

FUNDAMENTALS OF PUBLIC HEALTH

Methods of Public Health Research — Strengthening Causal Inference from Observational Data

Miguel A. Hernán, M.D., Dr.P.H.

Choosing wisely among possible courses of action requires knowledge about the effects of those actions. Public health and medical decision makers therefore need sound causal inferences

to know what works and what harms people. Decision makers prefer inferences based on randomized trials because random assignment of treatment strategies is expected to result in comparable treatment groups, permitting outcome differences to be attributed to the treatment rather than to preexisting differences between groups.

Indeed, for each question about causality, the target of inference can be specified as a randomized trial that would answer the question — the “target trial.” But we cannot conduct enough target trials to answer all causal questions about all treatment strategies and all outcomes in all pop-

ulation groups, and trials may take years to complete.

If an appropriate target trial does not exist when a decision must be made, causal inference may need to rely on observational population data (e.g., electronic medical records generated during routine medical care). Because causal inference from observational data can be viewed as an attempt to emulate a target trial, the question is not whether observational data should be used for causal inference, but rather how to use them most effectively.

Causal inference involves specifying a causal question and answering it. The question is articulated in the form of the target-trial

protocol, which incorporates eligibility criteria, treatment strategies, treatment assignment, start and end of follow-up, outcomes, causal contrasts (or estimands), and a data-analysis plan. These elements define the causal question and how it will be answered; then the target trial is conducted according to the protocol.

For researchers using observational data, a useful way to specify a question is to design the target trial that would answer it and then emulate the protocol as closely as possible. Emulating randomization generally requires data on prognostic factors that are associated with treatment decisions. If all such confounders were correctly measured and adjusted for, there would be no difference between a randomized trial and an observational analysis emulating it. The idea of target-trial emulation, or its logical

Outline of a Target-Trial Protocol: Specification and Emulation Using Observational Data.			
Protocol Component	Description	Example: Antiretroviral Therapy Initiation in HIV-Positive Persons ¹	
		Specification	Emulation Using Observational HIV Cohorts
Eligibility criteria	Who will be included in the study?	HIV-positive persons ≥ 18 yr of age with no prior use of antiretroviral therapy and no history of AIDS	Same as for specification <i>Required data for each person: age, history of therapy use, history of AIDS diagnosis</i>
Treatment strategies	What interventions will eligible persons receive?	Initiation of antiretroviral therapy: 1. Immediately 2. When CD4 cell count drops below 500 cells per cubic millimeter	Same as for specification <i>Required data: date of therapy initiation, clinical measurements of CD4 cell count</i>
Treatment assignment	How will eligible persons be assigned to the interventions?	Eligible persons will be randomly assigned to one strategy and will be aware of which strategy they were assigned to.	Eligible persons will be assigned to the strategies with which their data were compatible at the time of eligibility.
Outcomes	What outcomes in eligible persons will be compared among intervention groups?	Death	Same as for specification <i>Required data: date of death during the study</i>
Follow-up	During which period will eligible persons be followed in the study?	From treatment assignment until death, loss to follow-up, or administrative end of follow-up, whichever occurs first	Same as for specification <i>Required data: date of loss to follow-up</i>
Causal estimand	Which counterfactual contrasts will be estimated using the above data?	Intention-to-treat effect (effect of being assigned to treatment) Per-protocol effect (effect of receiving treatment as indicated in the protocol)	Observational analogue of the per-protocol effect
Statistical analysis	How will the counterfactual contrasts be estimated?	Intention-to-treat analysis Per-protocol analysis (requires adjustment for preassignment and post-assignment confounders)	Same as per-protocol analysis <i>Required data: preassignment and postassignment confounders</i>

equivalents, was central to causal inference methods developed during the 20th century¹; James Robins² generalized it to encompass treatment strategies that are sustained over time.

Target trials emulated using observational data are necessarily pragmatic trials lacking a placebo, blind treatment assignment, and blind outcome ascertainment — design features that don't occur in the real world. Therefore, observational data are not a good fit for causal questions that cannot be expressed in terms of a pragmatic trial. The table outlines the elements of the protocol of a target trial and its observational emulation. The principles

of target-trial emulation are applicable to any causal question that can be translated into a contrast between sufficiently well-defined interventions.

One example of causal inference from observational data can be found in the story of the HIV-treatment-as-prevention strategy. In 2010, millions of people worldwide had HIV infection. With no vaccine in sight, a possible strategy for reducing transmission was to immediately treat anyone with a positive HIV test (because antiretroviral therapy reduces viral concentration). U.S. guidelines, however, recommended starting therapy in asymptomatic people at CD4 cell counts of 350 per cu-

bic millimeter. Earlier initiation at 500 cells per cubic millimeter or higher was not recommended because of concerns about drug toxicity and accumulating resistance. A decision to change guidelines required evidence about the effectiveness and safety of early therapy initiation, but no randomized trials had generated that evidence.

In 2011, using observational data, U.S. clinical guidelines began recommending that antiretroviral therapy be initiated at 500 cells per cubic millimeter. Earlier, data from an international consortium of observational HIV cohorts had been used to emulate a target trial of various

strategies for initiating antiretroviral therapy.³ The analyses indicated that early therapy initiation resulted in the lowest risk of AIDS and death, a finding later validated by a randomized trial of early versus deferred treatment initiation.⁴

This example illustrates the complementarity of randomized trials and observational analyses for causal inference: the effect estimates from observational data were used for provisional decision making until estimates from randomized trials became available, and randomized trial estimates were used as a benchmark for the observational analyses. After an emulation produces results close to the benchmark, there is greater confidence in expanded observational analyses that answer causal questions not considered by the randomized trial. In this case, the observational cohorts were also used to emulate target trials with outcomes (death, drug resistance) and in subgroups (people over 50) that could not be studied with precision in the randomized trial. This interplay between study types may fail in the absence of adequate target-trial emulation: observational analyses that did not specify a target trial led to implausibly high estimates of the benefit of early antiretroviral therapy initiation.⁵

Traditional analyses of observational data are often based on allocation of person-time to “exposed” and “unexposed” groups, rather than on explicit specification and emulation of a target trial. Resulting estimates may not correspond to a relevant causal contrast and therefore may not be easily mappable onto real-world interventions.⁵ This lack of

actionable causal inference arises frequently in traditional analyses of both medical and nonmedical factors. An important reason to explicitly specify and emulate a target trial is to compare well-defined courses of action, which helps decision makers.

Another reason for emulating a target trial is to avert the selection and immortal time biases that arise from mishandling the start of follow-up (time zero) in analyses. As a rule, in causal analyses of both randomized trials and observational data, each participant’s time zero must be

Causal effect estimates from observational analyses are often distrusted because of the lack of randomization, which may lead to confounded estimates due to noncomparable treatment groups. Confounding is a serious concern, but many high-profile observational failures have resulted instead from mishandling of time zero. In fact, reanalyses of observational data that explicitly emulated a target trial (and thus handled time zero correctly) yielded estimates compatible with those from randomized trials in the examples above. That is, the

Explicit target-trial emulation increases the transparency and replicability of observational effect estimates.

the time when they meet the eligibility criteria and are assigned to a treatment strategy. This rule is automatically enforced in randomized trials and in observational analyses emulating target trials. Observational analyses deviating from this rule led to conclusions, which were later disproved, that statins reduce cancer risk and that estrogen-plus-progestin therapy reduces the risk of coronary heart disease in postmenopausal women. Though sometimes appropriate, deviations from this rule must be justified on a case-by-case basis; for example, given the long period between exposure and outcome, the effect of cigarette smoking on lung cancer may be approximately quantified even when people are not followed from the time they begin smoking.

observational data were sufficient to approximately emulate randomization (i.e., to adjust for confounding); the failures were caused by selection and immortal time biases that can be avoided by explicitly emulating a target trial. Alternatively, these biases can be avoided by studious application of principles of causal inference and study design, but the target-trial approach helps in implementing these principles.

By itself, however, target-trial emulation cannot overcome confounding bias from noncomparable treatment groups. Despite correctly emulating all other components of a target trial, observational analyses may be invalidated if confounding cannot be adequately adjusted for. Sophisticated adjustment methods are sometimes necessary, but they

work only when good data on confounders are available. Machine learning and artificial intelligence methods cannot compensate for missing data.

When confounder data are not available in an observational database, certain causal questions cannot feasibly be answered. For example, insurance claims databases may be inadequate for estimating effects of preventive interventions on all-cause mortality. It's important to build safeguards (such as negative controls) into observational analyses to alert investigators when the danger of confounding is too high. Confounding adjustment is also infeasible for causal questions involving interventions (e.g., antihypertensive therapy) that are used almost exclusively by people with risk factors for the outcome (e.g., cardiovascular disease).

When data on confounders are insufficient, researchers can sometimes use methods such as instrumental variable estimation, which replace the need to adjust for confounders with other

strong untestable conditions. For causal questions about the effects of an

intervention (such as a policy change or new program) newly implemented in a population, the requirement for data on confounders for each person can,

under certain conditions, be replaced by comparisons between preintervention and postintervention periods. For complicated causal questions involving interactions between persons in the population or systemwide effects, observational data sets cannot by themselves be used to emulate hypothetical target trials. For example, attempts to quantify individual and societal effects of interventions for controlling the U.S. opioid epidemic require simulation models integrating observational and experimental findings with assumptions about the structure of society and the health system.⁵

Explicit target-trial emulation increases the transparency and replicability of observational effect estimates. By including descriptions of target-trial protocols and their emulations in reports of observational analyses, investigators tell us precisely which causal effects they are estimating so that our replication attempts can be accurate. Also, explicit specification of the target trial imposes constraints on data analysis, reducing multiple comparisons and selective reporting of results. And it prevents data manipulations resulting in hard-to-interpret estimates that don't correspond to any relevant intervention.

Determining the effectiveness

and safety of many health interventions will continue to rely on observational data because randomized trials are not always feasible, ethical, or timely. Explicit emulation of a target trial using observational data helps eliminate unnecessary sources of bias so that concerns can focus on potential confounding bias due to nonrandomization.

Disclosure forms provided by the author are available at [NEJM.org](https://www.nejm.org).


From the CAUSALab and the Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston.

This article was published on October 2, 2021, at [NEJM.org](https://www.nejm.org).

1. Dorn HF. Philosophy of inferences for retrospective studies. *Am J Public Health* 1953;43:677-83.
2. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period — application to the healthy worker survivor effect. *Math Model* 1986;7:1392-512.
3. Cain LE, Logan R, Robins JM, et al. When to initiate combined antiretroviral therapy to reduce mortality and AIDS-defining illness in HIV-infected persons in developed countries: an observational study. *Ann Intern Med* 2011;154:509-15.
4. Lodi S, Phillips A, Lundgren J, et al. Effect estimates in randomized trials and observational studies: comparing apples with apples. *Am J Epidemiol* 2019;188:1569-77.
5. Kitahata MM, Gange SJ, Abraham AG, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. *N Engl J Med* 2009;360:1815-26.

DOI: 10.1056/NEJMp2113319

Copyright © 2021 Massachusetts Medical Society.

 An audio interview with Dr. Hernán is available at [NEJM.org](https://www.nejm.org)

The No Surprises Act and Informed Financial Consent

Barak Richman, J.D., Ph.D., Mark Hall, J.D., and Kevin Schulman, M.D.

At the close of 2020, the U.S. Congress passed the No Surprises Act to end the pernicious

practice of so-called surprise billing for out-of-network care.¹ But the country's recent experience

with balance-billing prohibitions for Covid-19 testing and treatment offers an ominous warn-