JAMA Guide to Statistics and Methods

Target Trial Emulation
A Framework for Causal Inference From Observational Data

Miguel A. Hernán, MD, DrPH; Wei Wang, PhD; David E. Leaf, MD, MMSc

Quantifying the effect of a treatment on a clinical outcome—causal inference—requires the comparison of outcomes under different courses of action. For example, to quantify the effect of tocilizumab on mortality in critically ill patients with COVID-19, the mortality risk could be compared between a group of patients administered tocilizumab and a group who are not. Ideally, eligible patients would be assigned to these groups at random. The key advantage of such a randomized trial is that both groups are expected to be comparable, and thus any differences in mortality can be attributed to tocilizumab rather than to prognostic differences between the groups.

There are additional reasons randomized trials support causal inference. In a randomized trial, the start of follow-up (time zero) for each participant is clearly specified (time of randomization), as is the assigned treatment group. This clarity regarding time zero and treatment assignment is often taken for granted when discussing the advantages of randomized trials. However, the importance of these features becomes clearer when considering failures in drawing causal conclusions from observational data.

One way to ensure that observational analyses preserve these desirable features of randomized trials is to design them so that they explicitly emulate a hypothetical randomized trial that would answer the question at hand: the *target trial*.[1] In a recent study using this approach, Gupta et al[2] used observational data from nearly 4000 critically ill patients with COVID-19 from 68 US hospitals to estimate the effect on mortality of tocilizumab administered within 2 days following admission to the intensive care unit (ICU).

What Is Target Trial Emulation in the Analysis of Observational Data?

Target trial emulation is a 2-step process. The first step is articulating the causal question in the form of the protocol of a hypothetical randomized trial that would provide the answer. The protocol must specify certain key elements that define the causal estimands (eligibility criteria, treatment strategies, treatment assignment, the start and end of follow-up, outcomes, causal contrasts) and the data analysis plan.[1] The randomized trial described in the protocol becomes the target study for the causal inference of interest.

The second step is explicitly emulating the components of that protocol using the observational data: finding eligible individuals, assigning them to a treatment strategy compatible with their data, following them up from assignment (time zero) until outcome or end of follow-up, and conducting the same analysis as the corresponding target trial, except that there is adjustment for baseline confounders in an attempt to emulate random treatment assignment. Sometimes there is ambiguity regarding assignment to a treatment group. For example, in the study by Gupta et al[2] the tocilizumab treatment group could have the drug started within 2 days of ICU admission. So, during the first 2 days, a patient not being administered tocilizumab could be considered a potential member of either treatment group. To avoid immortal time bias[3] and ensure time zero is considered correctly, such a patient may be "cloned" and, until 2 days have passed or tocilizumab is started, be represented in both treatment groups.[4]

Why Is Target Trial Emulation Used in the Analysis of Observational Data?

The goal of target trial emulation is to avoid making fundamental errors that can result in erroneous causal conclusions. For example, a randomized trial found an increased risk of coronary heart disease among postmenopausal women assigned to estrogen plus progestin hormone therapy compared with placebo, but observational analyses failed to detect this elevated risk.[5]

It is plausible that the observational estimate was biased because users and nonusers of hormone therapy had different prognostic factors. Instead, the bias resulted mostly from comparing women who had been using hormone therapy for some time (current users) with nonusers,[5] a comparison that would be avoided in both a randomized trial and in the analysis of observational data that emulates the target trial. By considering current users, the observational analysis implicitly set the start of follow-up long after therapy initiation. As a result, early coronary events were ignored and selection bias arose because the population of current users was partly depleted of women susceptible to heart disease.[6] If the randomized trial data were also incorrectly analyzed by deleting data from the early follow-up, an apparently beneficial estimate of hormone therapy would similarly have been found. After harmonizing the analysis of the randomized and observational data to eliminate this bias, the effect estimates were consistent.[5]

Limitations of Target Trial Emulation

Explicit target trial emulation alone cannot eliminate the bias that arises from lack of randomization—confounding from noncomparable treatment groups—even if the observational analysis correctly emulates all other components of the target trial. Thus, a successful target trial emulation requires detailed data not only on treatment and outcome but also confounders. Some sources of routinely collected data (eg, administrative claims databases) may have reasonably detailed data on treatments and outcomes but insufficient data on clinical factors that require adjustment. The key advantage of a correct target trial emulation is that it eliminates other common sources of bias so attention can be focused on confounding.

The use of observational data to emulate target trials is subject to other limitations. Observational databases include only information on treatment strategies actually used in clinical settings, and thus they cannot be used to emulate a target trial of novel treatments. Also, the protocol of emulated target trials cannot include blinded

treatment assignment (eg, using a placebo control) and blinded outcome ascertainment because these are not present in routine clinical practice. Target trials emulated with observational data are necessarily pragmatic trials.

Use of Target Trial Emulation in the Study by Gupta et al

The study by Gupta et al[2] emulated a target trial of tocilizumab in adults admitted to the ICU with COVID-19, using data specifically collected for COVID-19 research across 68 US hospitals. The investigators first outlined the protocol of the target trial: adults with laboratory-confirmed COVID-19 admitted to an ICU from March 4 to May 10, 2020, were randomly assigned to either initiation or no initiation of tocilizumab within 2 days following ICU admission. Individuals were then followed up until the first of in-hospital death, discharge, or June 12, 2020. An intention-to-treat analysis would compare the mortality between groups.

To emulate this target trial, Gupta et al[2] identified 3924 individuals who met the eligibility criteria and classified them into the tocilizumab group (433 individuals) or the control group depending on whether they did or did not start treatment with tocilizumab in the first 2 days after ICU admission. The investigators adjusted for measured confounders using inverse probability weighting.[7]

Interpreting the Results From the Study by Gupta et al

The study by Gupta et al[2] estimated a 30-day mortality risk estimate of 27.5% in the tocilizumab group and 37.1% in the control group (hazard ratio, 0.71 [95% CI, 0.56-0.92]). This lower mortality risk in the tocilizumab group was later replicated in a large, randomized trial of critically ill patients with COVID-19.[8]

The successful emulation of this target trial relied on 2 factors. First, the initiation of the 2 treatment strategies was synchronized with the eligibility criteria at time zero of follow-up. This synchronization of eligibility and treatment assignment at time zero is a key principle of study design that arises naturally in randomized trials but that has often been violated in observational analyses.[4] To emulate the target trial, Gupta et al had to synchronize eligibility and treatment assignment at time zero without using later information to classify individuals into a treatment strategy.

Second, adequate confounding adjustment was possible because detailed clinical data were collected at each of the participating hospitals. Compared with individuals in the control group, those in the tocilizumab group were younger and had a lower prevalence of comorbidities but had a higher prevalence of invasive mechanical ventilation and hypoxemia on ICU admission. Inverse probability weighting was used to balance these characteristics across groups, but other adjustment methods (eg, outcome regression, propensity scores[9]) would have been valid too, because potential confounders were all ascertained at time zero. The key issue for interventions like initiation of tocilizumab is the availability of data on confounders, not the adjustment method.

A reanalysis of the same data set that simply compared individuals who did and did not start tocilizumab at any time during the follow-up and that did not carefully adjust for clinical characteristics at time zero showed a mortality hazard ratio of 0.91 (95% CI, 0.79-1.06). That is, an observational analysis that does not adequately emulate a target trial suggests a considerably smaller mortality benefit of tocilizumab.

The explicit emulation of a tocilizumab target trial improves the interpretation of the results from the observational study, focusing on 2 questions: Does the observational analysis include adjustment for the most important confounders? If not, which additional confounders should have been adjusted for, and what is the expected direction of bias due to insufficient adjustment? In the tocilizumab study, the most important confounders were included in the adjustment.

The explicit emulation of a target trial had additional interpretational advantages. It allowed Gupta et al to construct a CONSORT-like flowchart of eligible individuals and to estimate the absolute risk of death under each treatment strategy. In contrast, many traditional observational analyses only provide hazard ratios, which are less useful to support clinical decision-making.[10]

ARTICLE INFORMATION

Author Affiliations: CAUSALab, Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts (Hernán); Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts (Wang); Division of Renal Medicine, Brigham and Women's Hospital, Boston, Massachusetts (Leaf).

Corresponding Author: Miguel Hernán, MD, DrPH, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115 (miguel_hernan@post.harvard.edu).

Section Editor: Roger J. Lewis, MD, PhD, JAMA Statistical Editor.

Published Online: December 12, 2022. doi:10.1001/jama.2022.21383

Conflict of Interest Disclosures: Dr Hernán reported receiving grants from the National Institutes of Health (NIH); serving as data science adviser for ProPublica; and serving as a consultant for Cytel. Dr Leaf reported receiving grants from the NIH (R01HL144566, R01DK125786). No other disclosures were reported.

REFERENCES

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758-764. doi:10.1093/aje/kwv254

2. Gupta S, Wang W, Hayek SS, et al; STOP-COVID Investigators. Association between early treatment with tocilizumab and mortality among critically ill patients with COVID-19. *JAMA Intern Med*. 2021;181(1):41-51. doi:10.1001/jamainternmed.2020.6252

3. Yadav K, Lewis RJ. Immortal time bias in observational studies. *JAMA*. 2021;325(7):686-687. doi:10.1001/jama.2020.9151

4. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-75. doi:10.1016/j.jclinepi.2016.04.014

5. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766-779. doi:10.1097/EDE.0b013e3181875e61

6. Stensrud MJ, Valberg M, Røysland K, Aalen OO. Exploring selection bias by causal frailty models: the magnitude matters. *Epidemiology*. 2017;28(3):379-386. doi:10.1097/EDE.0000000000000621

7. Adler AI, Latimer NR. Adjusting for nonadherence or stopping treatments in randomized clinical trials. *JAMA*. 2021;325(20):2110-2111. doi:10.1001/jama.2021.2433

8. Gordon AC, Mouncey PR, Al-Beidh F, et al; REMAP-CAP Investigators. Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N Engl J Med*. 2021;384(16):1491-1502. doi:10.1056/NEJMoa2100433

9. Haukoos JS, Lewis RJ. The propensity score. *JAMA*. 2015;314(15):1637-1638. doi:10.1001/jama.2015.13480

10. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13-15. doi:10.1097/EDE.0b013e3181c1ea43