



# The Target Cohort Approach: An Extension of the Target Trial Framework to Nested Case-Control Studies with Incidence Density Sampling

Hailey R. Banack<sup>1</sup> · Robert W. Platt<sup>2</sup> · Ellicott C. Matthay<sup>3</sup>

Accepted: 1 August 2024 / Published online: 23 August 2024  
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

## Abstract

**Recent Findings** The target trial framework is a well-known tool for estimating causal effects from observational data. The target trial approach can be used with data from any type of observational study, but it has most often been used to emulate a hypothetical target trial using data from a prospective cohort study.

**Purpose of this Review** In this manuscript, we present the target cohort framework for estimating causal effects from case-control studies. The target cohort approach extends the existing target trial framework for estimating causal effects using observational data but has an explicit case-control perspective.

**Summary of this Review** There are clear conceptual links from randomized trials to cohort studies and from cohort studies to case control studies. The target cohort framework uses a nested case control study to emulate a cohort study that has been designed to emulate a hypothetical pragmatic randomized controlled trial. Both target trial and target cohort frameworks require clear specification of eligibility criteria, treatment strategies, treatment assignment (randomization), follow-up period, outcomes of interest, causal contrast, and analysis plan. We demonstrate the target cohort approach using an example of an observational study to estimate the causal effect of semaglutide, a type of GLP-1 medication sold under the brand name Ozempic, on adverse gastrointestinal events.

**Keywords** Case-control · Observational study · Target trial approach · Target trial emulation · Incidence density sampling

## Introduction

A properly designed nested case-control study with incidence density sampling is expected to yield effect estimates that are equivalent to results from a cohort study using the same data source [1, 2]. Nested case-control studies with

incidence density sampling can be used to estimate causal effects, just like cohort studies, by emulating the approach that would be used to conduct a randomized controlled trial [2, 3]. Hernán and others developed the target trial framework as a heuristic tool for estimating causal effects from observational data [4–6]. By emulating a target trial, effect estimates from an analysis of observational data (under standard assumptions) should be identical to the effect estimates that would have been obtained from a randomized controlled trial (RCT) answering the exact same causal question, except for random variability [7, 8]. The target trial framework involves specifying the protocol of a hypothetical pragmatic randomized controlled trial (i.e., the target trial) [4]. The target trial protocol is then implemented using observational data, most often using a cohort study design [5, 7, 9, 10]. The target trial approach has been infrequently discussed in the context of observational data from case-control studies [2]. However, it is possible to extend

✉ Hailey R. Banack  
hailey.banack@utoronto.ca

<sup>1</sup> Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, 155 College Street, Toronto, Ontario M5T 3M7, Canada

<sup>2</sup> Departments of Epidemiology, Biostatistics, and of Occupational Health, and of Pediatrics, McGill University, Quebec, Canada

<sup>3</sup> Center for Opioid Epidemiology and Policy, Division of Epidemiology, Department of Population Health, New York University Grossman School of Medicine, New York, USA

the target trial heuristic to make causal inferences from case-control data.

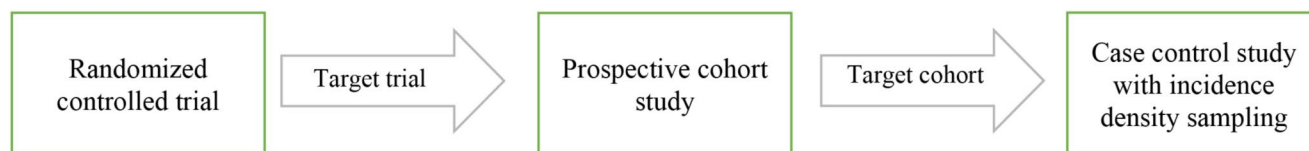
There are clear conceptual links from randomized trials to cohort studies and from cohort studies to case-control studies [11]. A nested case-control study with incidence density sampling can be conceptualized as a cohort study that uses an efficient sampling approach to form a comparison group [3]. In a cohort study, denominators of incidence rates are calculated by counting person time contribution for individuals in the exposed and unexposed groups. In a nested case-control study with incidence density sampling, cases and controls are matched on follow-up time, thus each set of cases and controls contribute the same amount of person-time [3, 12–14]. For readers interested in a brief primer on case-control studies please see the Appendix. In a cohort study design, incident cases of disease are compared to all non-cases in the cohort [15]. In a nested case-control study with incidence density sampling, incident cases of disease are compared to a control group comprised of a sample of non-cases drawn from the study cohort (i.e., the risk set) at the time each case occurred [16]. Both study designs can be used to estimate the average causal effect of an exposure on an outcome, by examining a causal contrast of exposed cases and non-cases (controls) with unexposed cases and non-cases (controls). In a case-control study, the comparison group is called the control group, comprised of a sample of non-cases from the study base, whereas in a cohort study, the comparison group is comprised of unexposed participants in the cohort [2, 17]. In a cohort study, the unexposed individuals act as a stand-in for the unobserved counterfactual group, representing the experience of the exposed participants had they not been exposed [18–20]. In contrast, in a case-control study, the non-cases act as a stand-in for an unobserved counterfactual group that represent the experience of exposed and unexposed cases from the study base had they not become cases. Thus, conducting a valid case-control study is predicated, in part, on identifying and sampling an appropriate group of controls from the study base [16, 21, 22]. Ultimately, the validity of causal effect estimates from both cohort and case control studies depend on how well each comparison group approximates the unobserved counterfactual group [23–25].

In this manuscript, we present a new framework for estimating causal effects from case-control studies called the target cohort approach. The target cohort approach extends the existing target trial framework and uses the same criteria

as the target trial approach. In the target cohort framework, we describe how to emulate a cohort study using a nested case-control design with incidence density sampling. Importantly, the cohort study being emulated in the target cohort approach has been developed using the target trial framework and therefore emulates a hypothetical randomized controlled trial. The target cohort framework is a heuristic tool for researchers seeking to make causal inferences from observational data when the study setting is suited to a nested case-control study with incidence density sampling. For example, in a study that utilizes outcome measures in an existing cohort, but involves collecting an expensive new exposure measure, the efficient sampling approach of a nested case-control study with incidence density sampling is appealing because it limits the exposure measurement to only a necessary subset of the full cohort. We demonstrate the target cohort approach using an example from a recently published randomized controlled trial examining the impacts of semaglutide, a type of GLP-1 medication sold under the brand name Ozempic, on adverse gastrointestinal events [26].

## The Target Cohort Framework

The target cohort framework extends the target trial approach to emulate a cohort study using data from a nested case-control study with incidence density sampling [4–6]. The target cohort approach emulates a hypothetical cohort study that was developed using the target trial approach. See Fig. 1 for a conceptual diagram linking the target cohort approach and the target trial approach. The target trial framework outlines the necessary components for making causal inferences from observational data [4]. The components of the target cohort framework are the same as the target trial, but tailored to considerations that arise specifically for the case-control study design with incidence density sampling. Both target trial and target cohort frameworks require clear specification of eligibility criteria, treatment strategies, treatment assignment (randomization), follow-up period, outcomes of interest, causal contrast, and analysis plan [6]. Ultimately, the goal of the target cohort approach is the same as that of the target trial framework: to make causal inferences using observational data. In the next section we introduce the example that will be used to demonstrate how the target



**Fig. 1** Conceptual diagram linking the target trial approach with the target cohort approach

cohort approach can be used to emulate a prospective cohort study and corresponding target trial.

### Example: GLP-1 Medication and Adverse Gastrointestinal Events

To illustrate the use of the target cohort approach in a case-control study using incidence density sampling, we describe a case-control study to estimate the causal effect of GLP-1 receptor agonist medications (e.g., semaglutide, brand name: “Ozempic”) on the risk of adverse gastrointestinal events. The use of GLP-1 medications has increased substantially in recent years owing to strong evidence from randomized controlled trials demonstrating weight loss, improved management of diabetes, and lower risk of cardiovascular disease [26–30]. A recently published RCT by Lincoff and colleagues reported a hazard ratio of 0.80 (95% CI: 0.72, 0.90), indicating a lower risk of cardiovascular events among individuals with pre-existing cardiovascular disease taking semaglutide compared to placebo.

Evidence from semaglutide trials has also demonstrated increased risk of adverse gastrointestinal events, including pancreatitis and bowel obstruction potentially related to the use of this medication [31]. Rapid weight loss using GLP-1 medications can also result in cholelithiasis (gallstones) [32]. Randomized trials are often not powered to determine rates of adverse events because the number of participants included in a trial is often limited and follow-up duration is often relatively short [33]. Adverse GI events are a relatively rare outcome among GLP-1 users and can take some time to develop [31]. Case-control studies are particularly useful in this setting—a rare outcome requiring long follow-up duration [22]. In this example, we will emulate Lincoff et al.’s RCT but examine a secondary safety outcome, adverse GI events, rather than the original primary outcome in their trial, cardiovascular events [26]. Given the increasing availability of GLP-1 medications in the population, there are now large secondary data sources (e.g., electronic health records databases) that can be used to examine the causal effect of GLP-1 medication on adverse GI events using observational data.

To begin a target cohort analysis, we must specify an appropriate research question [6, 34]. Our case-control study is designed to estimate the causal effect of initiating GLP-1 medication on risk of adverse gastrointestinal events among individuals with body mass index (BMI) greater than 27 kg/m<sup>2</sup> and pre-existing cardiovascular disease. We will describe a protocol for an RCT similar to Lincoff et al. [26], and describe how the elements of the trial can be emulated using a target trial and target cohort. Appendix Table S1 provides a worksheet that researchers can use to

operationalize the target cohort approach, linking the randomized trial, prospective cohort, and case-control study with incidence density sampling.

### Target Cohort Eligibility Criteria

The first step in the target cohort analysis is to describe the eligibility criteria to create the study cohort. Recall, a case-control study using incidence density sampling is conceptually the same as a cohort study but uses a sampling approach to assemble a group of non-cases. The protocol for describing eligibility criteria for the target cohort is the same as the protocol for describing the target trial. In Table 1, we describe the eligibility criteria for the RCT described by Lincoff and colleagues examining the use of semaglutide to prevent cardiovascular disease outcomes [26]. We also present the criteria that we would use to establish a target trial and target cohort emulating the RCT study population using a hypothetical EHR database. For the RCT, participants were at least 45 years of age, with body mass index (BMI)  $\geq 27$  kg/m<sup>2</sup>, and pre-existing cardiovascular disease, defined as previous myocardial infarction, stroke, or symptomatic peripheral artery disease [26]. Using data from an EHR database, we are able to emulate these inclusion criteria using data collected from routine clinical care encounters for the target trial and target cohort approaches. We can also emulate the exclusion criteria from the RCT using variables available in an EHR [26]. The exclusion criteria for the target trial and target cohort include observational analogs of self-report characteristics (e.g., planned coronary, carotid, or peripheral artery revascularization procedure vs. revascularization procedure documented in medical chart). There is a defined 5-year look-back window from the date eligibility is assessed for relevant inclusion and exclusion criteria related to medical history.

For the target cohort approach framed using a case-control design, there are additional eligibility criteria to consider related to who is eligible to be a case, who is eligible to be a control, and how many controls will be sampled per case. All individuals meeting the inclusion criteria to be in the study cohort are eligible to be a case and/or a control. Cases are any individuals in the study cohort that experience an incident outcome event (adverse gastrointestinal events) during the study period. The outcome of interest is described in greater detail in Step 5, below. Controls are individuals alive and sampled from the risk set (i.e., the eligible study cohort in the EHR) at the time the case event occurrence. The same individual can serve as a control multiple times and is eligible to become a case. There are differing opinions in the literature about how many controls to sample per case. Wacholder et al. recommended recruiting 4 controls per case based on the marginal gain in precision beyond a

**Table 1** Eligibility criteria for a randomized controlled trial, target trial, and target cohort

RCT	Target trial	Target cohort
<p><b>Inclusion Criteria:</b></p> <ul style="list-style-type: none"> <li>• Adults (<math>\geq 45y</math>)</li> <li>• BMI <math>\geq 27</math> kg/m<sup>2</sup></li> <li>• Cardiovascular disease (previous myocardial infarction, stroke, or symptomatic peripheral arterial disease)</li> </ul> <p><b>Exclusion criteria:</b></p> <ul style="list-style-type: none"> <li>• Diabetes (Type 1 or 2)</li> <li>• HbA1c <math>\geq 6.5\%</math></li> <li>• Treatment with glucose lowering medications within 90-days</li> <li>• Treatment with GLP-1 medication within 180 days</li> <li>• Class IV heart failure</li> <li>• End-stage kidney disease, dialysis</li> <li>• Planned revascularization procedure</li> <li>• Thyroid disease</li> <li>• Severe psychiatric disorder</li> <li>• Chronic pancreatitis</li> <li>• Malignant neoplasm (within 5 years)</li> <li>• Pregnancy or breastfeeding</li> </ul>	<p><b>Inclusion Criteria:</b></p> <ul style="list-style-type: none"> <li>• Adults (<math>\geq 45y</math>)</li> <li>• BMI <math>\geq 27</math> kg/m<sup>2</sup></li> <li>• Cardiovascular disease (previous myocardial infarction, stroke, or symptomatic peripheral arterial disease)</li> </ul> <p><b>Exclusion criteria:</b></p> <ul style="list-style-type: none"> <li>• Diabetes (Type 1 or 2)</li> <li>• HbA1c <math>\geq 6.5\%</math></li> <li>• Treatment with glucose lowering medications within 90-days</li> <li>• Treatment with GLP-1 medication within 180 days</li> <li>• Class IV heart failure</li> <li>• End-stage kidney disease, dialysis</li> <li>• Revascularization procedure</li> <li>• Thyroid disease</li> <li>• Severe psychiatric disorder</li> <li>• Chronic pancreatitis</li> <li>• Malignant neoplasm (within 5 years)</li> <li>• Pregnancy or breastfeeding</li> </ul>	<p>Same inclusion and exclusion criteria as in the target trial.</p> <p>Additional considerations required for sampling cases and controls.</p> <p><b>Sampling criteria</b></p> <ul style="list-style-type: none"> <li>• 1:4 ratio of cases to controls</li> <li>• <b>Eligible cases:</b> all individuals with a diagnosis of a major adverse cardiovascular event</li> <li>• <b>Eligible controls:</b> a random sample drawn from all individuals alive and in the risk set at the time a case was diagnosed.</li> </ul>

*Note* The target trial is emulating a hypothetical randomized controlled trial using data from a prospective cohort study, the target cohort is emulating the hypothetical prospective cohort study using data from a nested case control study with incidence density sampling

1:4 case to control ratio, but others have demonstrated the benefits of including between 10 and 50 controls [22, 35]. In our illustrative target cohort analysis, we will recruit 4 controls per case.

### Target Cohort Treatment Strategies

In the RCT, participants were randomly assigned to one of two treatment groups, subcutaneous semaglutide 2.4 mg per week or placebo injection [26]. An analogous treatment strategy must be identified for the exposure of interest in an observational context, either for the target trial or target cohort. As described in Table 2, in the observational context, medication initiation is indicated by first prescription date (2.4 mg semaglutide injection) recorded in the EHR. It is not possible to emulate a placebo-controlled trial using a target trial or target cohort approach. Non-placebo-controlled randomized trials may compare the treatment of interest to an active comparator, such as an alternative clinical treatment, or an inactive comparator, such as usual care [36, 37]. Huitfeldt et al. have discussed conditions for using an active comparator group in the context of a target trial [38]. The choice of active versus inactive comparators may have important implications for exchangeability. In this analysis, the comparison group is a group of individuals who are receiving usual care for weight management, most often diet or exercise advice. This approach emulates an unblinded pragmatic randomized controlled trial with treatment beginning at a defined time zero date (defined below, Step 4) for both the semaglutide and usual care group. In the

target trial approach, treatment is measured for all participants meeting inclusion criteria. Participants are assigned to the treated group (exposed) only if their electronic health record indicates a prescription for semaglutide at time zero, and otherwise they are considered unexposed. In the target trial, untreated participants are all individuals meeting the eligibility criteria but without a prescription for semaglutide. In the target cohort approach, treatment is assessed only for cases and the controls sampled from the study cohort. Cases and controls from the eligible study cohort are assigned to a treatment group based on their EHR-recorded treatment at time zero (i.e., if they match the defined treatment strategies under comparison). In this context, we define semaglutide initiation as any prescription recorded for the medication. Imposing a length-based exposure requirement, such as medication use for one year, or a fixed number of prescriptions, such as 6 continuous medication refills, could introduce immortal time bias [39, 40].

### Target Cohort Treatment Assignment

In an RCT, randomization is used to assign participants to a treatment group. If randomization is implemented correctly, treatment groups are assumed (on average) to be exchangeable and balanced according to both measured and unmeasured confounders. In the target trial and target cohort framework, analytic tools are used to emulate randomization and achieve conditional exchangeability [4]. Analytic approaches are used to achieve conditional exchangeability in observational data using measured covariates, ensuring

**Table 2** Strategies for defining and assigning treatment groups

	RCT	Target trial	Target cohort
<b>Treatment strategy</b>	<b>1. Treatment group A=1</b> -Semaglutide injection -Dose: 2.4 mg/week <b>2. Comparison group A=0 (placebo):</b> -Placebo injection	<b>1. Treatment group A=1</b> -Initiation of semaglutide injections -Dose: 2.4 mg/week -Continuous treatment -Medication initiation date is recorded as first prescription filled in EHR <b>2. Comparison group A=0</b> -No semaglutide use -Standard care for clinical management of obesity	Same as in target trial Among eligible cases and controls, treatment strategy will be assessed via EHR database
<b>Treatment assignment</b>	Randomization with 1:1 ratio	Use of analytic to emulating randomization with measured covariates Common approaches include regression adjustment, matching, propensity score methods, g-methods	Same as in the in target trial, emulating randomization using analytic approaches. Additional consideration required due to incidence density sampling approach. (e.g., propensity scores methods in nested case control studies requires that the sampling fraction of controls is known; see reference 37).
<b>Follow-up</b>	Follow-up period begins at randomization and continues until the occurrence of a cardiovascular event, censoring, withdrawal, or end of study at 48 months.	Beginning of follow-up at cohort entry (time zero) corresponding with exposure measurement and continuing until outcome, censoring, or end of study follow-up at 48 months.	Beginning of follow up: same as target trial End of follow-up: -For cases, end of follow-up occurs on the date they experience the outcome of interest -For all controls sampled from the risk set at the time the case occurs, their follow-up time ends on the date the case occurs (index date). -For controls not selected as controls matched to a case, they are not included in the study analysis and end of follow-up occurs upon censoring or the end of study follow-up at 48 months

balance between treatment groups and mimicking randomization [41–44]. At time zero, individuals are ‘assigned’ to a treatment group conditional on measured covariates. In this example, using electronic health records, time zero is an arbitrary calendar date chosen by the investigators. Individuals are considered to have been assigned to semaglutide if they have a prescription recorded on that date, otherwise will be considered to have been assigned to the usual care group at time zero. As a fictitious example, in an existing health record database, time zero could be January 1, 2022. This would be the date at which exposure status (treatment assignment) is assessed and the start of the accrual of person time for cases and controls.

In the target cohort approach, there are additional considerations related to achieving conditional exchangeability because of the case-control design and incidence density sampling approach [41, 45]. Specifically, confounder adjustment approaches must take into account the sampling approach: controls are matched to cases on follow up time and must be analyzed as pair-matched data, for example using conditional logistic regression [1]. Alternatively, if the sampling probabilities are known (i.e., the probabilities of being selected as a case or control from the eligible cohort), statistical analyses can account for the incidence density sampling by incorporating weights equal to the inverse of the sampling probabilities. In addition to matching on follow up time, matching on confounders is

also common in nested case-control studies with incidence density sampling to improve statistical efficiency, facilitate conditional exchangeability, and reduce the risk of positivity violations [46–49]. Modern approaches to achieve conditional exchangeability that are compatible with case-control incidence density sampling include propensity score methods, inverse probability of treatment weighting, parametric g-formula (standardization), and g-estimation [42–44, 50–52]. For example, Matthey and colleagues used the parametric g-formula to adjust for confounders within a nested case-control study with incidence density sampling by incorporating sampling weights [45].

### Target Cohort Follow-Up Period

The follow-up period for a typical RCT spans from randomization (treatment assignment) until the occurrence of the outcome of interest, censoring, withdrawal, or end of study. In a target trial, the start of follow-up is defined at cohort entry, often called study baseline, when the eligibility criteria are satisfied, and treatment assignment occurs (Table 2). The target trial framework explicitly describes that treatment assignment must occur at the same time as cohort entry to ensure time zero is well-defined and to avoid immortal time bias [5]. Time zero is the start of study follow up, the starting point for accrual of study outcomes. In a cohort study with primary data collection, time zero is typically the point



at which individuals begin their participation in the study. This is usually when the baseline measure of exposure is obtained and the starting point of follow-up for the outcome of interest. In a cohort study using a secondary data source, like an electronic health record, time zero is often defined by the investigators as a specific date at which eligibility is assessed, baseline exposure is measured, and study follow-up for the outcome begins. In a case-control study nested within a cohort, regardless of whether it is a primary data collection cohort or a secondary analysis of a large database, time zero must be clearly specified by the investigators. In the context of a case-control study, time zero can refer to a specific calendar date that is the same for all study participants (i.e., Jan 1, 2022) or to a time that is indexed to a defined entry event for each participant (i.e., becoming 65 years old).

In a target trial, follow-up time usually spans from time zero and until the occurrence of the outcome, loss to follow-up, or administrative censoring at the end of study follow-up [4, 53]. In the target cohort approach, follow-up still must begin at a well-defined time zero to ensure alignment of eligibility status, treatment assignment, and matching cases and controls on follow-up time. Although a case-control study is an outcome-dependent sampling design, time zero does not correspond to the time the case event occurs. This would be analogous to an illogical scenario in which investigators in a trial randomly assigned individuals to a treatment group once they experienced the outcome of interest. The crux of the nested case-control design with incidence density sampling is that each time a case occurs, a control, or set of controls, is sampled from the risk set at that point in time. The length of follow up for that grouping of cases and controls is identical by design: from time zero (set to time zero of the target cohort) to the date the case event occurred (for cases) and the date the controls were sampled from the risk set (for controls) [54]. For controls, the date on which the case-event occurred is sometimes referred to as the index date, so the follow-up time can be defined as the time period from time zero to index date.

In the context of a nested case-control study with incidence density sampling, a helpful heuristic is to consider each case-control grouping as emulating its own hypothetical RCT, a miniature version of a target trial if you will. The target cohort approach can be conceptualized as emulating a series of hypothetical mini-trials occurring within the study cohort. There are as many mini-trials as there are case-events. Figure 2 illustrates four so-called mini-trials, corresponding with four outcome events in the study cohort. The person-time contribution of cases is indicated by a solid line and the person-time contribution of controls is indicated by dashed line. In the first hypothetical mini-trial, a case occurred at 5.5 months. Four controls were sampled

from the risk set at the time the case occurred. Follow-up time for all cases and controls in mini-trial 1 spans from time zero ( $t_0$ ) to 5.5 months. Treatment assignment, defined as a prescription for semaglutide in the EHR ( $A=1$ ; blue) or not ( $A=0$ ; red), occurs at  $t_0$ . A similar structure applies for mini-trials 2–4. At the bottom of Fig. 2, there are three participants who remain in the study cohort from time zero to administrative censoring at the end of study follow-up. They do not experience the outcome of interest nor are they sampled as controls, so their person-time is not included in the analysis of a nested case-control study with incidence density sampling.

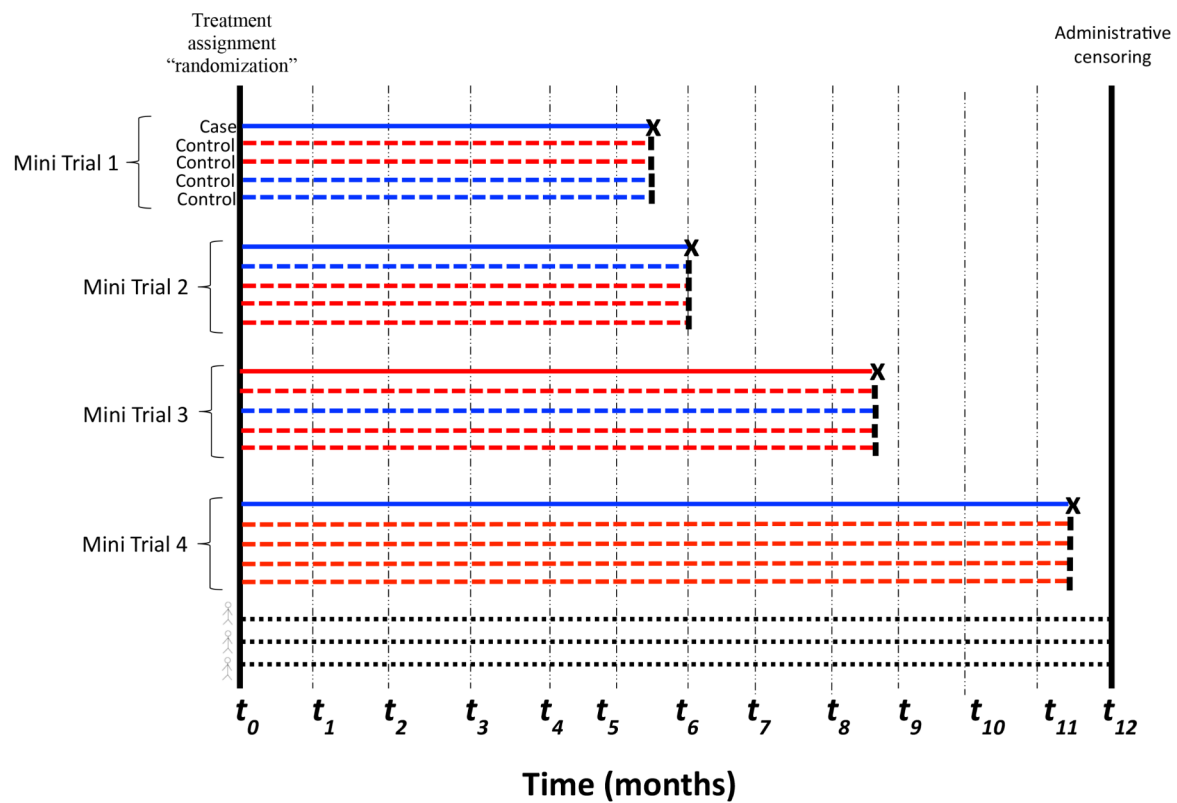
A nuance of the nested case-control design with incidence density sampling is that controls may go on to serve as a case at a later point in time or even as a control for another case. If a control does go on to become a case, they are matched with their own series of four controls at the time their case-event occurs. Extending the analogy of emulating hypothetical mini-trials, individuals can contribute person-time as a control to a mini-trial and also serve as a case in another mini-trial or contribute person-time as a control to more than one mini-trial (Fig. 3).

## Outcome of Interest

In RCTs, outcomes must be specified a priori. The CONSORT (Consolidated Standards of Reporting Trials) statement requires trial investigators to pre-specify primary and secondary outcome measures, whether physician adjudicated or patient-reported, and describe how they are assessed [55, 56]. Outcomes should be described in a similar level of detail in an observational study protocol. In our example, the outcome of interest in an RCT is incidence of any adverse GI outcomes, specifically biliary disease (including cholecystitis, cholelithiasis, and choledocholithiasis), pancreatitis (including gallstone pancreatitis), and bowel obstruction [31] (Table 3). In a target trial, these outcomes would be ascertained for exposed and unexposed members of the eligible cohort using ICD-10 diagnostic codes from the EHR. For our case-control study to emulate a target cohort, eligible cases are all individuals who have a diagnosis of the outcome of interest in the EHR, identical to the target trial. Cases are sampled from the study cohort during the follow-up period as described in steps 1–4 [57].

## Causal Contrast

RCTs commonly estimate intention-to-treat effects, comparing individuals randomized to Semaglutide treatment to those randomized to placebo or usual care. In our target cohort analysis, we are interested in the observational analog of an intention to treat effect, examining a causal



**Fig. 2** Illustration of target cohort framework as emulating series of hypothetical mini trials. This is a visual representation of the target cohort approach. The study cohort follow-up spans from baseline,  $t_0$ , to the end of study follow-up, administrative censoring, at 12mo. Consider each ‘grouping’ of cases and controls as an individual mini-trial. This illustration describes 4 case events (at 5.5 mo, 6 mo, 8.5 mo and 11.5 mo) and four controls sampled from the risk set at the time a case occurred. This results in four “mini trials” spanning from treat-

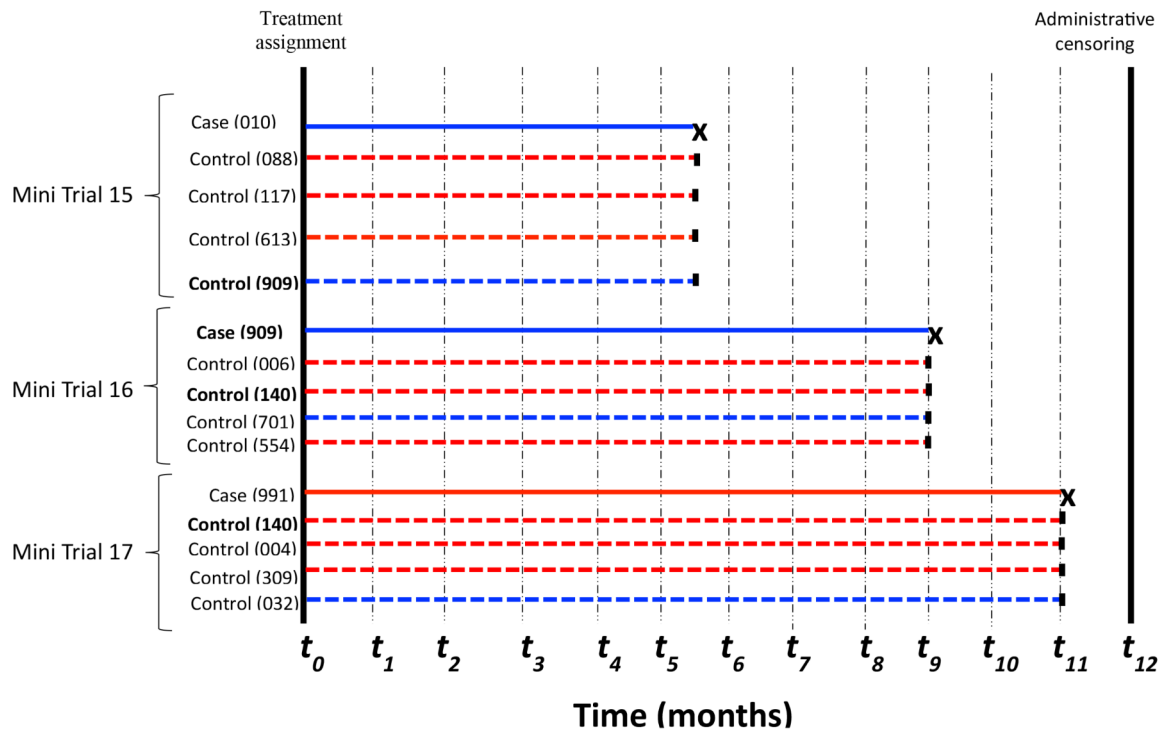
ment assignment, emulating randomization, to the end of follow-up when the case-event occurred. In the figure, person time contribution by cases is indicated by a solid line and person time contribution by controls is indicated by a dashed line. Blue represents Semaglutide use ( $A=1$ ) and red represents usual care ( $A=0$ ). The person time contribution of the three individuals in the study cohort who did not experience a case event and were not selected as control is represented by the black dashed line.

contrast between initiators and non-initiators of semaglutide at time zero [4, 26] (Table 3). However, there are other causal contrasts of potential interest in the context of target trial or target cohort approaches, including per protocol or as-treated effects [4, 53, 58]. Additionally, with a time-varying exposure, it is possible to examine contrasts considering changes in treatment throughout follow-up such as dynamic treatments using sequential nested trials [5, 59] and adaptive designs which involve change in treatment strategy in response to time-varying participant characteristics [53]. These extensions to the target cohort framework are possible so long as there are multiple measures of the exposure (treatment) of interest in the study cohort and treatment strategies are explicitly defined.

**Analysis plan**

In an RCT, assuming randomization was conducted appropriately, treatment groups are considered balanced on known and unknown confounders thus it is not necessary

to control for confounding at baseline. In the observational analog, whether target trial or target cohort, regression adjustment can be used to adjust for confounding at baseline. In the target cohort and target trial approach, the analysis plan is therefore inherently linked to step 3 in which treatment assignment and the method of achieving conditional exchangeability is defined. In our target cohort, we use conditional logistic regression adjusted for measured confounders and conditional on the time-matching of each mini-trial group to estimate the odds ratio, which approximates the incidence density ratio, to examine the causal contrast of interest [26]. This analysis compares the frequency of adverse GI events in initiators and non-initiators of semaglutide. We must control for a sufficient set of confounding variables to achieve conditional exchangeability and attempt to emulate randomization [18]. The effect estimate from an analysis using the target trial or target cohort approach has a causal interpretation if the formal identifiability assumptions for causal inference are satisfied—including conditional exchangeability, positivity,



**Fig. 3** Additional illustration of target cohort approach as a series of mini-trials. This figure describes three additional mini trials [14–16] including cases and controls sampled from the study cohort. For ease of explanation, the participant ID number for each case or control is included in parentheses in the figure (e.g., Case (010)). In a case-control study with incidence density sampling, a control may go on to be a case at a later point in time or a control may be selected as a control again at a later point in time. As illustrated, the participant with study

ID 909 serves as a control in mini trial 15 at 5.5 months but then goes on to experience the outcome of interest at 9 months. They contribute 5.5 mo of person time on Semaglutide ( $A=1$ ) to mini trial 15 and 9 mo of person time on Semaglutide ( $A=1$ ) to mini trial 16. Further, the participant with study ID 140 was selected to be a control in mini trial 16, at 9 mo, and then also selected to be a control in trial 17 at mo 11. They contribute 9 mo of unexposed person time ( $A=0$ ) to mini trial 16 and 11 mo of unexposed person time ( $A=0$ ) to mini trial 17

**Table 3** Description of follow-up period and measured outcomes of interest

	RCT	Target trial*	Target cohort**
<b>Outcomes of interest</b>	Adverse GI events assessed by physician diagnosis: • biliary disease (including cholecystitis, cholelithiasis, and choledocholithiasis) • pancreatitis • bowel obstruction	Same adverse GI events, measured via ICD codes recorded in electronic health record	Same as target trial. Case status defined by incident adverse GI event, controls are individuals in the risk set not experiencing a GI event at that time point.
<b>Causal contrast</b>	Intention to treat effect	Observational analog of intention to treat effect	Observational analog of intention to treat effect
<b>Analysis plan</b>	Cox proportional hazards model comparing time to event among individuals randomly assigned to receive Semaglutide ( $A=1$ ) vs. placebo ( $A=0$ ).	Cox proportional hazards analysis comparing time to event among individuals who initiate Semaglutide ( $A=1$ ) vs. non-initiators ( $A=0$ ). Adjustment for baseline covariates via regression model	Controls are matched to cases on follow up time and must be analyzed as pair-matched data (e.g., multivariable adjusted conditional logistic regression) -Odds ratio interpreted as an incidence rate ratio



consistency, no interference, and no model misspecification [18]. Of note, a conditional logistic regression model is appropriate in the present target cohort analysis, but many other analytic approaches from the causal inference literature can be applied to case-control studies with incidence density sampling [60, 61]. As discussed above, when the sample fractions for the cases and controls from the study base are known, the inverse of these fractions can be incorporated as weights in most statistical analyses to attain population-representative parameters [62].

## Discussion

In this manuscript, we present a novel framework for estimating causal effects from case-control studies using an adaptation of the target trial framework [4, 5]. We conceptualize nested case-control studies with incidence density sampling as a series of mini-trials within a study cohort. The goal of the target cohort approach is to help investigators make explicit decisions about the eligibility criteria, treatment and sampling strategies, treatment assignment (randomization), follow-up period, outcomes of interest, causal contrast, and analysis plan when attempting to draw causal inferences from a nested case-control study with incidence density sampling. Theoretically, if using the same eligible study cohort, the target cohort approach will produce a causal effect estimate that is equivalent to the effect estimate from a target trial, apart from sampling and estimation error.

Case-control studies have sometimes been described in the epidemiologic literature as an inferior, or less valid, type of study design. Within the hierarchy of the evidence-based medicine pyramid, case-control studies are considered to provide lower quality evidence than cohort studies [63]. In the 1970s, Feinstein [64] described case-control studies as “trohoc” studies (cohort spelled backwards) reflecting an apparently backward approach to conducting research. However, over the past 50 years, there has been considerable methodological work demonstrating the advantages of case-control studies [1, 3, 65]. In part, the methodological evolution of case control studies has been driven by the understanding that case control studies are not simply ‘backwards’ cohort studies. Poole (1999) emphasized the concept of the “trohoc” fallacy, which manifests in two ways: 1) misplaced concern about the comparability of cases and controls, when the real concern is the comparability of controls and the study base and 2) mistaken assumption that the control group must be healthy and free of disease [17]. Case control studies are better conceptualized as an efficient study design involving sampling from a cohort, whether it is a clearly defined cohort or a hypothetical cohort. Others have described the case-control design as one in which

information is missing at random for individuals in the population that are not included as cases or controls [66]. Many of the criticisms of case-control studies apply to traditional, cumulative case-control studies but not nested case-control studies using incidence density sampling [16]. For instance, in the cumulative case-control design, controls are sampled from individuals who remain non-cases when the study ends. This design makes it very difficult, if not impossible, to disentangle the effect of exposure on outcome from other factors such as loss-to-follow-up or selective survival [1]. However, using an incidence density sampling approach, the sampling probability for each control is proportional to the amount of time spent at risk for the outcome of interest and the sampling probability is independent of exposure [1]. Case-control studies with incidence density sampling can therefore be used to estimate causal effects with comparable validity to cohort studies.

Case-control studies with incidence density sampling are also an exceptionally efficient study design. Efficiency is defined by the amount of information a study produces relative to the size or cost [1]. A case-control study is more efficient than a cohort design because it includes all of the cases and a probability sample of controls that is representative of the study base that gave rise to the cases. This design over-represents cases relative to the total size of the cohort [1]. For a fixed number of cases, precision of a case-control study can be improved by recruiting a greater number of controls per case. With the increasing availability of secondary data sources, including “big data”, there are more opportunities to conduct case-control studies than ever before [65, 67, 68]. Case-control studies are a particularly efficient use of big data resources because they are efficient, in terms of cost, timing, and statistical efficiency. Although large secondary data sources may also provide increased opportunity for cohort studies as well, sampling non-cases from the study base can be a more efficient design.

Leveraging conceptual links between randomized trials, cohort studies, and case-control studies is key to estimating causal effects from observational data sources. By integrating foundational case-control concepts with modern approaches to causal inference, we aim to promote the use of case-control studies with incidence density sampling to estimate causal effects. Our goal in developing the target cohort approach is to improve the rigor and reproducibility of causal analyses from case-control study designs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40471-024-00353-3>.

**Author Contributions** H.R.B, E.C.M, and R.W.P. all contributed to the conceptualization of this manuscript. H.R.B wrote the initial draft of the manuscript text and received substantial input in the editing process from E.C.M and R.W.P.

**Data Availability** No datasets were generated or analysed during the current study.

## Declarations

**Human and Animal Rights and Informed Consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

**Competing Interests** The authors declare no competing interests.

## References

- Lash TL, J.VanderWeele T, Haneuse S, Rothman KJ. Modern Epidemiology. 4th Edition ed. Philadelphia, PA: Wolters Kluwer; 2021.
- Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Emulating a target trial in case-control designs: an application to statins and colorectal cancer. *Int J Epidemiol.* 2020;49(5):1637–46. <https://doi.org/10.1093/ije/dyaa144>.
- Vandenbroucke JP, Pearce N. Case-control studies: basic concepts. *Int J Epidemiol.* 2012;41(5):1480–9. <https://doi.org/10.1093/ije/dys147>.
- 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):758–64. <https://doi.org/10.1093/aje/kwv254>.
- 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;79:70–5. <https://doi.org/10.1016/j.jclinepi.2016.04.014>.
- Hernán MA, Wang W, Leaf DE. Target Trial Emulation: a Framework for Causal Inference from Observational Data. *JAMA.* 2022;328(24):2446–7. <https://doi.org/10.1001/jama.2022.21383>.
- Hernández-Díaz S, Huybrechts KF, Chiu YH, Yland JJ, Bateman BT, Hernán MA. Emulating a target trial of interventions initiated during pregnancy with Healthcare databases: the Example of COVID-19 vaccination. *Epidemiology.* 2023;34(2):238–46. <https://doi.org/10.1097/ede.0000000000001562>.
- Gupta S, Wang W, Hayek SS, et al. Association between Early Treatment with Tocilizumab and Mortality among critically ill patients with COVID-19. *JAMA Intern Med.* 2021;181(1):41–51. <https://doi.org/10.1001/jamainternmed.2020.6252>.
- 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):690–9. <https://doi.org/10.1097/ede.0000000000001626>.
- Barda N, Dagan N, Cohen C, et al. Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: an observational study. *Lancet.* 2021;398(10316):2093–100. [https://doi.org/10.1016/S0140-6736\(21\)02249-2](https://doi.org/10.1016/S0140-6736(21)02249-2).
- Lash T, VanderWeele Tyler J, Haneuse S, Rothman K. Modern epidemiology. 4th ed. Philadelphia: Wolters Kluwer; 2021.
- Labrecque JA, Hunink MMG, Ikram MA, Ikram MK. Do Case-Control studies Always Estimate odds Ratios? *Am J Epidemiol.* 2021;190(2):318–21. <https://doi.org/10.1093/aje/kwaa167>.
- Knol MJ, Vandenbroucke JP, Scott P, Egger M. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am J Epidemiol.* 2008;168(9):1073–81. <https://doi.org/10.1093/aje/kwn217>.
- Pearce N. What does the odds ratio estimate in a case-control study? *Int J Epidemiol.* 1993;22(6):1189–92. <https://doi.org/10.1093/ije/22.6.1189>.
- Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet.* 2002;359(9303):341–5. [https://doi.org/10.1016/S0140-6736\(02\)07500-1](https://doi.org/10.1016/S0140-6736(02)07500-1).
- Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of Controls in Case-Control studies: I. principles. *Am J Epidemiol.* 1992;135(9):1019–28. <https://doi.org/10.1093/oxfordjournals.aje.a116396>.
- Poole C. Controls who experienced hypothetical causal intermediates should not be excluded from case-control studies. *Am J Epidemiol.* 1999;150(6):547–51. <https://doi.org/10.1093/oxfordjournals.aje.a010051>.
- Hernán MA, Robins JM. Causal inference: what if? Boca Raton: Chapman & Hall/CRC; 2023.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol.* 2002;31(2):422–9. <https://doi.org/10.1093/intjpid/31.2.422>.
- Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health.* 2001;22:189–212. <https://doi.org/10.1146/annurev.publhealth.22.1.189>.
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol.* 1992;135(9):1029–41. <https://doi.org/10.1093/oxfordjournals.aje.a116397>.
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol.* 1992;135(9):1042–50. <https://doi.org/10.1093/oxfordjournals.aje.a116398>.
- Hernán MA. Beyond exchangeability: the other conditions for causal inference in medical research. *Stat Methods Med Res.* 2012;21(1):3–5. <https://doi.org/10.1177/0962280211398037>.
- Robins JM, Greenland S. Identifiability and exchangeability for Direct and Indirect effects. *Epidemiology.* 1992;3(2):143–55.
- Cole SR, Hernán MA. Fallibility in estimating direct effects. *Int J Epidemiol.* 2002;31(1):163–5. <https://doi.org/10.1093/ije/31.1.163>.
- Lincoff AM, Brown-Frandsen K, Colhoun HM, et al. Semaglutide and Cardiovascular outcomes in obesity without diabetes. *N Engl J Med.* 2023;389(24):2221–32. <https://doi.org/10.1056/NEJMoa2307563>.
- Garvey WT, Batterham RL, Bhatta M, et al. Two-year effects of semaglutide in adults with overweight or obesity: the STEP 5 trial. *Nat Med.* 2022;28(10):2083–91. <https://doi.org/10.1038/s41591-022-02026-4>.
- Jastreboff AM, Aronne LJ, Ahmad NN, et al. Tirzepatide once Weekly for the treatment of obesity. *N Engl J Med.* 2022;387(3):205–16. <https://doi.org/10.1056/NEJMoa2206038>.
- Saxena AR, Frias JP, Brown LS, et al. Efficacy and safety of oral small Molecule Glucagon-Like peptide 1 receptor agonist Danuglipron for Glycemic Control among patients with type 2 diabetes: a Randomized Clinical Trial. *JAMA Netw Open.* 2023;6(5):e2314493–e. <https://doi.org/10.1001/jamanetworkopen.2023.14493>.
- Wilding JPH, Batterham RL, Calanna S, et al. Once-weekly semaglutide in adults with overweight or obesity. *N Engl J Med.* 2021;384(11):989–1002. <https://doi.org/10.1056/NEJMoa2032183>.
- Sodhi M, Rezaeianzadeh R, Kezouh A, Etmiman M. Risk of gastrointestinal adverse events Associated with Glucagon-Like Peptide-1 receptor agonists for weight loss. *JAMA.* 2023;330(18):1795–7. <https://doi.org/10.1001/jama.2023.19574>.
- Wharton S, Davies M, Dicker D, et al. Managing the gastrointestinal side effects of GLP-1 receptor agonists in

- obesity: recommendations for clinical practice. *Postgrad Med.* 2022;134(1):14–9. <https://doi.org/10.1080/00325481.2021.2002616>.
33. Vandembroucke JP. What is the best evidence for determining harms of medical treatment? *CMAJ.* 2006;174(5):645–6. <https://doi.org/10.1503/cmaj.051484>.
  34. Goetghebeur E, le Cessie S, De Stavola B, Moodie EE, Waernbaum I, initiative obottgClotS. Formulating causal questions and principled statistical answers. *Stat Med.* 2020;39(30):4922–48. <https://doi.org/10.1002/sim.8741>.
  35. Katki HA, Berndt SI, Machiela MJ, et al. Increase in power by obtaining 10 or more controls per case when type-I error is small in large-scale association studies. *BMC Med Res Methodol.* 2023;23(1):153. <https://doi.org/10.1186/s12874-023-01973-x>.
  36. Lund JL, Richardson DB, Stürmer T. The active comparator, new user study design in pharmacoepidemiology: historical foundations and contemporary application. *Curr Epidemiol Rep.* 2015;2(4):221–8. <https://doi.org/10.1007/s40471-015-0053-5>.
  37. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiol (Cambridge Mass).* 2008;19(6):766–79. <https://doi.org/10.1097/EDE.0b013e3181875e61>.
  38. Huitfeldt A, Hernan MA, Kalager M, Robins JM. EGEMS (Wash DC). 2016;4(1):1234. <https://doi.org/10.13063/2327-9214.1234>. Comparative Effectiveness Research Using Observational Data: Active Comparators to Emulate Target Trials with Inactive Comparators.
  39. 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.* 2010;340:b5087. <https://doi.org/10.1136/bmj.b5087>.
  40. Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA. Electronic medical records can be used to emulate target trials of sustained treatment strategies. *J Clin Epidemiol.* 2018;96:12–22. <https://doi.org/10.1016/j.jclinepi.2017.11.021>.
  41. Månsson R, Joffe MM, Sun W, Hennessy S. On the Estimation and Use of Propensity scores in Case-Control and Case-Cohort studies. *Am J Epidemiol.* 2007;166(3):332–9. <https://doi.org/10.1093/aje/kwm069>.
  42. Austin PC. An introduction to Propensity score methods for reducing the effects of confounding in Observational studies. *Multivar Behav Res.* 2011;46(3):399–424. <https://doi.org/10.1080/00273171.2011.568786>.
  43. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med.* 2015;34(28):3661–79. <https://doi.org/10.1002/sim.6607>.
  44. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med.* 2009;28(25):3083–107. <https://doi.org/10.1002/sim.3697>.
  45. Matthey EC, Farkas K, Skeem J, Ahern J. Exposure to Community Violence and Self-harm in California: a Multilevel, Population-based, Case-Control Study. *Epidemiology.* 2018;29(5):697–706. <https://doi.org/10.1097/ede.0000000000000872>.
  46. Rose S, Laan MJ. Why Match? Investigating matched case-control study designs with Causal Effect Estimation. *Int J Biostatistics.* 2009;5(1). <https://doi.org/10.2202/1557-4679.1127>.
  47. Stürmer T, Brenner H. Degree of matching and Gain in Power and Efficiency in Case-Control studies. *Epidemiology.* 2001;12(1):101–8.
  48. Zhu Y, Hubbard RA, Chubak J, Roy J, Mitra N. Core concepts in pharmacoepidemiology: violations of the positivity assumption in the causal analysis of observational data: consequences and statistical approaches. *Pharmacoepidemiol Drug Saf.* 2021;30(11):1471–85. <https://doi.org/10.1002/pds.5338>.
  49. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol.* 2013;42(3):860–9. <https://doi.org/10.1093/ije/dyt083>.
  50. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168(6):656–64. <https://doi.org/10.1093/aje/kwn164>.
  51. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol.* 2011;173(7):731–8. <https://doi.org/10.1093/aje/kwq472>.
  52. Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol.* 2017;46(2):756–62. <https://doi.org/10.1093/ije/dyw323>.
  53. Jiao T, Platt RW, Douros A, Filion KB. Use of a statistical adaptive treatment Strategy Approach for emulating randomized controlled trials using Observational Data: the Example of blood-pressure control strategies for the Prevention of Cardiovascular events among individuals with hypertension at High Cardiovascular Risk. *Am J Epidemiol.* 2023. <https://doi.org/10.1093/aje/kwad091>.
  54. Penning de Vries L, Groenwold BB. Identification of causal effects in case-control studies. *BMC Med Res Methodol.* 2022;22(1):7. <https://doi.org/10.1186/s12874-021-01484-7>.
  55. Calvert M, Blazeby J, Altman DG, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO Extension. *JAMA.* 2013;309(8):814–22. <https://doi.org/10.1001/jama.2013.879>.
  56. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT Statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med.* 2001;134(8):663–94. <https://doi.org/10.7326/0003-4819-134-8-200104170-00012>.
  57. Miettinen OS. The case-control study: valid selection of subjects. *J Chronic Dis.* 1985;38(7):543–8. [https://doi.org/10.1016/0021-9681\(85\)90039-6](https://doi.org/10.1016/0021-9681(85)90039-6).
  58. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials.* 2012;9(1):48–55. <https://doi.org/10.1177/1740774511420743>.
  59. 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):2191–225. <https://doi.org/10.1002/sim.9718>.
  60. Rose S, van der Laan M. A double robust approach to causal effects in case-control studies. *Am J Epidemiol.* 2014;179(6):663–9. <https://doi.org/10.1093/aje/kwt318>.
  61. VanderWeele TJ, Vansteelandt S. A weighting approach to causal effects and additive interaction in case-control studies: marginal structural linear odds models. *Am J Epidemiol.* 2011;174(10):1197–203. <https://doi.org/10.1093/aje/kwr334>.
  62. van der Laan MJ. Estimation based on case-control designs with known prevalence probability. *Int J Biostat.* 2008;4(1):Article17. <https://doi.org/10.2202/1557-4679.1114>.
  63. Guyatt GH, Haynes R, Jaeschke RZ, et al. Users' guides to the medical literature: Xxv. Evidence-based medicine: principles for applying the users' guides to patient care. *JAMA.* 2000;284(10):1290–6. <https://doi.org/10.1001/jama.284.10.1290>.
  64. Feinstein AR. Clinical biostatistics; xx. The epidemiologic trochoc, the ablative risk ratio, and 'retrospective' research. *Clin Pharmacol Ther.* 1973;14(2):291–307. <https://doi.org/10.1002/cpt1973142291>.
  65. Suissa S. The Quasi-cohort Approach in Pharmacoepidemiology: upgrading thenested case-control. *Epidemiology.* 2015;26(2):242–6. <https://doi.org/10.1097/ede.0000000000000221>.

66. Wacholder S. The case-control study as data missing by design: estimating risk differences. *Epidemiology*. 1996;7(2):144–50. <https://doi.org/10.1097/00001648-199603000-00007>.
  67. Mooney SJ, Garber MD, Sampling, Sampling Frames in Big Data Epidemiology. *Curr Epidemiol Rep*. 2019;6(1):14–22. <https://doi.org/10.1007/s40471-019-0179-y>.
  68. Li CX, Matthey EC, Rowe C, Bradshaw PT, Ahern J. Conducting density-sampled case-control studies using survey data with complex sampling designs: a simulation study. *Ann Epidemiol*. 2022;65:109–15. <https://doi.org/10.1016/j.annepidem.2021.06.019>.
3. Poole C. Controls who experienced hypothetical causal intermediates should not be excluded from case-control studies. *American journal of epidemiology*. 1999;150(6):547–51. <https://doi.org/10.1093/oxfordjournals.aje.a010051>. Rationale: This manuscript describes the concept of a case-control study being nested in a cohort.

## Key References

1. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *American journal of epidemiology*. 2016;183(8):758–64. <https://doi.org/10.1093/aje/kwv254>. Rationale: This manuscript presents the underlying framework for the target trial approach.
2. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA. Emulating a target trial in case-control designs: an

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.