STATISTICS IN MEDICINE

"Target Trial Emulation" for Observational Studies — Potential and Pitfalls

Rebecca A. Hubbard, Ph.D., Constantine A. Gatsonis, Ph.D., Joseph W. Hogan, Sc.D., David J. Hunter, M.B., B.S., Sharon-Lise T. Normand, Ph.D., and Andrea B. Troxel, Sc.D.

Randomized, controlled trials (RCTs) are the standard for evaluating the relative safety and efficacy of medical interventions. Although analyses of observational data from epidemiologic studies and health care databases, including electronic health records (EHRs) and medical claims data, offer benefits that include large sample sizes, timely availability of data, and the ability to assess "real world" effectiveness, they are vulnerable to biases that diminish the strength of the evidence they can generate. Designing observational studies according to the principles of RCTs has long been proposed as a way of improving the validity of these studies' findings. Though many methodologic approaches seek to draw causal inferences from observational data, investigators are increasingly modeling the design of observational studies on hypothetical RCTs, under the rubric "target trial emulation."[1]

The target trial emulation framework entails design and analysis of an observational study in alignment with a hypothetical RCT that addresses the same study question. Although this approach has the potential to improve the quality of observational studies by providing a structured method for design, analysis, and reporting, investigations conducted in this manner remain vulnerable to many sources of bias, including confounding by unobserved covariates. Such studies require careful specification of design elements, analytic methods for addressing confounding, and reporting of sensitivity analyses.

In a study that uses the target trial emulation approach, investigators specify the hypothetical RCT that would ideally be conducted to address a given study question and then specify the design elements of an observational study aligned with the elements of that "target" RCT. Necessary design elements include eligibility criteria, participant selection, treatment strategies, treatment assignment, start and end of follow-up, outcome measure, efficacy assessment, and statistical analysis plan (SAP).[1,2] For example, Dickerman et al. used the target trial emulation framework to assess the comparative effectiveness of the BNT162b2 and mRNA-1273 vaccines in preventing SARS-CoV-2 infection, hospitalization, and death with the use of EHR data from the Department of Veterans Affairs (VA).[3]

A key aspect of target trial emulation is specification of "time zero," the point at which participant eligibility is assessed, treatment assignment is made, and follow-up begins. In the VA Covid-19 vaccine study, time zero was defined as the date when the first dose of vaccine was administered. Alignment of eligibility determination, treatment assignment, and the beginning of follow-up at time zero reduces important sources of bias — in particular, immortal time bias, which arises when treatment strategies are defined after the start of follow-up, and selection bias, which arises when follow-up starts after treatment assignment. In the VA Covid-19 vaccine study, assigning participants to treatment groups for analyses on the basis of receipt of the second vaccine dose while starting follow-up at the time of administration of the first dose would result in immortal time bias; assigning treatment group at the time of the first dose but starting follow-up at the time of the second would result in selection bias, since only people receiving both doses would be included.

Target trial emulation can also help avert lack of clarity in the treatment-effect definition, a common challenge in observational studies. In the VA Covid-19 vaccine study, participants were matched according to baseline characteristics, and treatment effectiveness was assessed on the basis of differences in risk of the outcomes at 24 weeks. This approach clearly defines the effectiveness estimate as the difference in Covid-19 outcomes between vaccinated populations that are otherwise balanced with respect to baseline characteristics — similar to the efficacy

estimate in an RCT with respect to the same question. As the authors of that study note, comparing two similar vaccines is probably much less susceptible to confounding than comparing outcomes in people who received a vaccine with those who did not.

Even with successful alignment, the validity of a study that uses the target trial emulation framework rests on various assumptions, choices about design and analysis, and the quality of the underlying data. Though the validity of RCT results also depends on the quality of design and analysis, results from observational studies are additionally threatened by confounding. As nonrandomized studies, they lack RCTs' protection against confounding, and lack of blinding among participants and clinicians could influence outcome assessment and study results. In the VA Covid-19 vaccine study, matching was used to balance the distribution of baseline participant characteristics including age, sex, race, and urbanicity of residence between groups. Differences in the distribution of additional characteristics, such as occupation, that might also be associated with Covid-19 infection risk, would cause residual confounding.

Many studies using the target trial emulation approach use "real world data" (RWD) such as those from EHRs. The advantages of RWD, including timeliness, generalizability, and capture of treatment patterns as they occur in routine practice, must be balanced against problems stemming from data quality, including missing data, inaccuracies and inconsistencies in ascertainment and definition of participant characteristics and outcomes, lack of uniformity in application or administration of therapies, differential frequency of follow-up assessment, and missing data and loss to follow-up as the result of participants moving between health systems.[4] Use of data from a single EHR limits concerns about data inconsistency in the VA study. However, incomplete ascertainment and recording of variables including coexisting conditions and outcomes remained a risk.

The selection of participants into the analysis sample is typically based on retrospective data, which may result in selection bias caused by excluding people with missing baseline information. Though these problems are not unique to observational studies, they constitute residual sources of bias that are not directly addressed by target trial emulation. In addition, observational studies typically are not preregistered, which exacerbates problems of design sensitivity and publication bias. Because results may vary dramatically among data sources and design and analytic approaches, prespecification of the study design, analysis, and rationale for the choice of data source is essential.

Several best practices for conducting and reporting studies using the target trial emulation framework can strengthen their quality and ensure that they are reported in adequate detail to support rigorous evaluation by readers. To begin with, a protocol and SAP should be prespecified before data analysis. The SAP should include details of the statistical approach that will be used to address bias due to confounding and sensitivity analyses to evaluate the robustness of results to key sources of bias, including confounding and missing data.

The title, abstract, and methods section should clearly identify the study design as observational to avoid confusion with RCTs and distinguish between the observational study that has been conducted and the hypothetical trial it seeks to emulate. Investigators should provide thorough descriptions of their data source, quality metrics such as reliability and validity of data elements, and if possible, references to other published studies using the data source. They should also include a table outlining the design elements of the target trial and its observational emulation, and should clearly specify the time at which eligibility is determined, follow-up starts, and treatment assignment is determined.

Studies using target trial emulation should incorporate methods for addressing immortal time bias in cases in which treatment strategies cannot be defined at baseline (e.g., studies of treatment duration or use of combination therapies).[5] Researchers should report meaningful sensitivity analyses investigating results' robustness to key sources of bias, including analyses quantifying the potential effect of unobserved confounding and exploring variability in results with alternative specifications of key design elements. Use of negative control outcomes — outcomes strongly believed to have no association with the exposure of interest — may also be useful for quantifying residual bias.

Although observational studies permit investigation of questions

that may be infeasible to study using RCTs and leverage the strengths of RWD, they have many potential sources of bias. The target trial emulation framework attempts to address some of these biases, but emulations must be undertaken and reported with care. Because of the risk of bias due to confounding, sensitivity analyses evaluating robustness of results to unobserved confounding are essential, and variability of results across alternative assumptions about confounding must be considered when interpreting results. When rigorously implemented, the target trial emulation framework provides a useful method for systematic specification of the design of observational studies, but is not a panacea.

Disclosure forms provided by the authors are available at NEJM.org.

From the Department of Biostatistics, School of Public Health, Brown University, Providence, RI (R.A.H., C.A.G., J.W.H.); the Departments of Epidemiology (D.J.H.) and Biostatistics (S.-L.T.N.), Harvard T.H. Chan School of Public Health, and the Department of Health Care Policy, Harvard Medical School (S.-L.T.N.) — both in Boston; the Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom (D.J.H.); and the Department of Population Health, Grossman School of Medicine, New York University, New York (A.B.T.).

This article was published on November 23, 2024, at NEJM.org.

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758-64.
2. Hernán MA. Methods of public health research — strengthening causal inference from observational data. N Engl J Med 2021;385:1345-8.
3. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative effectiveness of BNT162b2 and mRNA-1273 vaccines in U.S. veterans. N Engl J Med 2022;386:105-15.
4. Shortreed SM, Cook AJ, Coley RY, Bobb JF, Nelson JC. Challenges and opportunities for using big health care data to advance medical science and public health. Am J Epidemiol 2019;188:851-61.
5. Gaber CE, Hanson KA, Kim S, Lund JL, Lee TA, Murray EJ. The clone–censor–weight method in pharmacoepidemiologic research: foundations and methodological implementation. Curr Epidemiol Rep 2024;11:164-74.

DOI: 10.1056/NEJMp2407586

A Good Day

Kieran L. Quinn, M.D., Ph.D., and Allan S. Detsky, M.D., Ph.D.

"If we could help you with only one thing today, what would you want it to be?" we asked Mr. D. one morning 10 years ago as he lay in his hospital bed. We were following our standard routine: after reviewing the overnight admissions with the on-call residents and setting the daily objectives, we moved to the bedsides of our patients to talk with and examine them. The goal was to get to know them better. Mr. D. was a frail man in his 80s who had been readmitted for heart failure with severe shortness of breath. He had a very long problem list that included physical symptoms such as weakness and fatigue, medical diagnoses such as renal failure and peripheral vascular disease, a precarious living situation, and polypharmacy. Even the attending physician who had many years of experience caring for people with similarly complex conditions was overwhelmed, not knowing where to begin. So as the encounter was winding down, he attempted to achieve one win for the day, one way to make Mr. D. better off. We expected a response about one of his many clinical problems, perhaps something like "Please make my breathing better." Instead, Mr. D. answered, "Could you find my teeth? They lost them in the emergency department."

Medical care has become much more complex in the past 20 years, but its goals remain the same. People come to clinicians with problems related to their health, and we provide diagnoses, treatments, and advice; we want to help make them feel better. Much of the time, patients can tell us what's bothering them, but an asymmetry of information leaves them at the mercy of the medical team to decide what to do about it. When a patient has multiple complex problems, such decisions can be overwhelming for physicians. Sometimes we don't even know where to begin: Which of the many problems should be tackled first?

In recent years, the concepts of shared decision making and patient centeredness have become popular, yet our ability to translate these ideals into practice remains elusive. Often, their implementation amounts to little more than an illusion — mere hand-waving at the bedside of an overwhelmed patient.[1] Although true shared decision making and patient centeredness involve transmitting knowledge, being careful to avoid bias, and communicating in language that people can understand, the key step in this