Target Trial Emulation

A Primer on Improving Observational Research in Neurology

Samuel W. Terman,[1] Jaime L. Speiser,[2] Misha Eliasziw,[3] Tobias Kurth,[4] and Andrea Lauren Christman Schneider[5]

Correspondence
Dr. Terman
sterman@umich.edu

Neurology® 2025;105:e214269. doi:10.1212/WNL.0000000000214269

Abstract

Randomized controlled trials (RCTs) are often considered to be the gold standard for determining treatment effects, but they are often infeasible because of ethical, logistical, or financial constraints. Even when RCTs are possible, they take many years to complete and may not reflect the characteristics or care settings of routine clinical populations. High-quality observational data analyzed rigorously can thus fill important evidence gaps when RCTs are not available. However, analyses of nonrandomized data are susceptible to biases that could have been avoided in a well-designed trial. This review discusses the "target trial" framework to bridge this gap. Target trial emulation involves specifying the protocol for the ideal RCT (i.e., eligibility, interventions, treatment assignment, follow-up, outcomes, and analysis) and then emulating each component using observational data. We describe the concept of target trial emulation, in addition to an overview of the planning and execution of such studies. Although target trial emulation does not overcome data set limitations (e.g., measurement error or residual confounding), it does improve traditional observational analyses in numerous important aspects of study design, such as more precisely defining the research question and avoiding biases related to aligning the start of follow-up with eligibility and treatment assignment. We use examples to familiarize readers with how this methodology can be applied to neurologic conditions.

Introduction

Randomized controlled trials (RCTs) are foundational to evidence-based medicine. However, they are costly and time-consuming, involve intense regulatory oversight, and are not always ethically permissible. Moreover, enrollment criteria are often overly strict,[1] sometimes including only the most or least severely affected or most adherent patients[2] or underrepresenting key groups.[3] Trials test only limited comparisons during intensive scheduled monitoring, and patients are not blinded during routine care.

These challenges motivate using observational data when the ideal RCT is not feasible. Observational data may allow lower costs, faster data collection, longer follow-up, patient monitoring more typical of routine clinical care, a wider range of interventions, and larger sample sizes. Fortunately, extensive nonrandomized data exist (e.g., electronic medical records and claims databases). This is particularly important in neurologic care because recent years have witnessed a rapid increase in the number of new treatments.

Despite these advantages, observational analyses are prone to many biases. While some biases may be inherent to the data set (e.g., unmeasured confounding), others (e.g., immortal time bias)[4,5] can be avoided at the design phase with an appropriately structured framework.

In this study, we introduce the neurologic community to the target trial framework. This framework involves specifying the ideal RCT protocol (i.e., the target trial) and then systematically emulating each element using observational data when the ideal RCT is not available.[6,7] It is applicable anytime an investigator wishes to use observational data to make inferences about the

[1]Department of Neurology, University of Michigan, Ann Arbor; [2]Department of Biostatistics and Data Science, Wake Forest University School of Medicine, Winston-Salem, NC; [3]Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA; [4]Institute of Public Health, Charité-Universitätsmedizin Berlin, Germany; and [5]Department of Neurology, University of Pennsylvania, Philadelphia.

Glossary

ASM = antiseizure medication; **DAG** = directed acyclic graph; **ICH** = intracranial hemorrhage; **ITT** = intention-to-treat; **PP** = per-protocol; **RCT** = randomized controlled trial.

effects of a treatment. The target trial approach involves familiar epidemiologic concepts such as confounding and selection bias. However, conducting observational analyses (e.g., cohort studies) in this way adds important layers of rigor.

From Protocol to Emulation: Applying the Target Trial Framework

The first step in applying the target trial framework is to develop a detailed protocol for the hypothetical randomized trial that would ideally answer the causal question of interest. This protocol includes clearly defined eligibility criteria, treatment strategies, confounders that may differ between observational groups, the timing of treatment assignment and outcome measurement and start of follow-up (i.e., time zero), and the estimand of interest (e.g., the average treatment effect). These estimands are typically approached through specific analysis strategies, that is, intention-to-treat (ITT) or per-protocol (PP) analyses.

The Figure provides a schematic overview. The Table summarizes relevant definitions.

To demonstrate how this framework can be applied in practice, we highlight 3 recent studies published in *Neurology*® that use these methods:

1. Statins and dementia[9]: What is the effect of initiating statins vs not initiating statins on the risk of developing dementia among individuals aged 55 to 80 who had not used statins in the previous 2 years and had no baseline diagnosis of dementia?
2. Statins and intracranial hemorrhage (ICH)[10]: What is the effect of initiating statins vs not initiating statins on the risk of ICH among adults older than age 50 in China?
3. Lamotrigine and ventricular arrhythmias[11]: What is the effect of initiating lamotrigine vs initiating levetiracetam on the risk of ventricular arrhythmias among adults enrolled in Medicare with newly treated epilepsy?

Treatment and Analysis Strategies (Asking a Well-Defined Question)

RCTs assign participants to specific treatment strategies.[12] For example, "take 100 mg of Drug X vs placebo once daily for 2 years" could represent a comparison between 2 well-defined treatment strategies. Although subjective, the key is to specify

Figure Conceptual Diagram for the Target Trial Framework

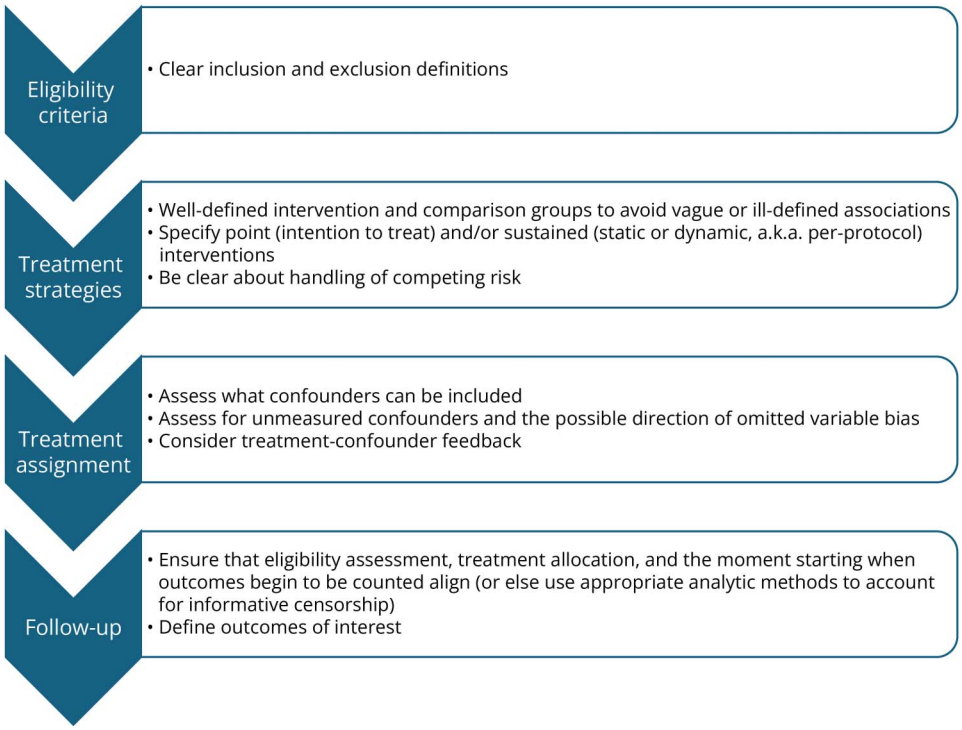

Table Definitions

Term	Definition
Cloning	A statistical technique involving duplicating observations in the data set. One use is when comparing initially indistinguishable interventions; thus, one can "assign" one treatment strategy to each "clone" and censor and weight each clone separately on deviation from their "assigned" strategy
Competing risk	Some outcomes (e.g., death from unrelated causes) may render the outcome of interest (e.g., dementia) impossible (or "compete" with the outcome from occurring)
Conditional exchangeability	An assumption that each treatment group has a similar average outcome risk considering all factors other than the treatment that they actually received. The assumption is that the treated group, had they alternatively been untreated counter to fact, would have had the same average outcome as the group that was actually untreated, and vice versa
Confounder	A common cause of both the exposure and the outcome but is itself not influenced by the exposure (otherwise, it would be a mediator)
Confounding	Observing a distorted causal relationship between the exposure and outcome when a set of variables exist that influence both the exposure and the outcome
Dynamic treatment strategy	This is a sustained intervention in which patients start an initial course of action but may deviate from that initial course of action under certain circumstances. An example could be, "Do not initiate statin therapy *unless* one subsequently develops heart disease"
Immortal time/prevalent user bias[8]	This refers to the result of design flaws leading to certain individuals who cannot possibly develop the outcome of interest for some time. This can be due to either misclassification (i.e., treatment is "assigned" after the start of follow-up) or selection bias (i.e., patients are "selected" after when follow-up should have started)
Informative censorship	Occurs when follow-up duration is related to a participant's chance of developing the outcome of interest. For example, if participants at the highest risk of the outcome dropout soonest, this could underestimate outcome occurrence
Intention-to-treat	The effect of the "coin flip"; in other words, the effect of being assigned to different courses of action, regardless of the degree to which patients subsequently adhere to their assigned course of action
Per-protocol	The effect of adhering to the initially assigned course of action
Point treatment strategy	This refers to the effect of being assigned to a given course of action, without regard to subsequent adherence, akin to intention-to-treat effects and in contrast to sustained interventions
Residual confounding	When the analysis has not adequately addressed confounding due to omitting or being unable to include sufficient variables to eliminate confounding
Selection bias	Observing a distorted causal relationship between the exposure and outcome when the outcome influences the probability of selection and the exposure also influences the probability of selection. Or else, when there are common upstream factors that influence both the outcome and selection and other upstream factors that influence both the exposure and selection
Sustained treatment strategy	This includes static or dynamic treatment strategies, in which adherence to a certain course of action after the "coin flip" is required; akin to per-protocol effects and in contrast to point interventions
Static treatment strategy	This is a sustained intervention in which patients start an initial course of action and are expected to continue it no matter what, unlike a dynamic treatment strategy

enough detail to the degree that experts believe likely matters. Some unimportant details will remain intentionally unspecified, for example, how many ounces of water are used to swallow the pill. Conversely, some details could matter yet may not be captured by the data, such as the degree of daily pill adherence. Say a less well-defined question could be more simply, "Is Drug X associated with Outcome Y?" Without specifying important details about the dose, frequency, comparator, duration, etc., study results may be difficult to interpret, in contrast to the ideal RCT's intervention.

RCTs also make the analysis strategy clear, such as the ITT (difference between randomized groups) vs the PP (difference between those who adhered to and did not adhere to their assigned protocol) effect. Observational analyses should be no different, yet in our experience, rarely is such detail sufficiently specified under typical observational analyses outside the target trial framework.

Two major types of interventions include the following:

1. A "point" intervention, for example, "What if all patients were assigned to start intervention A vs intervention B" is akin to an ITT question. This is the effect of the 1-time "coin flip." Thus, patients are analyzed according to their initial treatment regardless of subsequent nonadherence. This mimics the decision under the clinician's control—whether to recommend alternative treatments, knowing that nonadherence is possible.
2. A "sustained" intervention, for example, "What if all patients started *and continued* intervention A vs intervention B" is akin to a PP question. Unlike ITT analyses, this approach assumes that all patients start and indefinitely adhere to their initially assigned intervention.

Sustained interventions can be further divided into the following:

2a. "Static" interventions, for example, "What if all patients *always* took treatment A vs *never* took treatment A, no matter what" Or, "What if all patients *always* took treatment A vs *always* took treatment B, no matter what?"
2b. "Dynamic" interventions, for example, "What if all patients started and continued treatment A vs B *unless* certain conditions are met (e.g., intolerance or new contraindications of initially treated patients, or new indications for initially untreated patients)." This is often a more realistic contrast than static interventions because it allows patients to deviate from their initial treatment when valid reasons arise but requires that the data set contain reasons for deviation.

Being clear about which of these types of interventions are being compared is critical for interpretability. This provides one more reason why the question "Is Drug X associated with Outcome Y" remains inadequately specified, as results may differ depending on whether one aims to compare point vs sustained strategies. For example, in the statins and ICH article,[10] ITT analyses comparing time to first ICH in statin initiators vs noninitiators yielded a hazard ratio of 1.18 (95% CI 1.03–1.35). However, PP analyses yielded a stronger hazard ratio of 1.61 (95% CI 1.26–2.07). The stronger effect in the PP analyses was hypothesized to reflect a scenario in which all patients adhered to their initially assigned treatment, whereas the ITT effect is typically attenuated to the degree that patients switch treatments after initial assignment.

The statins and dementia study[9] illustrates an explicit comparison across sustained intervention strategies. Treatment strategies consisted of the following: (1) initiate and continue statin therapy (static strategy) vs (2) do not initiate statin therapy *unless* one subsequently develops heart disease (dynamic strategy). Note that in the static comparison, follow-up is censored when a baseline noninitiator starts statins or vice versa, because the treatment interventions of interest were "What if all patients started and continued a statin vs never started a statin?" By contrast, in the dynamic comparison, even if a baseline noninitiator started a statin after the follow-up period began, they would still be followed, provided that they initiated statins after developing a true indication (e.g., a myocardial infarction). However, had the comparison been framed as an ITT question, no such censorship would be necessary.

The key insight is that interpretability is enhanced when treatment strategies are more precisely specified than may otherwise be the case in observational analyses without the target trial framework. Patients are analyzed according to the treatment *strategy* of interest and are typically censored if/when they deviate from sustained treatment strategies.

Another important consideration when specifying a well-defined comparison is "competing risk."[13] Some outcomes (e.g., death from unrelated causes) may render the outcome of interest (e.g., dementia) impossible. The patient might have developed dementia had they not prematurely died from another unrelated cause, but we will never know whether or when that would have occurred. There is no consensus on the best way to handle this issue. If the investigator censors patients at death like any other cause of loss to follow-up (the "controlled direct effect"[13]), this acts as though it remains unknown whether the patient could ever develop dementia after death. This translates into an ITT question of "What is the effect on dementia of being assigned to statin vs no statin, assuming that patients were never lost to follow-up *and never died*?" While that question seems nonsensical, the alternatives also have drawbacks. The Fine-Gray approach (or "total effect"[13]) instead acknowledges that outcome risk goes to zero after death.[14] This analysis more realistically assumes only that patients "were never lost to follow-up" and acknowledges that the risk of the outcome is zero after death.[15] However, more lethal treatments will appear "protective" against non-death outcomes because death would preclude all other outcomes. A third approach was to combine death and the outcome of interest as a composite (e.g., "death or dementia"). While avoiding the abovementioned issues, this changes the causal question, no longer distinguishing between death and the outcome of interest. Fortunately, in our example articles, the handling of competing risk did not change conclusions and all illustrate the common practice of conducting competing risk analyses using several methods. Still, we raise this issue because choices have clear implications regarding interpretation, none inherently superior to others. Investigators often censor at death and then conduct a sensitivity analysis using the Fine-Gray approach, but it is important to recognize the strengths and limitations of each approach and frame results in terms of which type of question is of greater interest.

Finally, experts in target trial emulation tend to view adjusted survival curves as more informative than hazard ratios regarding a treatment's effect over time.[16] (1) Survival curves provide absolute, not relative, estimates. (2) Cox models require the proportional hazard assumption, not required with logistic regression approximations. (3) Hazard ratios inherently favor the "worse" group as study durations increase because they are calculated by conditioning on event-free individuals. Susceptible individuals accrue events most rapidly in the "worse" treatment, leaving behind only the least susceptible individuals over time.

Eligibility Criteria (Generalizability)

As with trials, observational studies of treatment effects should have clear inclusion and exclusion criteria. In the statins and dementia example,[9] the population was constructed identically to how one might envision criteria in the target trial. For example, if the outcome under investigation is not repeatable and the question pertains to development of incident disease, then it makes sense to exclude patients with prevalent disease. Thus, the investigators excluded patients with baseline

dementia to study incident dementia. In addition, because statins could be considered for primary prevention, which is particularly relevant to older adults, they included patients older than 50 years.

Note that applying eligibility criteria or accounting for nonparticipation (selection) does not necessarily imply selection bias.[17] Eligibility criteria affect generalizability of results to other populations. By contrast, selection bias affects accuracy of results within this population. It is introduced only to the extent that both (A) the outcome influences the probability of selection (or else common upstream factors influence both the outcome and selection) and (B) the exposure influences the probability of selection (or else common upstream factors influence both the exposure and selection).

Treatment Assignment (Confounding)

In randomized data, patients are assigned treatments independent of their outcome risk. However, during clinical care, clinicians prescribe treatments with knowledge of a patient's outcome risk. Hence, 1 group may have more favorable risk factors independent of assigned treatment. In observational analyses, we seek to account for such factors. This is best approached using directed acyclic graphs (DAGs),[18-20] which are visual depictions of how variables are suspected to influence each other, enabling the investigator to defend analytic choices. We encourage investigators to construct a DAG to justify which variables are needed to address confounding and also to identify variables that were unavailable and the potential direction of residual confounding (discussed in the statins and ICH article[10]).

Let us turn to the third example study regarding the influence of lamotrigine vs levetiracetam on ventricular arrhythmias.[11] What might the relevant confounders be? History of status epilepticus could favor levetiracetam because it can be started quickly intravenously, whereas mood dysfunction could favor lamotrigine. Both status epilepticus and mood dysfunction could also promote arrhythmias. Therefore, both could be confounders. For example, if status epilepticus favors levetiracetam and increases arrhythmias, then failing to account for status epilepticus could spuriously worsen outcomes in the levetiracetam group.

In the lamotrigine and arrhythmias example,[11] one could have compared "start lamotrigine vs start nothing." However, the researchers were concerned that residual confounding was too great to proceed, because the data set lacked potential confounders such as the number of convulsions or EEG findings (which could both promote treatment and influence arrhythmias). If frequent unprovoked convulsions promoted both lamotrigine initiation and arrhythmias, but convulsion frequency was not captured in the data set, then lamotrigine would appear spuriously harmful. Therefore, this comparison was abandoned. Rather, the chosen comparison was, "among patients with a treatment indication (to improve conditional exchangeability), start lamotrigine vs start levetiracetam. Levetiracetam is the most common antiseizure medication (ASM).[21] Moreover, because the study period was before the warning regarding arrhythmias was published and in absence of known enzymatic induction, it seemed unlikely that clinicians were choosing lamotrigine vs levetiracetam candidates based on cardiovascular risk profiles. Residual confounding seemed to be much less in the comparison "start lamotrigine vs start levetiracetam" than in the comparison "start lamotrigine vs start nothing." Moreover, the former question was felt to be more clinically applicable, given that the comparison of greatest interest seems to be *which* ASM to start, rather than *whether* to start an ASM.

Another consideration is treatment-confounder feedback. This occurs when an intermediate treatment response (e.g., ASM side effects) influences subsequent treatment (hence, "feedback"), and unmeasured confounders (e.g., brain atrophy) exist between the intermediate treatment response and the ultimate outcome of interest (seizures). When suspected, standard regression adjustment for these intermediates (factors influenced by previous treatment that are also causes of the outcome) would actually induce rather than mitigate bias. Instead, time-varying confounding with treatment-confounder feedback necessitates "g-methods" (e.g., inverse probability of treatment weighting) rather than regression adjustment for valid causal estimates.[22,23]

As implied, numerous methods exist to address confounding. For example, stratification is suited when only a limited number of categorical variables are sufficient to address confounding. However, stratification cannot handle continuous variables or higher order terms and may be less efficient because it divides the data. Alternatively, regression adjustment can be used when the investigator believes that they have measured a sufficient set of variables related to treatment that may also predict the outcome and they are willing to accept linear assumptions between predictors and the outcome. If the investigator is less confident regarding the functional form of predictor-outcome relationships, one can alternatively model the chance of receiving treatment. This can be performed by propensity score[24] matching to compare patients with similar probabilities of treatment, or else by inverse probability of treatment weighing to simulate pseudopopulations as if all patients received each treatment under consideration.[25] Such treatment modeling approaches seek to balance risk factors across treatment groups, as would have been achieved with randomization. Further doubly robust techniques combine modeling the outcome risk and modeling the chance of receiving treatment. This approach provides 2 opportunities to correctly model confounders,[26] although at the expense of potentially magnifying bias if both models are incorrectly specified.[27]

Follow-Up (Specifying Time Zero)

In an RCT, both groups are assessed for eligibility, treatment is assigned, and then follow-up begins. Target trial emulation seeks to explicitly reproduce this sequence of events to avoid

"immortal time bias,"[4,5,8,28-30] which could otherwise occur in observational studies when these time points are misaligned, potentially leading to findings of "protective" effects of harmful interventions.[4,29,30] Immortal time bias refers to design flaws where certain individuals cannot possibly develop the outcome of interest for some period by virtue of either selection bias (e.g., prevalent user bias) or exposure misclassification (e.g., improperly analyzing initially indistinguishable interventions), explained further.[8]

Imagine an observational study in which we compared current ("prevalent") lamotrigine users and patients not currently using lamotrigine. While this could initially seem reasonable, it poses an important threat to validity—"prevalent user bias." If lamotrigine increased serious ventricular arrhythmias, then one would enroll only those prevalent users who survived long enough for inclusion despite its toxic effects, thus enrolling only the remaining lamotrigine "super users" with lowest risk. Susceptible treated individuals are prematurely depleted from the risk set before outcomes begin to be counted. This problem occurred because eligibility for this hypothetical study was ascertained and outcome follow-up was started potentially years *after* treatment was started, which would have never occurred in the ideal RCT (which would have ascertained eligibility just *before* treatment assignment and would have begun outcome follow-up at the time of initiating lamotrigine rather than years afterward). This is a problem to the extent that unobserved factors driving the outcome (arrhythmias) also influence the probability of being selected into the study (survival).

This pitfall related to selection bias is worth highlighting because it is entirely avoidable by applying the target trial framework. Imagine randomizing half of the patients to "not start lamotrigine" vs to "have been using lamotrigine for the last 10 years." The former is easily performed by a coin flip. However, the latter is impossible without time travel. This faulty comparison is analogous to "Randomize half of patients to (1) do not start lamotrigine and begin arrhythmia monitoring immediately, vs (2) start lamotrigine but begin arrhythmia monitoring 10 years after randomization only among those patients who survived the first 10 years."[8] This illustrates a fundamental tenet of target trial emulation: treatments being compared in observational data should map to well-defined interventions that could be replicated in the target trial, with a fair, common starting line beginning when patients are assigned to treatment vs no treatment. The ideal trial would not assign treatments or determine eligibility retroactively, based on information available only after follow-up began. Only under advanced settings would this sort of comparison be advisable[31] or else in special situations discussed further with appropriate analytic techniques (e.g., sustained interventions). Thus, in our example, rather than comparing prevalent lamotrigine users vs nonusers, a better comparison was chosen—"start lamotrigine now" vs "start levetiracetam now," starting follow-up when the patient began their first ASM. The discussion of the statins and ICH study[10] also cites numerous past observational publications that suggested a protective effect of statins against ICH, in contrast to the target trial emulation and existing RCTs. They discuss in detail (page 3213489(8)) that discrepant findings could have been due to observational studies not adhering to concepts of target trial emulation (e.g., using hyperlipidemia diagnosis or first stroke as the index date for nonusers but statin initiation date for users and assessing covariates at the time of ICH for cases but the time of sampling for controls)

An important caveat pertains to analyzing sustained interventions that depend on postbaseline adherence. If the investigator wishes to compare point (ITT) strategies. "start lamotrigine" vs "start levetiracetam," then no special censorship is needed—simply compare groups according to their initial treatment and follow them until events, dropout, or the end of available follow-up. However, if comparing sustained (PP) interventions such as "start *and always take* lamotrigine" vs "start *and always take* levetiracetam," patients are typically censored on deviation from that strategy (e.g., when they subsequently stop their initial treatment). A clear problem arises when postbaseline reasons for nonadherence are also related to the outcome of interest (i.e., informative censorship). An RCT would not use information *after* randomization to determine follow-up duration. Therefore, 1 additional step is necessary, accounting for variables that may explain postbaseline nonadherence. A common approach is to conduct a model estimating the probability of receiving treatment across all available data, plus a second model estimating the probability of loss to follow-up over time, and then weight the main model using both of these quantities to handle both separate issues simultaneously.[32,33] This allows the analyst to censor patients when their data are no longer consistent with their initial treatment strategy, while still overcoming the bias introduced by informative censorship. In our lamotrigine example,[11] potential factors that could influence preferential stopping of lamotrigine or levetiracetam and also affect ventricular arrhythmias could include age or existing mood dysfunction; hence, both were included in the analysis. The key issue for the clinician designing research is that collecting baseline (non–time-varying) confounders helps overcome confounding for point (ITT) interventions. However, one also needs to consider postbaseline (potentially time-varying) factors influencing treatment crossover if one wishes to compare sustained (PP) interventions that depend on information after outcome follow-up begins.

In addition, along these lines, say one wished to compare interventions that are not distinguishable at the start of follow-up, such as "Start lamotrigine and titrate up to 100 mg twice daily vs 150 mg twice daily." On reaching 100 mg, it would not yet be clear which path they were following because the data would be consistent with either path. Worse, if lamotrigine did induce arrhythmias, then only the least arrhythmia-prone individuals would survive long enough to ever reach 150 mg, thus representing another way to introduce immortal time bias (only patients who remained event-free survive long

enough to reach the higher treatment; thus, patients assigned to the high-dose treatment could not have possibly experienced outcomes before reaching that dose). One preferred and efficient solution involves making 2 copies of each patient in the data set (i.e., cloning) when it is not initially clear which treatment to "assign" a given patient, given that their initial data could have been consistent with either treatment path. One clone then gets "assigned" to either treatment, and then, each clone is censored on deviation from their "assigned" treatment and weighted separately to address the bias otherwise induced by using postbaseline information for censorship.[34,35] Other approaches such as time-varying treatment indicators address immortal time bias but have less clear interpretation and may not recapitulate estimates from the target trial for reasons discussed elsewhere.[8,16]

In addition, consider that without a single moment of randomization, choosing what moment to start follow-up can sometimes be arbitrary, especially when a comparator group involves patients *not* initiating treatment. Furthermore, few individuals might initiate treatment at any one point in time leading to imprecise estimates. One approach is to choose a specific calendar date to determine eligibility, treatment assignment, and start of follow-up for all patients. Another approach to reduce arbitrariness and improve precision is for investigators to conduct not just a single trial emulation, but rather a *sequence* of emulations. In other words, imagine a new trial starting each week or month of follow-up from the same individuals, which are then all pooled together.[33,36-38] This was the case for both of our statin examples,[9,10] because it was rare to initiate a statin in any given month.

In short, aligning the start of follow-up with the moment of eligibility assessment and treatment assignment is a core concept in target trial emulation. Immortal time bias occurs due to investigator design choices, not data limitations, when either selection (e.g., prevalent user bias) or treatment classification (e.g., initially indistinguishable treatments) occur after outcome follow-up begins. The target trial framework thus encourages investigators to think critically about optimal study design to ask more valid research questions.

Limitations of Target Trial Emulation

Residual confounding is a key concern regarding target trial emulation. The statins and ICH article[10] clearly depicts their hypothesized causal diagram, illustrating which unmeasured variables could distort the relationships of interest despite best efforts. For example, while the investigators of the statins and ICH article noted that, ideally, they would have adjusted for certain hypothesized confounders (e.g., smoking, alcohol, body mass index, and low density lipoprotein levels), variables at hand included only surrogates that could have incompletely addressed confounding (e.g., International Classification of Diseases codes for substance use, obesity, and hypercholesterolemia). If such factors promoted both statin initiation and ICH, results could overestimate the effect of statins on ICH. Unfortunately, "no residual confounding" is an untestable assumption. This can only be argued based on content expertise, potentially supported by "negative control" ("falsification") analyses or quantitative bias assessment that infer how much omitted variable bias would have been necessary to overturn the results. While alternatives exist in select situations to unlink treatment from confounders (e.g., instrumental variables[39] or quasi-experiments), such methods require particular assumptions for validity. Thus, *Neurology* classifies target trial emulations typically as Class III evidence because of the possibility of residual confounding, in absence of true randomization.

Type I (false-positive) and II (false-negative) errors occur in target trial emulation, just as in any other observational or randomized setting. Thus, investigators should limit the number of comparisons to those of greatest scientific interest, and prespecified power calculations can help interpret the probability of missing a clinically meaningful difference.

Studies may incompletely describe their intended target trial's protocol.[40] Guidelines remain in development to enhance reporting.[41]

Even if an intervention is found to be harmful for 1 outcome, clinicians still must weigh the absolute effect sizes of harms against expected benefits.

Missing data represent a particularly important issue in observational data. RCTs measure predefined data at predetermined intervals, whereas observational data are often collected only during clinical encounters, the frequency of which is determined by likely unmeasured factors related to disease severity, provider availability, socioeconomic factors, etc. Multiple imputation can address bias from missing confounders or eligibility criteria,[42,43] and missing eligibility criteria can further be addressed by controlling for collider stratification bias,[44] although such methods require assumptions that missingness is predictable based on nonmissing data.

Differential measurement error can emerge from lack of blinding, or from differential detection due to nonrandomly spaced clinic visits, which may differ between treatment arms (e.g., patients who start a statin might be more likely to seek attention for cognitive concerns than those who do not start a statin).

Even RCTs asking essentially the same question may yield contradictory results due to either chance or systematic differences between populations and interventions.[45,46]

Finally, finding an effect does not clarify the mechanism, which is the purpose of mediation analysis.[47]

Fortunately, concordance does tend to be high between randomized and observational results when the target trial is emulated closely.[48]

Conclusions

Although target trial emulation involves familiar steps from observational analysis (e.g., defining eligibility and adjusting for confounding), it is distinguished by its protocol-based structure. Explicitly mapping each element of a hypothetical randomized trial onto observational data imposes a systematic design that improves interpretability.

Although it cannot resolve inherent data limitations such as residual confounding, the framework helps prevent common biases, especially those related to misaligned timing of eligibility, treatment, and outcome, and leads to better-defined research questions.

As available treatments advance within neurologic conditions, the potential applications of these techniques will likely continue to expand.

Many excellent reviews[5-7,49,50] exist for further reading.

Author Contributions

S.W. Terman: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; analysis or interpretation of data. J.L. Speiser: drafting/revision of the manuscript for content, including medical writing for content. M. Eliasziw: drafting/revision of the manuscript for content, including medical writing for content. T. Kurth: drafting/revision of the manuscript for content, including medical writing for content. A.L.C. Schneider: drafting/revision of the manuscript for content, including medical writing for content.

Study Funding

The authors report no targeted funding.

Disclosure

S.W. Terman is a member of the *Neurology*® Methods and Statistics editorial board, is supported by National Institute on Aging of the NIH under award number K23 AG081463, and has been a paid consultant for Xenon Pharmaceutical and Jazz Pharmaceutical for unrelated activities. J.L. Speiser is a member of the *Neurology*® Methods and Statistics editorial board and is supported by the National Institute on Aging of the NIH under Award Number K25AG068253. M. Eliasziw and T. Kurth are members of the *Neurology*® Methods and Statistics editorial board and report no other disclosures relevant to the manuscript. A.L.C. Schneider is an associate editor for the *Neurology*® Methods and Statistics editorial board and is supported by the National Institute of Neurological Disorders and Stroke of the NIH under award number K23 NS123340. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology*® June 24, 2025. Accepted in final form August 11, 2025. Submitted and externally peer reviewed. The handling editor was Editor-in-Chief José Merino, MD, MPhil, FAAN.

References

1. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA*. 2007;297(11):1233-1240. doi:10.1001/jama.297.11.1233
2. Goldenholz DM, Goldenholz EB, Kaptchuk TJ. Quantifying and controlling the impact of regression to the mean on randomized controlled trials in epilepsy. *Epilepsia*. 2023;64(10):2635-2643. doi:10.1111/epi.17730
3. Kong WY, Saber H, Marawar R, Basha MM. Racial and ethnic trends in antiseizure medications trial enrolment: a systematic review using ClinicalTrials.gov. *Epilepsy Res*. 2021;173:106613. doi:10.1016/j.eplepsyres.2021.106613
4. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Avoidable flaws in observational analyses: an application to statins and cancer. *Nat Med*. 2019;25(10):1601-1606. doi:10.1038/s41591-019-0597-x
5. 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
6. 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
7. Matthews AA, Young JC, Kurth T. The target trial framework in clinical epidemiology: principles and applications. *J Clin Epidemiol*. 2023;164:112-115. doi:10.1016/j.jclinepi.2023.10.008
8. Hernán MA, Sterne JAC, Higgins JPT, Shrier I, Hernández-Díaz S. A structural description of biases that generate immortal time. *Epidemiology*. 2025;36(1):107-114. doi:10.1097/EDE.0000000000001808
9. Caniglia EC, Rojas-Saunero LP, Hilal S, et al. Emulating a target trial of statin use and risk of dementia using cohort data. *Neurology*. 2020;95(10):e1322-e1332. doi:10.1212/WNL.0000000000010433
10. Ji D, Dong S, Wang T, et al. Statin use and risk of intracerebral hemorrhage in Chinese population: a target trial emulation study. *Neurology*. 2025;104(8):e213489. doi:10.1212/WNL.0000000000213489
11. Terman SW, Josephson CB, Goyal P, et al. Lamotrigine and cardiac arrhythmias: a target trial approach. *Neurology*. 2025;105(1):e213640. doi:10.1212/WNL.0000000000213640
12. Hernán MA. Does water kill? A call for less casual causal inferences. *Ann Epidemiol*. 2016;26(10):674-680. doi:10.1016/j.annepidem.2016.08.016
13. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure time settings with competing events. *Stat Med*. 2020;39(8):1199-1236. doi:10.1002/sim.8471
14. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc*. 1999;94(446):496-509. doi:10.1080/01621459.1999.10474144
15. Allison P. *For Causal Analysis of Competing Risks, Don't Use Fine & Gray's Subdistribution Method [online]*. Statistical Horizons; 2018. Accessed November 4, 2021. statisticalhorizons.com/for-causal-analysis-of-competing-risks
16. Hernán MA. The hazards of hazard ratios. *Epidemiology*. 2010;21(1):13-15. doi:10.1097/ede.0b013e3181c1ea43
17. Holmberg MJ, Andersen LW. Collider bias. *JAMA*. 2022;327(13):1282-1284. doi:10.1001/jama.2022.1820
18. Tennant PWG, Murray EJ, Arnold KF, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol*. 2021;50(2):620-632. doi:10.1093/ije/dyaa213
19. Lipsky AM, Greenland S. Causal directed acyclic graphs. *JAMA*. 2022;327(11):1083-1084. doi:10.1001/jama.2022.1816
20. Pearl J. Causal diagrams for empirical research. *Biometrika*. 1995;82(4):669-688. doi:10.1093/biomet/82.4.669
21. Terman SW, Lin CC, Kerr WT, DeLott LB, Callaghan BC, Burke JF. Changes in the use of brand name and generic medications and total prescription cost among medicare beneficiaries with epilepsy. *Neurology*. 2022;99(8):e751-e761. doi:10.1212/WNL.0000000000200779
22. Hernan MA, Robins JM. Treamtent-confounder feedback and G-methods for time-varying treatments. In:. *What If*. Chapman & Hall/CRC; 2020:2047-2276.
23. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9-12):1393-1512. doi:10.1016/0270-0255(86)90088-6
24. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. doi:10.1093/biomet/70.1.41
25. Thomas L, Li F, Pencina M. Using propensity score methods to create target populations in observational clinical research. *JAMA*. 2020;323(5):466-467. doi:10.1001/jama.2019.21558
26. Shardell M, Chen C, McCoy RG. Doubly robust causal modeling to evaluate device implantation. *JAMA Intern Med*. 2024;184(7):834-835. doi:10.1001/jamainternmed.2024.1181
27. Kang JDY, Schafer JL. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat Sci*. 2007;22(4):523-539. doi:10.1214/07-sts227
28. Suissa S. Immortal time bias in pharmacoepidemiology. *Am J Epidemiol*. 2008;167(4):492-499. doi:10.1093/aje/kwm324
29. Danaei G, Tavakkoli M, Hernán MA. Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. *Am J Epidemiol*. 2012;175(4):250-262. doi:10.1093/aje/kwr301
30. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care*. 2012;35(12):2665-2673. doi:10.2337/dc12-0788

31. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002; 58(1):21-29. doi:10.1111/j.0006-341x.2002.00021.x
32. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Evaluating metformin strategies for cancer prevention: a target trial emulation using electronic health records. *Epidemiology*. 2023;34(5):690-699. doi:10.1097/EDE.0000000000001626
33. 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
34. Hernán MA. How to estimate the effect of treatment duration on survival outcomes using observational data. *BMJ*. 2018;360:k182. doi:10.1136/bmj.k182
35. García-Albéniz X, Hernán MA, Logan RW, Price M, Armstrong K, Hsu J. Continuation of annual screening mammography and breast cancer mortality in women older than 70 years. *Ann Intern Med*. 2020;172(6):381-389. doi:10.7326/M18-1199
36. Gran JM, Røysland K, Wolbers M, et al. A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Stat Med*. 2010;29(26):2757-2768. doi: 10.1002/sim.4048
37. Keogh RH, Gran JM, Seaman SR, Davies G, Vansteelandt S. Causal inference in survival analysis using longitudinal observational data: sequential trials and marginal structural models. *Stat Med*. 2023;42(13):2191-2225. doi:10.1002/sim.9718
38. Caniglia EC, Zash R, Fennell C, et al. Emulating a sequence of target trials to avoid immortal time bias in pregnancy studies: an application to antibiotic initiation and preterm delivery. *Epidemiology*. 2023;34(3):430-438. doi:10.1097/EDE.0000000000001601
39. Angrist JD, Imbens GW, Rubin DB, Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 1996;91(434):444-455. doi:10.1080/01621459.1996.10476902
40. Hansford HJ, Cashin AG, Jones MD, et al. Reporting of observational studies explicitly aiming to emulate randomized trials: a systematic review. *JAMA Netw Open*. 2023;6(9):E2336023. doi:10.1001/jamanetworkopen.2023.36023
41. Hansford HJ, Cashin AG, Jones MD, et al. Development of the TrAnsparent ReportinG of observational studies Emulating a Target trial (TARGET) guideline. *BMJ Open*. 2023;13(9):e074626. doi:10.1136/bmjopen-2023-074626
42. Little RJ, Rubin DB. *Statistical Analysis with Missing Data*. 3rd ed. Wiley; 2019.
43. Rubin DB, Schenker N, Rubin DB, Schenker N. Ignorable nonresponse multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *J Am Stat Assoc*. 1986;81:366-374.
44. Tompsett D, Zylbersztejn A, Hardelid P, De Stavola B. Target trial emulation and bias through missing eligibility data: an application to a study of palivizumab for the prevention of hospitalization due to infant respiratory illness. *Am J Epidemiol*. 2023; 192(4):600-611. doi:10.1093/aje/kwac202
45. Jane-wit D, Horwitz RI, Concato J. Variation in results from randomized, controlled trials: stochastic or systematic? *J Clin Epidemiol*. 2010;63(1):56-63. doi:10.1016/j.jclinepi.2009.02.010
46. Omerovic E. Can contradictory results from RCTs that tested the same hypothesis be true simultaneously? Analyses of the VALIDATE-SWEDEHEART and BRIGHT-4 trials. *Eurointervention*. 2023;19(10):E802-E805. doi:10.4244/EIJ-D-23-00151
47. Hilkens NA, Hammerton G, De Vries NM, Bloem BR, Ben-Shlomo Y, Darweesh SKL. Guide to decomposition of causal effects into mediation, interaction, and direct effects: case study on aerobic exercise and Parkinson disease. *Neurology*. 2024;103: e209547. doi:10.1212/WNL.0000000000209547
48. Wang SV, Schneeweiss S, Franklin JM, et al. Emulation of randomized clinical trials with nonrandomized database analyses: results of 32 clinical trials. *JAMA*. 2023; 329(16):1376-1385. doi:10.1001/jama.2023.4221
49. Hernán MA, Dahabreh IJ, Dickerman BA, Swanson SA. The target trial framework for causal inference from observational data: why and when is it helpful? *Ann Intern Med*. 2025;178(3):402-407. doi:10.7326/ANNALS-24-01871
50. Matthews AA, Danaei G, Islam N, Kurth T. Target trial emulation: applying principles of randomised trials to observational studies. *BMJ*. 2022;378:e071108. doi:10.1136/bmj-2022-071108