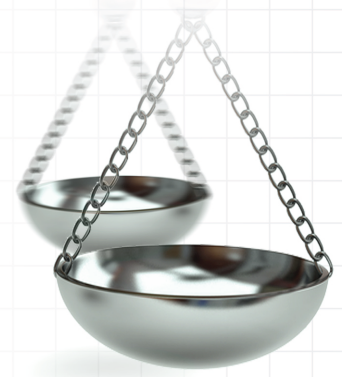




Target trial emulation: bridging observational studies and randomized trials for health decision-making

Journal of **Comparative Effectiveness Research**

Emilia Riggi^{*,1} , Eva Segelov^{2,3}  & Gian Luca Di Tanna¹ 

¹Department of Business Economics, Health & Social Care, University of Applied Sciences & Arts of Southern Switzerland, Stabile Piazzetta, Via Violino 11, 6928 Manno, Switzerland

²Department of Clinical Research, University of Bern, Freiburgstrasse 3, 3010 Bern, Switzerland

³Department of Radiation Oncology, Inselspital, Bern University Hospital, Freiburgstrasse 18, 3010 Bern, Switzerland

*Author for correspondence: Tel.: +41 (0)58 666 7738; emilia.riggi@supsi.ch

Randomized controlled trials (RCTs) are the gold standard generating evidence owing to their rigorous methodology. However, their logistical, financial and ethical limitations highlight the need for alternative approaches using real-world data. Target trial emulation (TTE) applies RCT design principles to estimate causal effects when trials are infeasible. TTE involves three steps: formulating a precise causal research question, explicitly specifying the protocol of the target trial, and rigorously replicating each component of the target trial, such as the eligibility criteria, treatment assignment and follow-up period, using available observational data. Statistical methods commonly used include propensity score matching, inverse probability weighting, G-methods and/or instrumental variables to address confounding and align observational data with the target trial design. Nonetheless, residual confounding, missing data and misclassification can bias results. Sensitivity analyses and transparent reporting are recommended. Notably, TTE frameworks utilizing continuously updated registry data enable ‘living protocols’ that can be iteratively refined as new data accumulate, representing an important evolution toward prospective-retrospective hybrid designs that maintain causal clarity while addressing emerging clinical questions. Though valuable, TTE complements rather than replaces RCTs, as both inform causal inference and clinical decisions.

Plain language summary: Using real-world data to answer clinical questions when randomized trials are not possible

What is this article about? This article describes how researchers can use real-world medical data from registries to design studies that mimic the structure of randomized controlled trials. This approach, called target trial emulation (TTE), applies the same principles as a clinical trial but uses existing health data instead of enrolling new participants. Unlike traditional retrospective analyses that use fixed datasets, registry-based TTE studies can take advantage of continuously updated data, creating opportunities for ‘living protocols’ that evolve as new information becomes available.

What were the results or methods described? The article highlights how TTE can improve the quality and reliability of registry-based research. By requiring researchers to predefine study criteria and use robust statistical methods to adjust for confounding factors, TTE reduces bias, prevents selective reporting and encourages transparent and reproducible science.

What do the results mean and why is this important? Registry-based TTE transforms observational research into a structured, causal framework that produces more credible and clinically relevant findings. The ‘living protocol’ approach allows studies to adapt as data accumulate, supporting timely and trustworthy evidence generation. By applying TTE principles, registry research can become a cornerstone of evidence-based medicine while maintaining scientific rigor and public trust.

First draft submitted: 28 October 2025; Accepted for publication: 13 February 2026; Published online: 19 March 2026

Keywords: causal inference • comparative effectiveness research • health decision-making • real-world data • target trial emulation



The causal effect, which assesses the consequences of actions, is fundamental to decision-making. Robust evidence to inform clinical and policy decisions is crucial for an evolving health research landscape. Randomized controlled trials (RCTs) remain the gold standard for determining causal effects and generating evidence because of their rigorous methodology [1–4]. Their main advantages include reducing bias, balancing confounders through randomization, ensuring adequate study power via sample size calculations, and systematic data collection that improves validity and reliability. Blinding, when present, further minimizes observer bias. However, research has highlighted significant concerns about the quality of many RCTs owing to methodological flaws or concerns about bias [5]. When well-designed and properly implemented, RCTs excel in internal validity; however, their controlled settings, strict criteria, and design can limit external validity, meaning results may not always generalize to real-world settings. Despite their strengths, with the increasing recognition of the logistical, financial and ethical constraints of RCTs, alternative approaches to causal inference using real-world data (RWD) have gained prominence. For instance, evaluating a drug for a rare disease may require lengthy recruitment, long follow-up, or raise ethical issues in randomization. Trials can also be prohibitively costly, demand excessive monitoring, or require complex infrastructure. Pragmatic trials using routine health data collected from registries, electronic health records and reimbursement/claims databases, known as RWD, are increasingly being used to provide insights into treatment effects under routine practice conditions, often using larger, more inclusive populations. Pragmatic trials and observational studies both aim to generate evidence in real-world clinical settings, but they differ fundamentally in their design and ability to support causal inference. Pragmatic trials leverage RWD to assess interventions as they occur in everyday clinical practice yet retain baseline randomization to preserve internal validity and minimize bias. Because treatment allocation is controlled, causal effects can be estimated more reliably, even as the trial is embedded in routine care. However, because pragmatic trials often allow flexible adherence, treatment delivery and patient management to reflect usual practice, they may sacrifice some degree of internal validity due to reduced control over confounding factors [6–8]. In contrast, observational studies lack randomization, and treatment decisions are determined by clinical practice or patient choice, making them more susceptible to confounding and bias. While purely observational designs can offer greater external validity and capture a wider, more representative patient population, they require careful design and analytical strategies to approximate causal effects reliably. The susceptibility of observational data to bias (Table 1 & Figures 1 & 2), such as selection bias and immortal time bias [9,10], can further limit their reliability, making it challenging to draw definitive causal conclusions [11–13]. This trade-off between internal validity in RCTs and external validity in pragmatic trials highlights the need to balance rigor with relevance when evaluating healthcare interventions for real-world decision-making. Target trial emulation (TTE) provides a transparent framework for researchers to estimate causal effects using observational data while explicitly mimicking the design of a hypothetical target trial. TTE helps ensure clarity of assumptions and supports robust causal inference.

Aim

TTE has emerged as a powerful methodological approach that combines the best of both worlds by applying RCT design principles to observational data [9,18,19]. TTE allows researchers to estimate the causal effects in a transparent, systematic and rigorous manner. This approach not only enhances the validity of observational studies but also bridges the gap between real-world evidence and traditional trials, offering a promising framework for more informed and timely health decision-making.

This paper serves as a practical guide for researchers and practitioners who may not have extensive expertise in advanced causal inference techniques. It summarizes recent advances in the field and presents them in an accessible manner, while offering practical advice on how to put these methods into practice. It highlights not only the theoretical underpinnings of TTE but also practical advice for applying these methods, making these tools more accessible to a wider range of researchers and practitioners.

Causal inference using potential outcomes

To lay the groundwork for the effective application of TTE, it is essential to provide a basic overview of causal inferences, particularly in the context of potential outcomes. Causal inference is the process of drawing conclusions about causal relationships from data. In this framework, potential outcomes, also known as counterfactuals, serve as fundamental concepts for understanding causal effects [20,21].

Table 1. Definition of common biases.

Confounding bias	Ref.
A confounder is a variable that influences both exposure (or intervention) and the outcome in a study, leading to a spurious association between them. Confounders can distort or obscure the true relationship, making it difficult to determine whether the observed effect is due to the intervention itself or the confounding variable. Key features of a confounder: 1. Associated with the exposure: the confounder is related to the factor being studied (e.g., a treatment or risk factor). 2. Associated with the outcome: the confounder independently affects the outcome of interest. An independent relationship means that a variable (in this case, a confounder) has a direct association with the outcome that is not caused by or dependent on the exposure being studied. In other words, the confounder influences the outcome on its own, regardless of whether the exposure is present. 3. Not on the causal pathway: the confounder is not an intermediate step in the causal relationship between the exposure and the outcome. Example: You are interested in studying whether coffee consumption (exposure) affects heart disease (outcome). Smoking is a potential confounder because: • People who drink more coffee may also be more likely to smoke (associated with exposure). • Smoking increases the risk of heart disease (associated with the outcome). • Smoking is not caused by coffee consumption (not part of the causal pathway). If smoking is not accounted for, the estimated effect of coffee on heart disease could be exaggerated or obscured.	[6–8]
Selection bias (internal and external validity) Selection bias (internal validity) refers to systematic differences in the characteristics between those included in the study and those who do not, or between study groups (treated and control), and those characteristics are related to either the exposure/treatment or outcome under investigation. This bias distorts the observed relationships between exposure/treatment and outcome, making the results unreliable or unrepresentative of the broader population. There are several potential sources of selection bias, commonest during the recruitment phase and in the processes of retaining participants in a study, due to: • Flaws in the selection process, such as unclear or not predefined inclusion and exclusion criteria; • Inherent personal characteristics that make some people more likely to be willing to participate in a study than others; • Whether some participants are more likely to be selected than others. Examples: 1. A researcher who only includes patients in the study if follow-up data are available introduces selection bias, if unmeasured factors affect both follow-up and the outcome; or 2. in a study evaluating the effect of a new exercise program on cardiovascular health, participants who voluntarily enroll may be more health-conscious, motivated, or have fewer comorbidities than those who do not enroll. This self-selection creates bias because the study population is not representative of the target population, and the estimated effect of the program may be exaggerated. Selection bias can threaten external validity when the study sample is not representative of the target population. This type of bias does not necessarily distort causal effect estimates within the study sample (internal validity may still hold), but it limits the generalizability of the results. Common causes include restricted recruitment (e.g., only urban hospitals, specific occupational groups), volunteer bias or exclusion of certain demographic subgroups. Potential sources: • Recruitment from specific geographic areas, institutions or subpopulations; • Demographic or occupational restrictions; • Volunteer or self-selection biases. Example: • Comparing disease rates in a particular occupational group with the general population can lead to the healthy worker effect, since employed individuals are generally healthier than the overall population.	[14]
Immortal time bias Immortal time bias is a distortion that modifies an association between an exposure and a health outcome, caused when participants assigned to the treatment or exposure group have a period of follow-up during which they cannot experience the outcome and are essentially rendered 'immortal', leading to an overestimation of the intervention effect. Examples: 1. As shown in Figure 2, Follow-up begins at the time of prescription for a drug, but patients must survive to receive that prescription. This creates a period during which the outcome cannot occur; or 2. in a study where groups are defined by requiring multiple events over time, such as the number of times a diagnosis was made, or a medication was prescribed. For example, in this published cohort study, the authors conducted a population-based retrospective cohort study using healthcare administrative data to compare the rates of acute myocardial infarction in elderly patients. To be considered as exposed subjects were required to have at least two prescriptions of a given cyclooxygenase-2 inhibitor, and they were followed from their first prescription for up to 1 year until death or hospitalization for acute myocardial infarction. On the other hand, nonusers of these drugs were followed from a date selected randomly from the observation period. In the 1-year follow-up, the study reported that the rate of acute myocardial infarction in subjects was similar with cyclooxygenase-2 inhibitor compared with that for the unexposed subjects. Subjects who died after their first prescription of cyclooxygenase-2 inhibitor were not included in the exposed group. On the other hand, the unexposed subjects could have died at any time during the follow-up period. Therefore, in this design, the time between the two prescriptions that define the exposed subjects is necessarily 'immortal'.	[15–17]

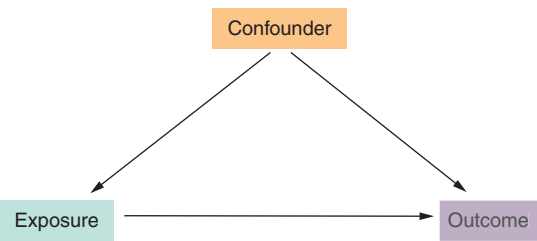


Figure 1. The principle of confounding.

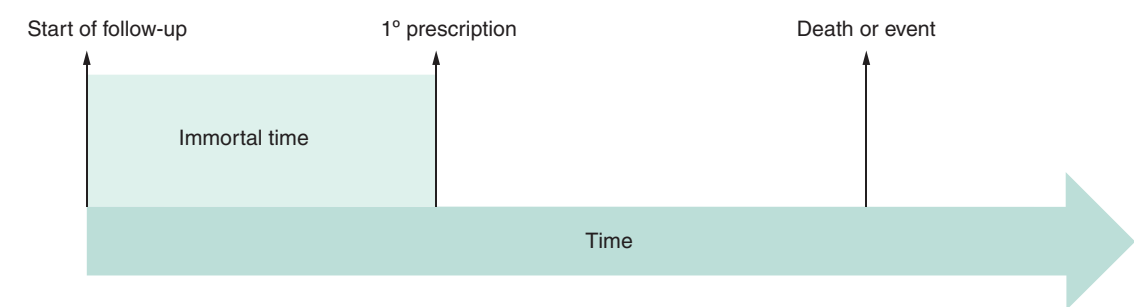


Figure 2. Immortal time bias.

Table 2. Causal estimands under the potential outcomes framework.				
Units	Potential outcomes		Individual-level causal effect	Summary causal effect
	Treatment $Y^t = 1$	Control $Y^t = 0$		
1	$Y_1^{t=1}$	$Y_1^{t=0}$	$Y_1^{t=1} - Y_1^{t=0}$	Comparison of $Y_i^{t=1} - Y_i^{t=0}$ for a common set of units
i	$Y_i^{t=1}$	$Y_i^{t=0}$	$Y_i^{t=1} - Y_i^{t=0}$	
N	$Y_N^{t=1}$	$Y_N^{t=0}$	$Y_N^{t=1} - Y_N^{t=0}$	

Table 3. Glossary definitions.	
Definition	Guiding question
The estimand (θ) is the quantity of interest to answer the trial’s research objective	What hypothetical quantity, or parameter, are we interested in?
An estimator ($\hat{\theta}(X)$) is a method for estimating approximately the estimand using data	Can we write this causal quantity in terms of observable data?
An estimate ($\hat{\theta}(X = .x)$) is a numerical value of the estimand that results from the use of a particular estimator	What algorithm will best approximate this statistical quantity?

These potential outcomes represent the set of outcomes that an individual can experience under different treatment conditions, where only one is observed, depending on the treatment received.

For instance, in a study assessing the effectiveness of oral aspirin in relieving headache pain, the potential outcome $Y^{t=1}$ represents whether relief is achieved one hour after taking aspirin ($T = 1$), whereas $Y^{t=0}$ represents the same outcome had aspirin not been taken ($T = 0$). The individual causal effect for the treatment is computed as $Y_i^{t=1} - Y_i^{t=0}$, which captures the difference in the potential outcomes for an individual (Table 2). When this effect is aggregated across a population, we derive a summary causal effect of the treatment in the population of interest, one of the most common of which is called the average treatment effect (ATE): $ATE = E[Y^{t=1}] - E[Y^{t=0}]$. Here the estimand (Tables 3 & 4) is the expected value of the difference in potential outcomes across all individuals.

However, the fundamental problem of causal inference arises because we cannot simultaneously observe both potential outcomes for the same individual [22,23]. Due to the presence of missing data, it is not possible to measure ‘directly’ the causal effect, but we must make assumptions. These include:

- Stable unit treatment value assumption (SUTVA) [24], which encompasses both consistency and no interference. The consistency assumption implies that an individual’s potential outcome, given their observed exposure history,

Table 4. Examples of treatment effects in observational data and randomized controlled trials.

Effect	Potential outcome notation
Average treatment effect (ATE)	$E[Y^{t=1}] - E[Y^{t=0}]$
Average treatment effect in the treated	$E[Y^{t=1}] - E[Y^{t=0}]T = 1$
Complier average causal effect (CACE)	$(E[Y^{t=1}] - E[Y^{t=0}]) \left(\begin{matrix} T^{A=0} = 0 \\ T^{A=1} = 1 \end{matrix} \right)$
T denotes the treatment/exposure (t = 1 treated/exposed, t = 0 untreated/unexposed). A denotes assignment to the treatment/exposure, which may or may not have been adhered to.	

is the outcome observed for that person: $Y_i^t = Y_i^T = Y_i$, if $T_i = t$. If an individual has been treated with aspirin ($T = 1$), the potential outcome under treatment $Y^{t=1} = 1$ is equal to the observed outcome of relief from headache 1 hour later ($Y = 1$). This assumption can be violated when there are multiple versions of the treatment that are not distinguished in the analysis, or when the intervention is applied inconsistently across individuals. Violations of consistency can lead to ambiguous causal effects and biased estimates, highlighting the importance of precisely defining and standardizing the treatment under study.

The no interference assumption allows that the potential outcomes for any unit do not vary with the treatments assigned to other units, and for each unit, there are no different forms or versions of each treatment level (no hidden version of treatments), which leads to different potential outcomes. For instance, in evaluating a new blood pressure medication, a patient's outcome should depend only on their own treatment, not on whether other patients receive the medication, and each patient should receive a well-defined version of the treatment to avoid different potential outcomes.

- Conditional exchangeability (unconfoundedness) assumes that, conditional on measured covariates Z , the potential outcomes under treatment t are independent of treatment assignment ($Y_i^t \perp\!\!\!\perp T_i \mid Z_i$ for all t). In other words, among individuals with the same values of the confounders, the treated would have experienced the same average outcome as the untreated had they received the alternative treatment, and *vice versa*. For example, in an study comparing two antihypertensive drugs, patients receiving drug A would have, on average, the same outcomes as those receiving drug B if they had received the other drug, conditional on covariates such as age, sex, baseline blood pressure and comorbidities. This assumption allows unbiased estimation of causal effects when confounders are properly measured and adjusted for, as is implemented in TTE frameworks.
- Positivity: Every individual has a nonzero probability of receiving each treatment level given their covariates ($0 < P(T_i = t \mid Z_i) < 1$ for all t). In the context of TTE, positivity ensures that all treatment strategies specified in the target trial protocol are empirically represented in the observed data. Violations of positivity may occur when certain subgroups are deterministically assigned to a specific treatment due to clinical guidelines, contraindications or structural features of the healthcare system. Without positivity, causal contrasts cannot be estimated for subgroups of the population.
- Unbiased measurement of key variables: outcomes, treatments and confounders must be measured accurately. Misclassification or measurement error can bias effect estimates even if other assumptions are held.

The reason why RCT can be used to estimate average causal effects is because the relationship between treatment and outcome is not confounded. Randomization ensures that the independent predictors of the outcome are equally distributed between the treated and untreated groups.

Target trial emulation framework

The TTE framework builds on these causal principles to design real-world studies that emulate hypothetical randomized trials that address a specific causal question. As the name suggests, by mimicking an ideal RCT using RWD, TTE adheres to the rigorous design principles of RCTs while capitalizing on the scale and relevance of the observational data. The process to design a TTE involves three main steps [9,19,25,26] (Figure 3).

Step 1: causal research question phase

The first step is to formulate the causal question of the research. It is useful to bear in mind that association inferences deal with real-world questions, whereas causal inferences deal with 'what-if' questions in counterfactual

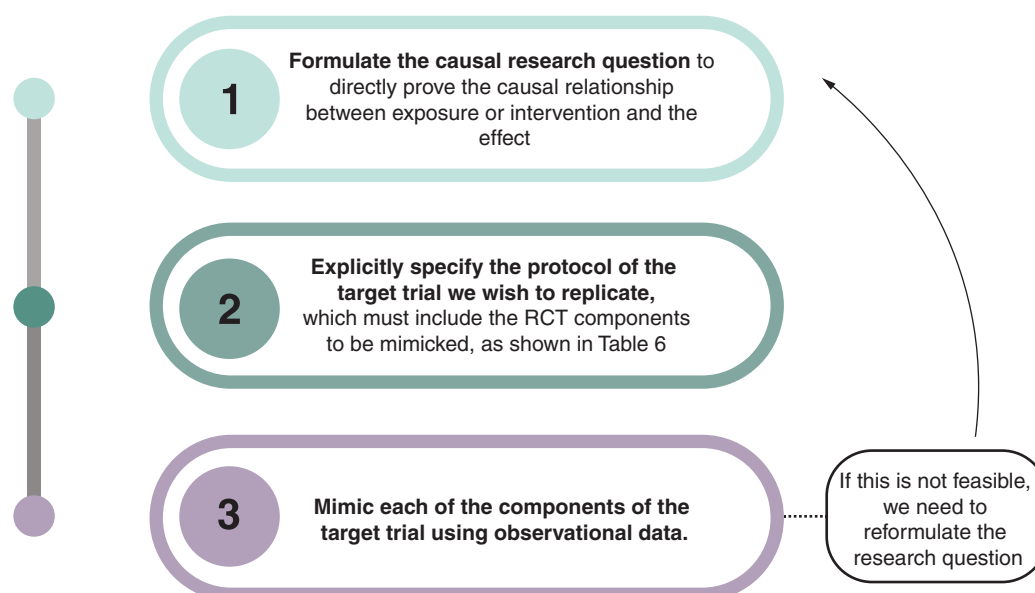


Figure 3. Target trial emulation framework.

worlds; that is, what would be the risk if everyone had been treated? what would be the risk if everyone had not been treated?

To formulate the causal question, as Hernàn suggests [27], it is helpful to define the causal effect in the study population as that which would have been observed in a hypothetical trial where an individual participant had been randomly assigned to a specific ‘treatment’ for some period.

For example, when studying the potential ‘causal effect’ of high BMI on cardiovascular outcomes, the question to be posed is: how would this study be conducted as an RCT? As it is impossible to randomize and assign participants to different BMI levels, the researcher needs to reformulate the question in terms of intervention, for example, targeting a certain BMI level through interventions such as diet, exercise regimens, medication or bariatric surgery. By doing so, the results will be interpretable and useful for decision making, as the researcher precisely specifies how the decrease in BMI is achieved. Each causal research question should have its own target trial protocol, and it is noteworthy that these may require different observational designs and analyses. The main advantage of explicitly specifying the hypothetical ‘target trial’ is that it forces investigators to articulate a well-defined intervention and causal question [28,29], and to assess whether the intervention can be meaningfully specified using the available data. While TTE is most suitable for questions that involve well-defined interventions, many other research questions remain important and informative even if they cannot be directly addressed using this framework. Explicitly defining the target trial clarifies the assumptions required for valid inference and supports transparent, evidence-based decision-making. As reported in these reviews [30,31], many studies identified as TTE did not explicitly define the target trial. This omission increases concern because specifying the components of the target trial serves the dual purpose of increasing transparency and reducing bias.

Step 2: target study protocol or design phase

Once the causal question has been formulated, the second step is to explicitly specify the protocol of the target study, addressing all components listed in Table 5. This is crucial because the flawed emulation of protocol components is more likely to lead to incorrect inferences [27]. As the goal is to be transparent, the target trial must be pragmatic, because observational data cannot emulate a placebo-controlled trial [9]. For the same reason, we can only emulate target trials without blind assignment. By focusing on the protocol and carefully thinking about each component, researchers can avoid many common design pitfalls of dealing with observational data.

Examples of published target trial protocols are presented in Table 6, where the authors assessed in the first one the impact of special educational needs (SEN) provision on health and education outcomes for a well-defined population [32]; while in the second paper the authors assessed the effectiveness of health interventions from RWD in the context of health technology assessment using a case study, the ‘Emergency Surgery or Not’ (ESORT)

Table 5. Target trial protocol components.

Protocol component	Description
Eligibility criteria	Who will be included in this study?
Treatment or intervention strategies	Which well-defined treatment or intervention will eligible individuals receive?
Treatment or intervention assignment procedure	How will eligible individuals be assigned to the treatment or intervention strategies?
Outcomes of interest	What outcomes will be measured during follow-up?
Causal estimand	Which causal estimand will be estimated with the observational data?
Start and end of follow-up	When does follow-up start and when does it end?
Statistical analysis plan	Which statistical analyses will be used to estimate the causal estimand?

study [33]. It is important to note that multiple choices can be made when emulating a target trial, including defining eligibility, treatment assignment, follow-up periods and analytical methods. For example, in the ESORT study, alternative approaches such as clone-censor-weighting could have been used to handle the 7-day grace period, and different strategies could address unmeasured confounding. Explicitly defining the target trial protocol clarifies which operational decisions were made, making the assumptions, analysis, and interpretation transparent and facilitating replication or alternative analyses in future work.

Step 3: emulation phase using observational data

The final step is to mimic each component of the target trial protocol using observational data. Following a properly conducted design phase, the emulation of the target study and the subsequent statistical analysis can be guided by the explicit protocol, defined in the previous step and the prespecified statistical analysis plan. As illustrated in Figure 3, this step should be conducted iteratively, since not all components of an ideal target trial may be fully feasible with the available data. When feasibility issues arise, investigators may need to return to the previous step to adapt the protocol, ensuring that the emulation remains transparent and methodologically rigorous. To ensure proper emulation of the target trial protocol, a key step is to align the following three components at so-called time-zero, or the baseline:

- Eligibility criteria (all the included individuals who meet the specified inclusion criteria);
- Treatment strategies (eligible individuals are assigned well-defined treatment or intervention);
- Start date and timing for end of follow-up.

A common mistake is the use of information collected during the follow-up period to select eligible individuals which introduces selection bias. To avoid this, researchers should not use the information collected post-baseline (future information after time zero) to determine which individuals should be included in the study.

Recognizing that when dealing with observational data, confounding is often a key concern, all potential measured covariates that contribute to the causality between the point intervention and outcome should be identified and included in the design phase of the target trial protocol and must be measured before treatment assignment. Conceptual tools such as directed acyclic graphs [34] can support this process by making causal assumptions explicit and guiding the selection of an appropriate adjustment set while avoiding overadjustment or collider bias. However, depending on the study context and data source, other sources of bias, such as selection bias or informative censoring, may represent equally important or even dominant threats to validity.

However, under what conditions can RWD be used for causal inference? A simple and common solution is to analyze RWD as if the treatment had been randomly assigned conditional on measured covariates Z . Thus, an observational study can be viewed as a conditionally randomized experiment when several causal inference assumptions hold, including consistency, conditional exchangeability and positivity (Table 7) [35–37].

A practical guide is given in Table 8 which shows that a target trial can be rigorously replicated, and causal effects can be estimated from the observational data.

Specific instruments and methods are required to adjust for all baseline confounding factors in order to emulate the random assignment of strategies at time zero. Causal effect estimation methods typically focus on reducing the impact of confounding factors by conditioning on a set of potential confounders (common causes of treatment/exposure and the outcome). Conditioning can be achieved by employing at least one of the following techniques:

Table 6. Target trial emulation protocol's examples.

1) Trial emulation to estimate the causal effect of special educational needs by Year 1 on unplanned hospitalizations by Year 6 in children with cleft lip and/or palate (without other congenital anomalies) [32].

Protocol component	Ideal target trial	Emulated target trial
Eligibility criteria	Region: England Year 1 started between 2008 and 2018. Diagnosed with cleft lip and/or palate prior to year 1 Born in England	Geography: England Year 1 started in a state school between 2008 and 2018. Identified in HES with cleft lip and/or palate before start of year 1 Has a birth record in HES Linked to NPD
Recruitment period	Year 1 started between the academic years 2008/2009 and 2018/2019	Year 1 Started between the academic years 2008/2009 and 2018/2019
Follow-up duration	From: randomization to the intervention To: the end of primary school OR loss of follow-up (e.g., emigration) OR death OR end of study	From: January Census in year 1 To: the end of primary school OR loss of follow-up in NPD OR death OR end of study/end of data (for HES: 31 August 2019)
Outcomes	Unplanned hospital utilization as defined by days in AE or APC Medical related absences as defined using half-day sessions. Unauthorized absences as defined using half day sessions	Unplanned hospital utilization as defined by days in AE or APC Medical related absences as defined using half-day sessions. Unauthorized absences as defined using half day sessions
Intervention to be compared	One of three categories of SEN (none, SEN, EHCP) to be delivered following randomization (between start of reception and end of year 1	One of three categories of SEN (none, SEN, EHCP) as recorded by the January census in year 1
Causal contrast	The average treatment effect of initiating SEN vs noninitiating SEN at all by year 1 on the number of unplanned hospital days expressed as a rate ratio. The average treatment effect of initiating EHCP vs initiating SEN by year 1 on the number of unplanned hospital days expressed as a rate ratio.	The average treatment effect of recording SEN vs noninitiating SEN at all by year 1 on the number of unplanned hospital days expressed as a rate ratio. The average treatment effect of recording EHCP vs recording SEN by year 1 on the number of unplanned hospital days expressed as a rate ratio. These estimands will be defined for the whole population and for the sub-populations of 'treated' and 'untreated' children, that is the children who were (or were not) recorded to receive the relevant intervention.
Analysis plan	Poisson or negative binomial regression (depending on the degree of overdispersion) of the number of events accountings for duration of follow-up. Clustering by school and/or local authority to be dealt with using either mixed effects models or robust inference (e.g., GEE).	Appropriate methods for confounding adjustment (such as regression adjustment and standardization, or propensity score-based methods) involving Poisson or Negative Binomial Regression (depending on the degree of overdispersion) of the number of events accountings for duration of follow-up. Clustering by school and/or local authority to be dealt with using either mixed effects models or robust inference (e.g., GEE).

EHCP: Education and HealthCare Plan; GEE: Generalized estimating equation; HES: Hospital episode statistic; NPD: National Pupil Database; SEN: Special educational need.

2) Target trial emulation of emergency surgery vs nonemergency surgery for acute appendicitis and acute gallstone disease [33].

Protocol component	Description of target trial of ES	How was the protocol element emulated in the ESORT?
Eligibility criteria	Inclusion criteria: - Patients were at least 18 years old at admission. - Emergency admission, via emergency department or primary care. - The condition was the reason for admission into hospital. - The diagnosis was confirmed by a consultant. Exclusion criteria: - According to clinical condition-specific exclusion criteria. - Emergency admission for the condition in the previous year. - Surgery for the condition within the previous 90 days. - Patient transferred between hospitals before surgical assessment.	Inclusion criteria: - Emulated directly from HES data. - Emulated directly from HES data. - Expert panel defined diagnostic (ICD-10) codes with equipoise between comparator strategies. - Emulation directly from HES data. Exclusion criteria: - Expert panel designated exclusion criteria with (ICD-10) codes. - Emulated directly from HES data. - Emulated directly from HES data (using definitions of treatment strategies below). - Emulated directly from HES data. Additional criteria according to data availability: - Patients were admitted to an ineligible hospital for ESORT. - Admission lacked information on admission or discharge status or date.

ES: Emergency surgery; ESORT: Emergency surgery or not; FCE: Finished consultant episode; HES: Hospital episode statistics; HRQoL: Health-related quality of life; ICD: International Classification of Diseases; ITT: Intention-to-treat; IV: Instrumental variable; LIV: Local instrumental variable; NES: Nonemergency surgery; ONS: Office for National Statistics; OPCS Office of Population Censuses and Surveys; PP: Per-protocol; QALY: Quality-adjusted life-year.

Table 6. Target trial emulation protocol's examples (cont.).

2) Target trial emulation of emergency surgery vs nonemergency surgery for acute appendicitis and acute gallstone disease [33].		
Protocol component	Description of target trial of ES	How was the protocol element emulated in the ESORT?
Recruitment period	Time zero is analogous to the time of randomization and is when all the eligibility criteria are met, the assignment to ES or NES occurs, and follow-up starts.	Emulation assumed time zero was the start date of the first FCE for the first admission, in which the specialty code was general surgery, colorectal surgery or upper gastrointestinal surgery.
Follow-up duration	Follow-up ends at the earliest of 1 year, death, or end of study period.	Emulation censored patients at the date of death, if that was within 1 year from day zero. Complete follow-up data were available for all patients.
Outcomes	- Life-years at 1 year from randomization. - QALYs at 1 year from randomization - total costs at 1 year from randomization. - Net monetary benefit at 1 year from randomization.	- Emulated directly from HES data (linked to ONS death data). - Emulation required adjusting life-years using published age- and sex-adjusted HRQoL scores from similar populations. - Emulation required calculating resource use for categories considered to be main drivers of total costs (length of stay, including critical care; operative and diagnostic procedures and readmissions up to 1 year) and valuing resource use data using relevant estimates of unit costs taken from national unit cost databases. - Emulated combining cost and QALY data.
Intervention to be compared	- ES defined as urgent, expedited, or immediate surgery for the condition. - NES: (1) medical management with no surgery for the condition and (2) surgery that did not meet the criteria for ES, either because of the not relevant procedure or after the 7-day time window, possibly preceded by medical management.	- Expert panel defined the 2 criteria for ES: (1) the procedure constituted "surgery for the condition" according to selected OPCS codes, and (2) to be considered "emergency," the panel designated a time window of 7 days from the date of assessment (see below). - Emulation assumed patients assigned NES if they did not meet ES criteria.
Causal contrast	- ITT effect (effect of assignment of patients to interventions at baseline). - PP effect (effect of complying with the trial protocol).	- ITT effect could not be emulated because information on the initial treatment assignment was not available from HES. - Emulation of the PP effect required taking differences between the treatment groups in estimated total costs, life-years, QALYs, and net monetary benefits at 1 year.
Analysis plan	- ITT analysis and PP analysis with adjustment for baseline prognostic factors. - Subgroup analyses by baseline age, sex, frailty, and number of comorbidities.	- Emulation of the PP analysis required using a LIV approach to mitigate the risk of confounding because of unmeasured prognostic factors associated with ES receipt. IV was the hospital's tendency to operate. Models were adjusted for a wide range of case-mix measures (age, sex, frailty level, comorbidity profile, ethnicity, index of multiple deprivation), fixed effects for each financial year, and proxies of quality of acute care (rates of emergency admission and mortality for each hospital and acute condition in 2009–2010 and in the year before the admission). - Emulated directly from HES data.
ES: Emergency surgery; ESORT: Emergency surgery or not; FCE: Finished consultant episode; HES: Hospital episode statistics; HRQoL: Health-related quality of life; ICD: International Classification of Diseases; ITT: Intention-to-treat; IV: Instrumental variable; LIV: Local instrumental variable; NES: Nonemergency surgery; ONS: Office for National Statistics; OPCS Office of Population Censuses and Surveys; PP: Per-protocol; QALY: Quality-adjusted life-year.		

Table 7. The basic assumptions to estimate causal effects from real-world data.

Consistency:	An individual's observed outcome is the same as their potential outcome under the observed treatment [$Y_i^t = Y_i^{t'} = Y_i$, if $T_i = t$]. This assumption requires well-defined treatment, specifically one that can be administered in multiple ways, but that will still lead to the same potential outcome.
Conditional exchangeability:	An individual's counterfactual outcome is independent of actual treatment given all confounders (Z) [$Y_i^t \perp T_i Z_i$ for all t]. This assumption is often called the no unmeasured confounders or unconfoundedness assumption.
Positivity:	All individuals in the target population have a nonzero chance of being assigned any treatment regardless of their characteristics ($0 < P(T_i = t Z_i) < 1$ for all t). In simple words, positivity holds for Z because there are people at all levels of treatment (i.e., $T = 0$ and $T = 1$) in every level of Z (i.e., $Z = 0$ and $Z = 1$).

- Restricting the study sample. This method involves narrowing the study sample to individuals who share the same value for potential confounders. Since these values do not differ among the subjects of your study, they cannot correlate with your independent variable and thus cannot confound the cause-and-effect relationship you are studying. For example, we are interested in study whether a low-carb diet can cause weight loss; and we know that age, sex, level of education and exercise intensity are all factors that may be associated with weight

Table 8. Practical guidelines to run a target trial emulation.

1. Define the target trial		Ref.
Research question	Clearly define the causal question of interest.	
Eligibility criteria	Specify inclusion and exclusion criteria for the study population.	
Treatment strategies	Define the treatment strategies.	
Assignment procedures	Describe how treatment would be assigned in an ideal RCT.	
Outcome	Define the primary outcome and the time point(s) at which it will be measured.	
Follow-up period	Specify the duration of follow-up and how losses to follow-up will be handled.	
2. Emulate the target trial using observational data		
Data source	Identify a suitable observational dataset (e.g., electronic health records, claims data and registries).	
Cohort selection	Apply the eligibility criteria to create a study population that resembles the target trial population.	
Baseline covariates	Identify baseline covariates that could influence treatment assignment and outcomes.	
Treatment assignment	Use statistical methods to account for confounding.	
Handle time-varying confounding	If treatment changes over time, use methods like g-methods to account for time-varying confounding.	
Assess assumptions	Consistency: ensure that treatment assignments are accurately recorded and that there is no confusion or inconsistency in how treatments are applied in practice. Conditional exchangeability: this assumption requires that all variables that could affect both treatment assignment and the and outcome are measured and included in the analysis. Perform balance checks on covariates to ensure that treatment groups are comparable. Positivity: check that both treatment strategies are possible for each covariate profile. Look at the distributions of the covariates within each treatment group to make sure there's overlap. If certain groups of individuals receive only one treatment and not the other, the assumption of positivity may be violated.	
Estimate the causal effect	Estimate the causal effect by comparing outcome(s) between treatment strategies.	
Sensitivity analyses	Conduct sensitivity analyses to assess the robustness of results to unmeasured confounding or other biases.	
3. Validate the TTE and report results		
Compare to RCTs	If possible, compare your results to existing RCTs on the same topic to assess the validity of your emulation.	
Transparency	Clearly report the target trial protocol and how the emulation was conducted.	
Limitations	Discuss potential limitations, such as unmeasured confounding, selection bias or measurement error.	
Tools and software		
Protocol	Outlines the steps to design a study that mimics the structure of a randomized controlled trial (RCT) using observational data (Table 5).	
Software	Use statistical software like R for causal inference. Some package examples: “MatchIt”: used for propensity score matching and other matching techniques, which are commonly employed to reduce bias in observational studies. “ipw”: it is designed for Inverse Probability Weighting, a method commonly used in causal inference. “gfoRmula”: it is a tool for performing g-formula estimations in causal inference, particularly in the context of longitudinal data. “EValue”: for sensitivity analyses for unmeasured confounding, selection bias, and measurement error.	[38–41]
TARGET guideline	This guideline provides a standardized approach for transparent reporting of such studies. It ensures that researchers and readers can properly evaluate the design, methods and results of observational studies that mimic the framework of a target trial.	[42,43]
Common pitfalls		
Misspecification of the target trial	A common problem is misspecification of the ‘target’ trial. If the design, treatments or outcomes of the target trial are incorrectly specified, the results of the replicated observational study will not be meaningful or valid. Solution: start by clearly defining all the components (mentioned above) of the target study trial.	
Ignoring or underestimating confounding	Observational data are often subject to confounding, and failure to properly adjust for confounding can lead to biased estimates. In TTE, confounders must be carefully controlled for, and the methods used to control for confounding should be well specified. Solution: carefully identify potential confounders, ensuring that both time-invariant and time-varying confounders are considered.	
RCT: Randomized controlled trial; TTE: Target trial emulation.		

Table 8. Practical guidelines to run a target trial emulation (cont.).

3. Validate the TTE and report results	
Ignoring selection bias due to loss to follow-up	In observational studies, loss to follow-up is common and can lead to selection bias, where individuals who drop out of the study differ from those who remain. This can invalidate the causal inference by affecting the treatment assignment and outcome. Solution: implement methods to handle missing data, such as multiple imputation or IPW for missing data, and perform sensitivity analyses to assess how different assumptions about missing data affect the estimates.
Inappropriate handling of post-treatment variables	Post-treatment variables (e.g., variables affected by the treatment) should be handled carefully, as their inclusion in the analysis may lead to post-treatment bias. Solution: post-treatment variables should not be directly included in the model as time-zero confounders.
RCT: Randomized controlled trial; TTE: Target trial emulation.	

loss, as well as with the diet our subjects choose to follow. In this case, we restrict our subject pool to 45-year-old women with bachelor's degrees who exercise at moderate levels of intensity between 100 and 150 min per week. However, while effective, this method can significantly reduce sample size and limit generalizability.

- **Stratification.** This method divides the population of interest into subgroups (strata) based on confounder levels. The treatment effect is then analyzed within each stratum. For example, participants in a study on aspirin use for headache relief can be stratified by baseline headache severity. Stratification helps compare treated and untreated groups within strata where confounding variables are similar. A pooled analysis across strata can provide an overall estimate, though this method may be less effective when there are many confounders or confounders with continuous values.

Both restricting the study sample and stratification are typically applied to a limited number of key variables, while more flexible methods are often preferred when many confounders must be addressed.

- **Matching.** This technique involves selecting individuals from the treated and untreated groups so that the distribution of confounders is balanced across these groups. One common approach is propensity score matching, where a propensity score is calculated for each participant representing the probability of receiving the treatment given their observed characteristics (e.g., age, BMI and disease duration). Participants in the treatment and control groups with similar propensity scores are then matched, creating comparable groups that reduce confounding. For example, in studying the effect of a new diabetes medication, researchers may use propensity score matching to ensure that participants in the treatment and control groups have similar baseline characteristics, such as age, BMI and duration of diabetes. Matching can reduce bias but may exclude unmatched participants, which can reduce effective sample size.
- **Multivariable regression adjustment.** It adjusts for confounding by including potential confounders as covariates in a regression model. For instance, in analyzing the impact of a public health campaign on smoking cessation, factors such as age, education, income, and prior smoking history can be included as covariates in the regression model. Standard regression models generally assume a linear relationship between covariates and the outcome (on the scale of the model's link function), but this assumption can be relaxed by including nonlinear terms, polynomials or interaction effects when appropriate. Multivariable regression is widely used, flexible, and can handle multiple confounders simultaneously, but care must be taken to specify the model correctly.
- **G methods [21,44,45].** Time-varying confounding arises when covariates that influence treatment assignments also change over time and are themselves affected by prior treatment. In this setting, standard regression adjustment may induce bias, as conditioning on post-treatment variables can block or distort causal pathways between treatment and outcome. For example, in an observational study emulating a target trial to estimate the effect of long-term statin use on cardiovascular mortality, low-density lipoprotein (LDL) cholesterol levels may influence treatment decisions at each follow-up while also being affected by prior statin use. LDL cholesterol is therefore a time-varying confounder that lies on both the causal pathway and the treatment assignment mechanism (Figure 4). Adjusting for LDL cholesterol using standard regression methods would result in biased estimates by partially blocking the treatment effect. G-methods address this challenge through approaches such as:

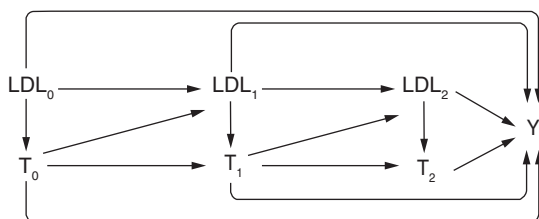


Figure 4. Time-varying confounder.

- Inverse probability weighting, typically implemented through marginal structural models, creates a pseudo-population in which treatment assignment is independent of measured baseline and time-varying confounders by weighting individuals by the inverse probability of their observed treatment history. This approach appropriately accounts for treatment–confounder feedback.
- The G-Formula and G-Estimation are both advanced methods, which provide alternative frameworks for estimating causal effects in the presence of time-varying confounding by explicitly modeling treatment strategies and outcome processes over time.

All the above methods rely on the assumption of no unmeasured confounding, which is often not plausible given the observational nature of the study design.

To overcome this problem, investigators may rely on a limited set of advanced causal inference methods. While valuable in specific contexts, these approaches depend on strong identification assumptions that must be carefully assessed:

- Instrumental variables, this method is used to control unobserved confounding by exploiting external variables (called instruments) that affect the treatment or exposure but do not directly influence the outcome, except through the treatment. The instrument needs to satisfy three conditions: relevance, the instrument must be strongly correlated with the treatment or exposure, exclusion restriction, the instrument should not be associated with the outcome, except through its effect on the treatment or exposure and no unmeasured common causes of the instrument and the outcome [46].
- Regression discontinuity design, a quasi-experimental research design that is useful for estimating causal effects in situations where treatment assignment is determined by whether an observed covariate crosses a specific threshold [47]. For example, a scholarship awarded based on a test score cutoff allows the comparison of individuals slightly above and below the threshold, assuming that those near the cutoff are similar in all respects except for receiving the scholarship. Causal effects are identified by comparing individuals close to the cutoff, under the assumption that units near the threshold are otherwise comparable.
- Interrupted time series design, a quasi-experimental research design too, used to evaluate the effect of an intervention or treatment on an outcome over time, particularly when the intervention happens at a clearly defined point [48]. Their validity relies on alternative identification assumptions, such as the absence of concurrent interventions and stable pre-intervention trends.

By carefully selecting and applying these methods, researchers can emulate randomization and obtain valid causal estimates, even in observational settings. The choice of method should align with the research question, study's data characteristics and the feasibility of assumptions. Most of the studies often used advanced statistical techniques of causal inference, such as G-methods for time-varying confounding or inverse weighting probability, followed by matching with propensity score [30,49,50].

Analogous to RCT, given the observational nature of treatment assignment, similar estimands can be computed in TTE. They are labelled with the term observational-analog as follows:

- Observational-analog intention-to-treat (ITT), because the closest observational analog ITT is a comparison of initiators of the different strategies, assuming adequate adjustments for baseline confounders, regardless of whether individuals continue the strategies after baseline;
- Observational-analog per-protocol (PP), where adjustment for baseline and time-varying confounding is necessary when the treatment strategies under study are sustained over time. G-methods are the best option for validating the PP [9].

Confidence in the results of the TTE framework may be increased by comparison with existing RCTs and the potential use of negative control outcomes (outcomes for which the well-defined intervention of interest is not expected to have an effect).

Strength & limitations of TTE

Although the TTE concept is relatively new and was first proposed as a framework in 2016, it is increasingly being used to infer causality from observational data. This approach is a popular method used in epidemiology and causal inference to estimate causal effects, especially when RCTs are not feasible [30,31,49,50]. The most common study topics were cancer, cardiovascular, cerebrovascular and infectious diseases.

The TTE framework pushes researchers to ask meaningful causal questions about actions, treatments or interventions, leading to practical evidence useful for decision-making as an actionable causal inference. This forces us to define the causal effect in a hypothetical trial, the target trial, as the inferential target. Focusing mainly on the emulation of the hypothetical target trial, specifying the target trial protocol prevents self-inflicted avoidable biases due to a flawed study design. As reported by Wang [51], real-world evidence studies can reach conclusions similar to those of RCTs when the design and measurements can be closely emulated.

Although this approach has gained traction for estimating causal treatment effects in real-world settings, it does come with several limitations.

Despite efforts to mimic an RCT, the use of observational data inherently presents challenges in controlling for all possible confounders. In an ideal RCT, randomization ensures that confounders are equally distributed across treatment groups; however, in observational data, confounding variables may be unmeasured or incompletely captured. Even with advanced statistical methods, researchers may not account for all relevant confounders, which can lead to residual confounding and bias causal estimates. To acknowledge this limitation and improve the reliability of causal estimates, researchers can employ various strategies as sensitivity analysis to evaluate the robustness of the findings to potential unmeasured confounding. Although sensitivity analysis cannot eliminate residual confounding, it can provide an indication of how much it might influence the study's results.

Researchers can use hypothetically simulated bias analysis scenarios to examine the effect of potential confounding factors. Specifying a plausible distribution for an unmeasured confounder helps to evaluate whether residual confounding is likely to impact on the treatment-outcome relationship and, consequently, on the conclusions drawn from the data [52].

Another approach can be used to assess whether results are likely to be robust even in the presence of unmeasured confounders is the E-value analysis [53]. The E-value quantifies the minimum strength of association that an unmeasured confounder would need to have with both the treatment and the outcome (beyond measured confounders) to explain the observed treatment effect [41].

Missing data presents significant challenges for accurate causal inference. RWD, such as electronic health records or claims data, often suffer from missingness in key variables, such as baseline confounders (information such as socioeconomic status or lifestyle factors might be unrecorded), exposure (timing or dosage of treatments could be inconsistently captured) and/or outcome (may not always be systematically tracked, especially if patients move to different care settings). As with RCTs, specifying a missing data strategy when designing the emulated trial to avoid *post hoc* adjustments and selective reporting is highly recommended. A transparent approach is essential, with clear reporting of the extent of missing data, the methods used to deal with it, such as complete case analysis and advanced methods such as multiple imputation [54], and the sensitivity of results to different assumptions about missingness. Transparent and standardized reporting is essential for TTE studies to ensure reproducibility, interpretability, and credibility of causal inferences derived from observational data. In addition to clearly describing the extent of missing data and methods used to address it, TTE studies should comprehensively report all components of the target trial protocol. Inconsistent or incomplete reporting of eligibility criteria, treatment strategies, follow-up, outcomes, causal estimands and analytic choices has been identified as a major limitation in current practice [49]. Existing observational study reporting frameworks, such as STROBE and RECORD, do not fully capture the specific requirements of TTE. The recently developed TARGET (Transparent Reporting of Observational Studies Emulating a Target Trial) [42,43] guideline addresses this gap by providing a 21-item checklist that emphasizes explicit specification of the target trial, mapping to observational data, and transparent description of analysis methods. Adopting TARGET, including structured tables summarizing key trial components and prespecified sensitivity analyses, can promote consistent and clear reporting, enhance the comparability of TTE studies and strengthen the reliability of real-world evidence for informing clinical and policy decisions.

Despite these limitations, when carefully designed and appropriately analyzed, the TTE approach remains a valuable tool for understanding causal relationships in contexts where randomized trials are not feasible. Recognizing and mitigating these limitations is critical for ensuring the validity and interpretability of results derived from this approach. However, it should not be considered as the ‘default gold standard’ for all observational studies; other observational study designs may be more appropriate and may provide additional insights into causality [55].

As guidance on RWE continues to evolve [56–58], TTE is expected to play an increasingly central role, offering a robust framework to ensure that real-world analyses meet the methodological rigor necessary for regulatory evaluation.

Systematic reviews [59,60] have highlighted growing concerns about research integrity arising from the proliferation of single-factor analyses using large health datasets such as NHANES and the UK Biobank, many of which rely on simplistic associations and inadequate control of bias. In this context, TTE allows for a broader regulatory shift toward integrating high-quality observational data in a way that preserves scientific validity while addressing practical and ethical constraints of clinical research.

Strengthening registry-based research with target trial emulation

An important methodological consideration in registry-based TTE studies is the dynamic nature of large, continuously updated data sources. Unlike traditional retrospective analyses based on fixed datasets, registries evolve over time as clinical practices change, emerging treatments are introduced and patient populations shift. These features create both opportunity and risk: while registries enable large-scale, timely analyses, they are also particularly vulnerable to analytic flexibility and time-related biases, including temporal confounding.

The concept of ‘living protocols’ naturally emerges, where study protocols may be updated as new data accrue or clinical standards evolve. For example, a living protocol emulating the effect of newly approved cancer therapies may need to update eligibility criteria over time to reflect new clinical indications, incorporate newly treated patients and recalculate treatment assignment probabilities and inverse probability weights. These ongoing adjustments require repeated sensitivity analyses and careful monitoring of assumptions to ensure valid causal inference.

TTE offers a promising solution to many of these challenges, providing a structured causal framework that mitigates the key sources of bias and analytic flexibility:

- Prevents data dredging and hypothesizing after the results are known: TTE requires researchers to pre-specify eligibility criteria, exposures, comparators, outcomes and follow-up periods, effectively serving as a prospective protocol that limits *post hoc* fishing for significant results.
- Reduces false discoveries: robust causal inference methods within the TTE framework, including inverse probability weighting and g-methods, help control for time-varying confounding and produce effect estimates with stronger causal interpretation.
- Promotes transparency and reproducibility: when paired with protocol registration and code sharing, TTE creates a fully auditable research workflow and makes replication feasible.

Adopting TTE more widely across registry-based research would therefore not only improve methodological rigor but also help safeguard the credibility of open data initiatives. It transforms registry studies from a source of noisy associations into a powerful engine for causal inference, aligning the goals of Open Science with trustworthy and clinically meaningful evidence.

Conclusion

TTE is a valuable tool for regulatory evidence-based decision-making, especially because it allows the comparison of pragmatic strategies in real practice. The core strategy is to use available health datasets to mimic hypothetical RCT in a transparent and systematic way, minimizing potential biases.

Applying this framework proficiently could extend the benefits of using RWD data, especially for conditions in which interventions are already widely implemented. Additionally, TTE can serve as an exploratory tool to test initial hypotheses and generate evidence that can inform the design and implementation of future RCTs. Structured reporting is essential for quality assessment, evidence synthesis and translation of evidence into policy and practice.

However, it is important to understand the scenarios in which TTE is an appropriate method, its limitations, and how it complements rather than replaces RCTs. TTE relies heavily on detailed and high-quality observational data. If such data are incomplete, lack key confounders or have measurement errors, the validity of TTE results

may be compromised. In situations where it is possible and ethical to conduct an RCT, it remains the first choice because minimize confounding through randomization, an advantage that TTE can only replicate.

Each approach, RCT or TTE, has unique strengths and is best viewed as complementary tools in a broader methodological framework. TTE should be viewed as filling gaps and enriching the causal evidence base. The most robust causal conclusions come from the integration of both methods, using RCTs to anchor the analysis and TTE to extend the findings to broader real-world settings.

Summary points

- Randomized controlled trials (RCTs) remain the gold standard for establishing causal effects but are often limited by cost, logistics and ethical constraints.
- Target trial emulation (TTE) provides a structured approach to estimate causal effects using real-world data when RCTs are not feasible.
- TTE involves explicitly defining the research question, specifying a 'target trial' protocol, and replicating each trial component using observational data.
- Common statistical methods in TTE include propensity score matching, inverse probability weighting, G-methods and instrumental variables to minimize confounding.
- Registry-based TTE frameworks introduce the concept of 'living protocols,' where study designs evolve as registry data are continuously updated.
- This dynamic approach enables adaptive, ongoing evidence generation while maintaining causal clarity and methodological rigor.
- The shift toward prospective-retrospective hybrid TTE designs represents a key methodological evolution in modern clinical research.
- By aligning registry research with TTE principles, researchers can transform real-world data into robust, clinically meaningful evidence.
- Broad adoption of TTE in registry-based research can enhance research integrity, improve reproducibility and strengthen trust in open data science.

Financial disclosure

This research received no specific grants from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests disclosure

The authors have no competing interests or relevant affiliations with any organization or entity with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Writing disclosure

No funded writing assistance was utilized in the production of this manuscript.

CRedit authorship contribution statement

Riggi Emilia: writing – review and editing, writing – original draft, conceptualization. Segelov Eva: writing – review and editing. Di Tanna Gian Luca: writing – review and editing, conceptualization, supervision.

Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/>

References

Papers of special note have been highlighted as: • of interest

1. Caparrotta TM, Dear JW, Colhoun HM, Webb DJ. Pharmacoepidemiology: using randomised control trials and observational studies in clinical decision-making. *Br. J. Clin. Pharmacol.* 85(9), 1907–1924 (2019).
2. Gale RP, Zhang MJ, Lazarus HM. The role of randomized controlled trials, registries, observational databases in evaluating new interventions. *Best Pract. Res. Clin. Haematol.* 36(4), 101523 (2023).
3. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research. *BJOG Int. J. Obstet. Gynaecol.* 125(13), 1716 (2018).

4. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. ERIC Notebook. *Randomized controlled trials (second edition no.10)*. UNC Gillings School of Global Public Health. <https://sph.unc.edu/epid/eric/>
5. Wilkinson J, Heal C, Antoniou GA *et al*. Assessing the feasibility and impact of clinical trial trustworthiness checks via an application to Cochrane reviews: Stage 2 of the INSPECT-SR project. *J. Clin. Epidemiol.* 184, 111824 (2025).
6. Jager KJ, Zoccali C, MacLeod A, Dekker FW. Confounding: what it is and how to deal with it. *Kidney Int.* 73(3), 256–260 (2008).
7. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. ERIC Notebook. *Confounding bias part I (second edition no.11)*. UNC Gillings School of Global Public Health. <https://sph.unc.edu/epid/eric/>
8. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. ERIC Notebook. *Confounding bias part II (second edition no.12)*. UNC Gillings School of Global Public Health. <https://sph.unc.edu/epid/eric/>
9. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* 183(8), 758–764 (2016).
- **Seminal paper that introduced the target trial emulation (TTE) framework, outlining how to design observational analyses using the structure and rigor of randomized controlled trials.**
10. 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.* 79, 70–75 (2016).
11. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 188(8), E158–E164 (2016).
12. Wilson BE, Booth CM. Real-world data: bridging the gap between clinical trials and practice. *eClinicalMedicine* 78, 102915 (2024).
13. Rosenbaum PR. *Observational studies*. Springer, New York, NY (2002).1–17
14. Alexander LK, Lopes B, Ricchetti-Masterson K, Yeatts KB. ERIC Notebook. *Selection bias (second edition no.13)*. UNC Gillings School of Global Public Health. <https://sph.unc.edu/epid/eric/>
15. Lévesque LE, Hanley JA, Kezouh A, Suissa S. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 340, b5087 (2010).
16. Yadav K, Lewis RJ. Immortal time bias in observational studies. *JAMA* 325(7), 686–687 (2021).
17. Mamdani M, Rochon P, Juurlink DN *et al*. Effect of selective cyclooxygenase-2 inhibitors and naproxen on short-term risk of acute myocardial infarction in the elderly. *Arch. Intern. Med.* 163(4), 481–486 (2003).
18. Delitto A. Pragmatic clinical trials: implementation opportunity, or just another fad? *Phys. Ther.* 96(2), 137–138 (2016).
19. Seewald NJ, McGinty EE, Stuart EA. Target trial emulation for evaluating health policy. *Ann. Intern. Med.* 177(11), 1530–1538 (2024).
- **Demonstrates the application of TTE beyond clinical trials, showing its utility for assessing real-world policy interventions and public health outcomes.**
20. Holland PW. Statistics and causal inference. *J. Am. Stat. Assoc.* 81(396), 945–960 (1986).
21. Hernán MA, Robins JM. *Causal inference: What if*. CRC Press, Boca Raton, FL (2020).
22. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66(5), 688–701 (1974).
23. Holland PW. Statistics and causal inference. *J. Am. Stat. Assoc.* 81(396), 945–960 (1986).
24. Rubin DB. Randomization analysis of experimental data: the Fisher randomization test comment. *J. Am. Stat. Assoc.* 75(371), 591–593 (1980).
25. Hernán MA. Methods of public health research—strengthening causal inference from observational data. *N. Engl. J. Med.* 385(15), 1345–1348 (2021).
26. Hernán MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA* 328(24), 2446–2447 (2022).
27. Hernán MA. The C-word: scientific euphemisms do not improve causal inference from observational data. *Am. J. Public Health* 108(5), 616–619 (2018).
28. Fu EL. Target trial emulation to improve causal inference from observational data: what, why, and how? *J. Am. Soc. Nephrol.* 34(8), 1221–1228 (2023).
29. Petersen ML, van der Laan MJ. Causal models and learning from data. *Epidemiology* 25(3), 418–426 (2014).
30. Simon-Tillaux N, Martin GL, Hajage D *et al*. Conducting observational analyses with the target trial emulation approach: a methodological systematic review. *BMJ Open* 14(11), e2336023 (2024).
- **Provides a comprehensive systematic review of TTE applications, documenting methodological trends, reporting quality and future research needs.**
31. Hansford HJ, Cashin AG, Jones MD *et al*. Reporting of observational studies explicitly aiming to emulate randomized trials: a systematic review. *JAMA Netw. Open* 6(9), e2336023 (2023).
32. Nguyen VG, Lewis KM, Gilbert R, Dearden L, De Stavola B. Early special educational needs provision and its impact on unplanned hospital utilisation and school absences in children with isolated cleft lip and/or palate. *NIHR Open Res.* 3, 54 (2023).

33. Moler-Zapata S, Hutchings A, O'Neill S, Silverwood RJ, Grieve R. Emulating target trials with real-world data to inform health technology assessment: findings and lessons from an application to emergency surgery. *Value Health* 26(8), 1164–1174 (2023).
- **Illustrates a practical example of TTE applied to real-world health technology assessment, bridging research and regulatory decision-making.**
34. Pearl J. Causal diagrams for empirical research. *Biometrika* 82(4), 669–688 (1995).
35. Zivich PN, Cole SR, Westreich D. Positivity: identifiability and estimability. *arXiv*. <https://arxiv.org/abs/2207.05010> (2022).
36. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology* 20(1), 3–9 (2009).
37. Anders H. Sequence announcement: applied causal inference (online) (2014).
38. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J. Stat. Softw.* 42(8), 1–28 (2011).
39. van der Wal WM, Geskus RB. ipw: an R package for inverse probability weighting. *J. Stat. Softw.* 43(13), 1–23 (2011).
40. Lin V, McGrath S, Zhang Z. gfoRmula: parametric g-formula. *R Package Version 1.1.0* (2024). <https://CRAN.R-project.org/package=gfoRmula>
41. VanderWeele TJ, Ding P. Sensitivity analysis in observational research: Introducing the E-value. *Ann. Intern. Med.* 167(4), 268–274 (2017).
42. Hansford HJ, Cashin AG, Jones MD *et al.* Development of the TrAnsparent ReportinG of observational studies emulating a target trial (TARGET) guideline. *BMJ Open* 13(9), e074626 (2023).
43. Cashin AG, Hansford HJ, Hernán MA *et al.* Transparent reporting of observational studies emulating a target trial: The TARGET statement. *BMJ* 390, e087179 (2025).
- **Introduces the TARGET reporting guideline, establishing best practices for transparent and reproducible TTE research.**
44. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period. *Math. Model.* 7(9–12), 1393–1512 (1986).
45. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial. *Biometrics* 56(3), 779–788 (2000).
46. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 17(4), 360–372 (2006).
47. Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. *BMJ* 352, i1216 (2016).
48. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions. *Int. J. Epidemiol.* 46(1), 348–355 (2017).
49. Zuo H, Yu L, Campbell SM, Yamamoto SS, Yuan Y. The implementation of target trial emulation for causal inference: a scoping review. *J. Clin. Epidemiol.* 162, 29–37 (2023).
50. Scola G, Chis Ster A, Bean D, Pareek N, Emsley R, Landau S. Implementation of the trial emulation approach in medical research. *BMC Med. Res. Methodol.* 23(1), 186 (2023).
51. Wang SV, Schneeweiss S, Franklin JM *et al.* Emulation of randomized clinical trials with nonrandomized database analyses. *JAMA* 329(16), 1376–1385 (2023).
52. Groenwold RHH, Sterne JAC, Lawlor DA, Moons KGM, Hoes AW, Tilling K. Sensitivity analysis for the effects of multiple unmeasured confounders. *Ann. Epidemiol.* 26(9), 605–611 (2016).
53. Chung WT, Chung KC. The use of the E-value for sensitivity analysis. *J. Clin. Epidemiol.* 163, 92–94 (2023).
54. Li P, Stuart EA, Allison DB. Multiple imputation: a flexible tool for handling missing data. *JAMA* 314(18), 1966–1967 (2015).
55. Pearce N, Vandenbroucke JP. Are target trial emulations the gold standard for observational studies? *Epidemiology* 34(5), 614–621 (2023).
56. US Food and Drug Administration. Real-world data: assessing registries to support regulatory decision-making for drug and biological products. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products>
57. HMA-EMA roadmap advances EU regulatory guidance on real-world evidence. <https://becarispublishing.com/digital-content/blog-post/hma-ema-roadmap-advances-eu-regulatory-guidance-real-world-evidence>
58. European Medicines Agency. Reflection paper on use of real-world data in non-interventional studies. <https://www.ema.europa.eu/en/r eflection-paper-use-real-world-data-non-interventional-studies-generate-real-world-evidence-scientific-guideline>
59. Suchak T, Aliu AE, Harrison C, Zwiggelaar R, Geifman N, Spick M. Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US National Health Database. *PLoS Biol.* 23(5), e3003152 (2025).
- **Warns of the growing crisis of low-quality, data-mined research; underscores the need for structured causal frameworks like TTE to preserve research integrity.**
60. Spick M, Onoja A, Harrison C, Stender S, Byrne J, Geifman N. Quantifying new threats to health and biomedical literature integrity from rapidly scaled publications and problematic research. *MedRxiv*. Preprint posted online 9 July 2025. <https://www.medrxiv.org/content/10.1101/2025.07.07.25331008v1>