaccination on pneumonia hospitalizations in Ontario (Luca et al. 2018). Machine learning algorithms and “big data” will likely also play a role, particularly for electronic health records and genomics. However, as with secondary data, these newer approaches must be validated to ensure construct validity and reliability and minimize the risk of measurement error, selection bias, and other researcher-introduced biases.

Conclusion

To move the field of vaccine evaluation forward using observational epidemiology, a hybrid approach is needed that integrates these new data sources and tools with traditional epidemiological approaches. We must embrace new data sources and the advantages they offer with respect to data availability and computational power. We must utilize new tools and methodologies to improve automation, data linkage, and knowledge transfer. However, in doing so, we must appropriately acknowledge their limitations and carefully consider their potential biases. Part of the way forward will require a return to traditional epidemiological methods and the principles of field evaluation (Orenstein et al. 1988), for example, employing active surveillance methods or linking cohort studies to administrative data. Our governments and funding agencies must also continue to invest in primary data collection and longitudinal research. Ultimately, integrating the principles of epidemiology that emphasize minimizing bias through

thoughtful study design and careful measurement with new data sources and tools will improve evidence-informed decision making for our COVID-19 immunization programs.

Acknowledgements I would like to thank Drs. Ann Burchell, Shelley Deeks, Jeff Kwong, and Laura Rosella for their thoughtful review and helpful comments on earlier versions of this manuscript.

Availability of data and material Not applicable.

Code availability Not applicable.

Author contributions CC conceived of this commentary, drafted the article, and critically revised it for important intellectual content.

Funding CC is a Vanier scholar (Canadian Institutes of Health Research, #415141) and Canadian Immunization Research Network trainee.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The author declares no competing interests.

References

- Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Roupael, N., Creech, C. B., McGettigan, J., Kehtan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., Gilbert, P., Janes, H., Follmann, D., Marovich, M., Mascola, J., Polakowski, L., Ledgerwood, J., Graham, B. S., Bennett, H., Pajon, R., Knightly, C., Leav, B., Deng, W., Zhou, H., Han, S., Ivarsson, M., Miller, J., Zaks, T., & COVE Study Group. (2020). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMoa2035389>.
- Buchan, S. A., Booth, S., Scott, A. N., Simmonds, K. A., Svenson, L. W., Drews, S. J., Russell, M. L., Crowcroft, N. S., Loeb, M., Warshawsky, B. F., & Kwong, J. C. (2018). Effectiveness of live attenuated vs inactivated influenza vaccines in children during the 2012–2013 through 2015–2016 influenza seasons in Alberta, Canada: A Canadian Immunization Research Network (CIRN) study. *JAMA Pediatrics*, *172*(9), e181514.
- Chen, X., Chen, Z., Azman, A. S., Deng, X., Sun, R., Zhao, Z., Zheng, N., Lu, W., Zhuang, T., Yang, J., Viboud, C., Ajelli, M., Leung, D. T., & Yu, H. (2021). Serological evidence of human infection with SARS-CoV-2: A systematic review and meta-analysis. *Lancet Global Health*, *9*(5), e598–e609.
- Chua, H., Feng, S., Lewnard, J. A., Sullivan, S. G., Blyth, C. C., Lipsitch, M., & Cowling, B. J. (2020). The use of test-negative controls to monitor vaccine effectiveness: A systematic review of methodology. *Epidemiology*, *31*(1), 43–64.
- Crowcroft, N. S., & Klein, N. P. (2018). A framework for research on vaccine effectiveness. *Vaccine*, *36*(48), 7286–7293.
- Crowcroft, N. S., & Levy-Bruhl, D. (2017). Registries: An essential tool for maximising the health benefits of immunisation in the 21st century. *Euro Surveillance*, *22*(17), 30523.
- Crowcroft, N. S., Schwartz, K. L., Chen, C., Johnson, C., Li, Y., Marchand-Austin, A., Bolotin, S., Jamieson, F. B., Drews, S. J., Russell, M. L., Svenson, L. W., Simmonds, K., Mahmud, S. M., & Kwong, J. C. (2019). Pertussis vaccine effectiveness in a frequency matched population-based case-control Canadian Immunization Research Network study in Ontario, Canada 2009–2015. *Vaccine*, *37*(19), 2617–2623.
- Hanquet, G., Valenciano, M., Simondon, F., & Moren, A. (2013). Vaccine effects and impact of vaccination programmes in post-licensure studies. *Vaccine*, *31*(48), 5634–5642.
- Health Canada. (2021). Drug and vaccine authorizations for COVID-19: List of applications received. Government of Canada. <https://www.canada.ca/en/health-canada/services/drugs-health-products/covid19-industry/drugs-vaccines-treatments/authorization/applications.html>.
- Luca, D. L., Kwong, J. C., Chu, A., Sander, B., O'Reilly, R., McGeer, A. J., & Bloom, D. E. (2018). Impact of pneumococcal vaccination on pneumonia hospitalizations and related costs in Ontario: A population-based ecological study. *Clinical Infectious Diseases*, *66*(4), 541–547.
- Orenstein, W. A., Bernier, R. H., & Hinman, A. R. (1988). Assessing vaccine efficacy in the field. Further observations. *Epidemiologic Reviews*, *10*, 212–241.
- Patel, M. M., Jackson, M. L., & Ferdinands, J. (2020). Postlicensure evaluation of COVID-19 vaccines. *JAMA*. <https://doi.org/10.1001/jama.2020.19328>.
- Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. V., Cooper, D., Frenck, R. W., Hammitt, L. L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D. B., Mather, S., Dormitzer, P. R., Şahin, U., Jansen, K. U., Gruber, W. C., & C4591001 Clinical Trial Group. (2020). Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMoa2034577>.
- Sadoff, J., Gray, G., Vandebosch, A., Cárdenas, V., Shukarev, G., Grinsztejn, B., Goepfert, P. A., Truyers, C., Fennema, H., Spiessens, B., Offergeld, K., Scheper, G., Taylor, K. L., Robb, M. L., Treanor, J., Barouch, D. H., Stoddard, J., Ryser, M. F., Marovich, M. A., Neuzil, K. M., Corey, L., Cauwenberghs, N., Tanner, T., Hardt, K., Ruiz-Guiñazú, J., Le Gars, M., Schuitemaker, H., Van Hoof, J., Struyf, F., Dougouih, M., & ENSEMBLE Study Group. (2021). Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJMoa2101544>.
- Singh, J. A., & Upshur, R. E. G. (2020). The granting of emergency use designation to COVID-19 candidate vaccines: Implications for COVID-19 vaccine trials. *Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30923-3](https://doi.org/10.1016/S1473-3099(20)30923-3).
- Smith, L. M., Strumpf, E. C., Kaufman, J. S., Lofters, A., Schwandt, M., & Lévesque, L. E. (2015). The early benefits of human papillomavirus vaccination on cervical dysplasia and anogenital warts. *Pediatrics*, *135*(5), e1131–e1140.
- Voysey, M., Clemens, S. A. C., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., Angus, B., Baillie, V. L., Barnabas, S. L., Bhorat, Q. E., Bibi, S., Briner, C., Cicconi, P., Collins, A. M., Colin-Jones, R., Cutland, C. L., Darton, T. C., Dheda, K., Duncan, C. J. A., Emary, K. R. W., Ewer, K. J., Fairlie, L., Faust, S. N., Feng, S., Ferreira, D. M., Finn, A., Goodman, A. L., Green, C. M., Green, C. A., Heath, P. T., Hill, C., Hill, H., Hirsch, I., Hoddgson, S. H. C., Izu, A., Jackson, S., Jenkin, D., Joe, C. C. D., Kerridge, S., Koen, A., Kwatra, G., Lazarus, R., Lawrie, A. M., Lelliott, A., Libri, V., Lillie, P. J., Mallory, R., Mendes, A. V. A., Milan, E. P., Minassian, A. M.,

- McGregor, A., Morrison, H., Mujadidi, Y. F., Nana, A., O'Reilly, P. J., Padayachee, S. D., Pittella, A., Plested, E., Pollock, K. M., Ramasamy, M. N., Rhead, S., Schwarzbald, A. V., Singh, N., Smith, A., Song, R., Snape, M. D., Sprinz, E., Sutherland, R. K., Tarrant, R., Thomson, E. C., Török, M. E., Toshner, M., Turner, D. P. J., Vekemans, J., Villafana, T. L., Watson, M. E. E., Williams, C. J., Douglas, A. D., Hill, A. V. S., Lambe, T., Gilbert, S. C., Pollard, A. J., & Oxford COVID Vaccine Trial Group (2021). Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet*, 397(10269), 99–111.
- Wilson, S. E., Quach, S., MacDonald, S. E., Naus, M., Deeks, S. L., Crowcroft, N. S., Mahmud, S. M., Tran, D., Kwong, J. C., Tu, K., Johnson, C., & Desai, S. (2017a). Immunization information systems in Canada: Attributes, functionality, strengths and challenges. A Canadian Immunization Research Network study. *Canadian Journal of Public Health*, 107(6), e575–e582.
- Wilson, S. E., Quach, S., MacDonald, S. E., Naus, M., Deeks, S. L., Crowcroft, N. S., Mahmud, S. M., Tran, D., Kwong, J., Tu, K., Gilbert, N. L., Johnson, C., & Desai, S. (2017b). Methods used for immunization coverage assessment in Canada, a Canadian Immunization Research Network (CIRN) study. *Human Vaccines and Immunotherapeutics*, 13(8), 1928–1936.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.