VIEWPOINT

Detecting Selection Bias in Observational Studies—When Interventions Work Too Fast

Ghulam Rehman Mohyuddin, MD
Division of Hematology, Huntsman Cancer Center, University of Utah, Salt Lake City.

Vinay Prasad, MD
Division of Epidemiology, University of California, San Francisco.

Observational studies have an important role in advancing medical knowledge. They may clarify the incidence and prevalence of disease, provide useful information on natural history and prognosis, and facilitate the development of clinical risk scores. However, when it comes to assessing the efficacy of interventions, such as surgeries, drugs, medical devices, or radiotherapy, observational studies have well-recognized limitations. Observational studies can find associations, not cause-and-effect relationships. If misapplied, the findings of observational studies may lead to overuse, as well as underuse, of medical interventions. To validate the findings of observational studies, randomized clinical trials are often needed.

Attempts have been made to improve the reliability of observational research regarding causal inferences, including the development of target trial framework[1] and the inclusion of falsification end points.[2] The target trial emulation method asks investigators to formulate their research question as a hypothetical randomized clinical trial. Typically, there are a logical and clear enrollment period and rules. In some cases, such as the association of steroids with mortality related to COVID-19, targeted trial emulations have faithfully replicated the results of randomized clinicaltrials.[3] Falsification end points are outcomes that are considered unlikely to be associated with a therapy but reveal persistent confounding. If these end points are prespecified and null, it strengthens the case for an inference that the intervention is associated with the outcome of interest. For example, COVID-19 vaccination should be associated with lower rates COVID-19–related deaths but have no association with the rates of car crashes. Imbalances in car crashes between groups in a study of the effects of COVID-19 vaccination would suggest residual confounding in the observational data set.

Early separation of the Kaplan-Meier curves is another mechanism for detecting residual confounding in observational studies. The Kaplan-Meier survival curve is a graphical representation of time-to-event end points and allows for maximal use of each participant's time-related data. In a Kaplan-Meier analysis, participants contribute to the survival estimate until the event of interest occurs (eg, death, progression of disease) or until they are censored (eg, loss to follow-up or mandatory data lock).

Many interventions cannot work immediately. For example, vaccinations may take 1 to 2 weeks to elicit an immune response and show therapeutic efficacy. After surgical interventions aimed at reducing long-term complications from a procedure, differences in outcomes may not emerge for weeks. However, sometimes observational studies of such interventions show a clear separation of Kaplan-Maier curves for outcomes, such as overall survival or hospitalizations, within the first day of follow-up. It is biologically inconceivable for such an effect to take place. Instead, the results suggest residual confounding. Examples include the association of bisphosphonates with mortality[4] and the effect of geriatrician involvement following serious traumatic injury,[5] among others.[6] The **Figure** shows a visual representation of this construct, in which the curves separate from the beginning of the follow-up period.

Three recent examples illustrate the importance of recognizing situations in which apparent clinical benefits may accrue too fast. First, in an observational study of COVID-19 booster vaccinations, researchers examined the efficacy of a fourth vaccine dose on mortality.[7] Although the study found a reduced risk of death associated with a fourth vaccine dose, the Kaplan-Meier curves begin to separate on day 8 after vaccination. Given that COVID-19 boosters do not immediately activate the immune system and protect against COVID-19 infection, and that the time from contracting COVID-19 to death is often days or weeks, a clinical benefit that begins 8 days after vaccinations seems implausibly early. Perhaps the people who promptly sought out the fourth vaccine dose were more health conscious, took more precautions against COVID-19, and were healthier than those who received the fourth vaccine dose later or did not receive it at all.

Second, consider an observational study that compared patients with lymphoma receiving chimeric antigen receptor T-cell (CAR-T) therapy as part of a clinical trial with a historical cohort of patients receiving standard-of-care salvage chemotherapy.[8] Although CAR-T therapy has been transformative for patients with lymphoma, the therapy requires time to manufacture. Moreover, in the initial days after therapy, patients are vulnerable to the cytokine release syndrome, a runaway immune response which may lead to severe illness and death. When compared retrospectively with chemotherapy, CAR-T therapy appeared to show benefit with the Kaplan-Meier curves for patient survival separating at day 0, with a clear difference in survival observed within the first few weeks of follow-up. However, such a finding is implausible for a CAR-T therapy with its method of delivery, action, and risk of early treatment-related mortality. A more likely explanation is that the patients in the historical cohort were in poorer overall health and had different cancer characteristics than the patients in the clinical trial.

The third example involves a percutaneous left atrial appendage occlusion device, which is used to seal the left atrial appendage and lower the risk of stroke among patients with atrial fibrillation. A recent analysis of a 2021

Corresponding Author: Ghulam Rehman Mohyuddin, MD, Huntsman Cancer Institute, University of Utah, 1950 Circle of Hope Dr, Salt Lake City, UT 84093 (g.mohyuddin@hci.utah.edu).

Figure. Hypothetical Curve Demonstrating Implausibly Early Separation of Curves

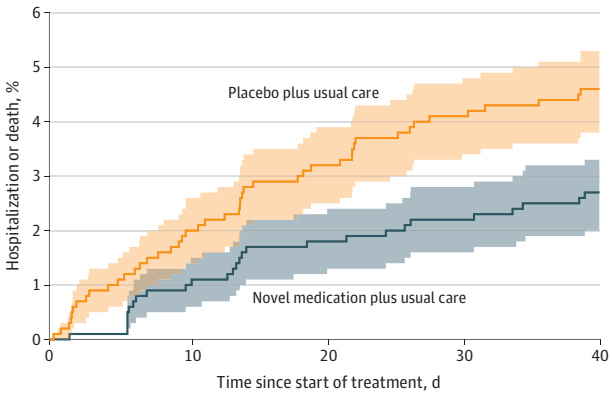

Illustration of a hypothetical curve in which the event of interest happens implausibly early and the curves separate from the beginning. In this hypothetical scenario, the gray line represents rates of hospitalizations following administration of an oral outpatient medication for COVID-19, and the orange line represents rates of hospitalizations for a control group. The curves separate from day 0, which is biologically impossible, as the medication cannot prevent hospitalization immediately after administration. The shaded area represents the confidence interval.

observational study that compared the use of left atrial appendage occlusion and direct oral anticoagulation found that the Kaplan-Meier curves for all-cause mortality separated at day 0 after the devices were inserted and the size of the effect increased over time.[9,10] However, it is biologically impossible for a stroke prevention device to save lives soon after it is inserted; in patients with atrial fibrillation, the risk of stroke increases over time. During the immediate postprocedure period, the survival curves for patients who received the procedure or were treated with direct oral anticoagulation should be superimposable. Residual confounding or confounding by indication (patients with more severe disease may not receive the left atrial appendage occlusion device) are more plausible explanations for the early separation of the Kaplan-Meier curves.

A limitation to this approach to detecting selection bias in observational studies is that it requires knowledge of the pathophysiology and natural history of a disease, the expected efficacy of the interventions, or other expertise in the subject of the study. Another is that when observational studies are published, Kaplan-Meier curves are often not included. Instead, there are tables showing relative risks or odds ratios for the comparisons between the groups of patients that are being observed; such tables do not show changes in the relative risks or odds ratios that may occur over time. To allow for the ascertainment of bias, such as imbalances between the characteristics of cohorts and residual confounding or confounding by indication, observational studies that examine a time-to-event end point or contain time-to-event data should report this information in a graphical form. When interventions appear to work too fast, the findings of a study may be too good to be true.

ARTICLE INFORMATION

Published Online: June 12, 2023. doi:10.1001/jamainternmed.2023.2067

Conflict of Interest Disclosures: Dr Prasad reported grants from Arnold Ventures during the submitted work as well as personal fees from Johns Hopkins Press, Medscape, MedPage, UnitedHealthcare, OptumRx, Substack, YouTube, and Patreon outside the submitted work. No other disclosures were reported.

REFERENCES

1. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA*. 2022;328(24):2446-2447. doi:10.1001/jama.2022.21383

2. Prasad V, Jena AB. Prespecified falsification end points: can they validate true observational associations? *JAMA*. 2013;309(3):241-242. doi:10.1001/jama.2012.96867

3. Hoffman KL, Schenck EJ, Satlin MJ, et al. Comparison of a target trial emulation framework vs Cox regression to estimate the association of corticosteroids with COVID-19 mortality. *JAMA Netw Open*. 2022;5(10):e2234425. doi:10.1001/jamanetworkopen.2022.34425

4. Bergman J, Nordström A, Hommel A, Kivipelto M, Nordström P. Bisphosphonates and mortality: confounding in observational studies? *Osteoporos Int*. 2019;30(10):1973-1982. doi:10.1007/s00198-019-05097-1

5. van Dalen SL, Goudzwaard JA, van den Hoogan MWF. Geriatrician assessment and immortal time bias in the FiTR 2 study. *Lancet Healthy Longev*. 2022;3(11):e735. doi:10.1016/S2666-7568(22)00196-9

6. van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. *J Clin Epidemiol*. 2004;57(7):672-682. doi:10.1016/j.jclinepi.2003.12.008

7. Arbel R, Sergienko R, Friger M, et al. Effectiveness of a second BNT162b2 booster vaccine against hospitalization and death from COVID-19 in adults aged over 60 years. *Nat Med*. 2022;28(7):1486-1490. doi:10.1038/s41591-022-01832-0

8. Neelapu SS, Locke FL, Bartlett NL, et al. Comparison of 2-year outcomes with CAR T cells (ZUMA-1) vs salvage chemotherapy in refractory large B-cell lymphoma. *Blood Adv*. 2021;5(20):4149-4155. doi:10.1182/bloodadvances.2020003848

9. Ruzieh M. An easy way to spot bias in observational studies. Accessed March 15, 2023. https://sensiblemed.substack.com/p/an-easy-way-to-spot-bias-in-observational

10. Nielsen-Kudsk JE, Korsholm K, Damgaard D, et al. Clinical outcomes associated with left atrial appendage occlusion versus direct oral anticoagulation in atrial fibrillation. *JACC Cardiovasc Interv*. 2021;14(1):69-78. doi:10.1016/j.jcin.2020.09.051