




Use of causal inference methods in case–control studies: a methodology review

Miceline Mésidor^{*1,2} , Mengting Xu³, Awa Diop^{4,5}, Canisius Fantodji¹, Marie-Élise Parent^{1,3} , Alexander Keil⁶ 

¹Institut national de la recherche scientifique—Centre Armand Frappier Santé-Biotechnologie, Laval, Canada

²Faculté de pharmacie, Université de Montréal, Montréal, Canada

³École de santé publique, Université de Montréal, Montréal, Canada

⁴StatHarbor Analytics, Montréal, Canada

⁵Real World Data Science, Biopharmaceuticals Medical Evidence, AstraZeneca, Mississauga, Canada

⁶Occupational and Environmental Epidemiology Branch, National Cancer Institute, Rockville, MD, United States

*Corresponding author: Miceline Mésidor, Institut national de la recherche scientifique – Centre Armand Frappier Santé-Biotechnologie, 531 des Prairies, Laval, QC H7V 1B7 (miceline.mesidor@inrs.ca)

Abstract

The use of causal inference methods in cohort studies has increased considerably in recent years. However, their use has been limited in case–control studies. This report aimed at providing a detailed review of causal inference methods used in case–control studies and to review and examine their applications in previous studies. Several methods have been used to facilitate causal inference in case–control studies, including intercept-adjustment, propensity scores, and weight-based and doubly robust estimators. We used the Medical Literature Analysis and Retrieval System Online database to identify original peer-reviewed case–control studies conducted from March 2014 to March 2024 that applied these methods. We identified 418 studies, 23 of which met the inclusion criteria. Most studies involved case–control matching (individual or frequency) and included incident cases. The covariate-conditional odds ratio was the most frequently reported estimated parameter. Sixty-five percent of included studies considered an adjustment for sampling bias, most often using inverse-probability of observation weighting and case–control targeted maximum likelihood approaches. We are still in the early stages of development and application of causal inference methods for case–control studies. Their implementation and new techniques to address time-varying confounding can improve the validity of study findings and should be encouraged.

Key words: causal inference; case–control studies; sampling selection bias; marginal causal effects; review.

Introduction

Causal inference methods are a loose collection of statistical approaches that have been shown, under varying sets of assumptions, to directly estimate causal effects. In the potential outcomes framework, a causal effect can be expressed as the contrast in outcomes that would be expected between two (hypothetical) copies of the same population that are identical except that they differ in the exposure of interest. Randomized controlled trials (RCTs) are the gold-standard method for investigating many causal questions, as weaker assumptions are required for causal interpretation of a given estimate due to the study design and execution. However, for many scenarios, RCTs are not ethical or lack generalizability to target populations of interest,¹ which highlights the importance of methods to facilitate causal inference from observational studies. Although the last few decades have seen a meteoric rise in the development and application of causal inference methods for observational studies, the focus has primarily been on cohort studies. However, cohort studies can be impractical for studying rare outcomes or when exposure assays are costly.² In these scenarios, case–control study designs are often used for cost-efficiency.

By far, the most commonly used method for estimating exposure effects in case–control studies is logistic regression, in which case the case–control odds ratio (OR) is often the de facto estimand of interest.^{2,3} The case–control OR is commonly used because, under cumulative sampling that is nested in a cohort, the case–control OR estimates the OR that would be obtained in a cohort study.⁴ Thus, in some situations, case–control studies and cohorts give consistent inference on ORs, and if causal assumptions are met, such ORs can represent causal ORs.⁵ However, logistic regression parameters are interpretable as (at best) conditional causal effects, given a measured set of adjustment variables. Causal methods are useful for estimating marginal causal parameters that correspond to population average effects like the marginal risk differences, which equal the average of individual-level causal effects in the entire population of interest and have a strong implication for public health.⁶ Thus, causal inference methods can be useful both for addressing bias as well as estimating parameters for specific populations, ideas that have previously been demarcated as internal and external validity.

Several causal inference methods for case–control studies have been introduced^{7–9} but are rarely used in practice.¹⁰ To facilitate

Received: January 17, 2025. Accepted: August 13, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

broader adoption and clarify their purpose, we review available methods proposed for causal inference in case-control studies and their associated assumptions. We discuss applications in the literature, provide code, and weigh the strengths and weaknesses of each method to identify gaps and provide recommendations. Brief introductions of causal inference and design of case-control studies are presented in the Supplementary material (Appendix S1).

Methods

Since Rubin's framing of causal inference in observational studies through potential outcomes in 1974,¹¹ many methods have been developed under this framework for cohort studies. Commonly used methods included propensity scores/inverse probability of treatment weighting (IPTW), g-computation, and doubly robust methods (such as targeted maximum likelihood estimation). Each of these methods focuses on statistical control of measured confounders and ensuring results are relevant to a specific target population. For conciseness, we do not consider other approaches such as effect decomposition (causal mediation analysis),^{12,13} negative controls,¹⁴ or instrumental variable methods¹⁵ like Mendelian randomization that use alternative assumption sets.

For case-control studies, the majority of these methods have been adapted to control confounding (internal validity) while also addressing the issue of sampling bias (external validity). By "sampling bias," we mean bias that occurs because non-cases are under-sampled from the target population, relative to cases. Sampling bias manifests in 2 ways: (1) prevalence in the study population is high relative to the target population ("outcome sampling bias") and (2) covariates that are associated with the outcome will have different distributions in the study population relative to the target population ("covariate sampling bias"). Here, we review methods that typically fall under the umbrella term "causal inference approaches" or "g-methods," including propensity scores,^{16,17} case-control weighted estimators (CCW-MLE),^{8,18} and case-control targeted maximum likelihood estimators (CCW-TMLE).^{8,18-20} In addition, we review the more classical method of intercept-adjusted logistic regression,^{21,22} which can also be used to address sampling bias while addressing confounding, such that this approach falls under the "outcome regression" or "g-computation" family of causal inference approaches. More technical illustrations of sampling bias can be found elsewhere.^{7,8,10}

The causal inference methods we review here have been developed primarily for cumulative incidence sampling, where cases and controls (non-cases) are selected at the end of follow-up. Some approaches also work with case-cohort designs where non-cases are replaced with a sample (at baseline) from a cohort within which the cases are nested.²³ Incidence density sampling, in which controls are selected from among those at risk at the time each case arises, is briefly considered in the discussion.

In the following, we consider case-control data in which we have sampled cases and controls in the target population (Pop). For simplicity, we assume that the distribution of the observed outcome *Y* can be described using a logistic regression model:

$$\log\left(\frac{P(Y = 1|\mathbf{X}, Pop)}{P(Y = 0|\mathbf{X}, Pop)}\right) = \beta_0 + \sum_{j=1}^p \beta_j X_j$$

where $\mathbf{X} = (X_1, \dots, X_p)$ is a vector of *p* covariates (which includes the exposure *A*) sufficient to control confounding; β_0 and β_j for $j = 1, \dots, p$ are the log-OR parameters to be estimated.

Intercept-adjusted logistic regression

A simple approach to address outcome sampling bias has long been known for unmatched case-control studies,^{22,24,25} which allows the estimation of absolute risk. It involves adding an offset term c_0 to the standard logistic regression.^{22,24,25} This is equal to the logarithm of the sampling OR:

$$\log(c_0) = \log\left(\frac{\frac{P(Y=1|Pop)}{P(Y=0|Pop)}}{\frac{P(Y=1|Study)}{P(Y=0|Study)}}\right)$$

where $P(Y = 1|Pop)$ and $P(Y = 1|Study)$ represent the prevalence of the outcome in the target population and the case-control study data, respectively.

This approach, which requires knowledge of the population prevalence of the outcome, makes it possible to estimate some population-level causal effects.^{18,26} Matching can also be accommodated using stratum-specific sampling ORs.²⁵ Given sufficient control of confounders, the intercept-adjusted logistic model will estimate the conditional causal OR (an OR for a group of individuals with a specific set of covariates). It can be used to estimate marginal causal effects on other scales (eg, risk difference) but not for the target population due to residual covariate sampling bias.

To estimate marginal effects in the target population, intercept adjustment must also be accompanied by weighting or standardization to the covariate distribution of the target population. This holds because effect measures vary over the covariate distribution in the study, which is subject to sampling bias except when the covariates are not associated with the outcome. This issue also arises when oversampling certain groups in case-control studies, for example, if racial or ethnic minorities are over-sampled to power the estimation of disparities. In matched studies, one must weight or standardize across the distribution of matching variables.^{18,26} Case-control sample weighting can address this issue, as we discuss below.

Propensity scores

While logistic regression addresses confounding through regression adjustment, the propensity score method, a widely used set of methods, relies on using a propensity score (a covariate-conditional probability of exposure) for addressing confounding.¹⁷ To control confounding, different propensity score methods can be used: propensity score matching, IPTW using the propensity score, stratification on the propensity score, and covariate adjustment using the propensity score.^{16,17} These methods vary with respect to whether and how sampling bias is addressed.

Briefly, the propensity score is the probability of being exposed, given other covariates in the population:

$$\pi = P(A = 1|\mathbf{Z}, Pop)$$

where \mathbf{Z} is the covariate vector \mathbf{X} excluding exposure *A*. For estimating average effects of binary exposures, an inverse probability weight can be defined for each individual as follows:

$$w_i = \begin{cases} \frac{1}{P(A_i=1|\mathbf{Z}_i, Pop)}, & \text{if the individual is exposed} \\ \frac{1}{1-P(A_i=1|\mathbf{Z}_i, Pop)}, & \text{if the individual is unexposed} \end{cases}$$

Thorough introductions to propensity scores can be found elsewhere.^{16,27} Crucially, for case-control studies, the sample estimate of the propensity score $P(A = 1|\mathbf{Z}, Study)$ is subject to sampling

bias because, for example, if the exposure increases the probability of the outcome, given Z , exposed individuals will be more common in the case-control study relative to the cohort study. Generally, propensity score methods (and other causal inference methods) can rely on bootstrapping to generate standard errors, but specific estimators may allow less computationally intensive approaches.

Multiple ways of addressing sampling bias in the propensity score have been proposed for case-control studies.²⁸⁻³¹ Under rare outcomes, Robins proposed that the IPTW can be used in the context of marginal structural models to estimate marginal ORs in the target population²⁸ with the propensity score estimated only among controls, thus addressing covariate sampling bias but not outcome sampling bias. The weight is then applied to all individuals and allows that the population risk of the outcome can be unknown (aside from knowing that it is rare).²⁸ Mansson introduced several novel approaches to use of propensity scores in case-control studies, which involved (1) estimating propensity scores in a subset of the study that included all controls and a subset of cases and (2) including case status in the propensity score model and estimating propensity scores for all individuals as though they had been controls.²⁹ Similar to Robins' approach, both approaches are restricted to the estimation of the marginal OR because they address covariate sampling bias only through the propensity score estimates. The approach proposed by Zhu et al. estimates the propensity score model among cases and then matches controls on the predicted propensity scores before estimating conditional log-ORs.³¹ However, this approach does not yield valid estimates of any marginal parameter, so it is only tangentially related to causal inference vis-a-vis adjustment for confounding or sampling bias.

G-computation

G-computation, as a complement to propensity score methods that model exposures, consists of the modeling outcomes and predicting potential outcomes under different exposure conditions.^{32,33} For time-fixed exposures, the main steps of g-computation are:

- 1) Regress the exposure and the covariates on the outcome;
- 2) Using the regression model, predict outcomes for individuals under each exposure condition;
- 3) Estimate average causal effects by averaging the predicted outcomes across the study population.

The third step is equivalent to standardizing over the covariate distribution of the study population, which yields study-population-averaged effects. Confounding and sampling bias are addressed in steps 1 and 3 (described below). If the study population is a representative sample of the target population, then this estimates a population average (marginal) causal effect. For time-varying exposures with time-varying confounders impacted by prior exposures, the algorithm is more complicated (eg,^{34,35}).

The case-control weighted maximum likelihood estimator (CCW-MLE) is useful as a step in g-computation. CCW-MLE uses weighting similar to one of the methods demonstrated by Mansson et al.²⁹ This approach aligns with Miettinen's standardization approach³⁶ as applied by Newman.³⁰

A weighted logistic regression is preferably used in the first step with $P(Y = 1)$ and $\frac{P(Y=0)}{J}$ (which are assumed to be known from the target population of the case-control study participants and J is the ratio of controls to cases), respectively, the weights for cases and controls for unmatched studies.^{10,20} Intuitively, the weighted proportion of cases (and weighted distribution of

covariates) in the case-control data is then equal to the marginal proportion of cases (and marginal distribution of covariates) in the target population, so that more general models (outside logistic regression) can be fitted to the data and marginal effects can be estimated. Then, g-computation steps 2 and 3 from above can then be followed, with the modification that step 3 uses case-control weighted averages to fully account for covariate sampling bias. Bootstrap¹⁹ and delta²⁶ methods can be used to obtain standard errors. Matching can also be accommodated.²⁰ This method allows the estimation of marginal effects (eg, ORs, relative risks, and risk differences).

Targeted maximum likelihood estimation

Targeted maximum likelihood estimation (TMLE) is a causal inference estimation method that can use models for the exposure and the outcome. TMLE is doubly robust, meaning it has the advantage of providing a consistent estimator if either model is consistent, and can have improved variance if both models are consistent.³⁷ The typical steps involve getting a first-stage estimate of the outcome risk (like g-computation) and then updating that estimate using a modification of the propensity score. TMLE and other doubly robust approaches are discussed thoroughly elsewhere.^{38,39}

Specifically, the steps to estimate the case-control weighted TMLE estimator are as follows⁸:

- 1) Define case-control weights for cases and controls as in CCW-MLE. In the following, we will define $q_0 \equiv P(Y = 1)$ (such that $1 - q_0 = P(Y = 0)$).
- 2) Fit a CCW-MLE or intercept-adjusted logistic regression model predicting the outcome based on the exposure and covariates and calculate the initial probabilities of the outcome for each individual under each exposure condition, denoted by $\hat{Q}(A, Z)$.
- 3) Fit a CCW-MLE logistic regression model for A to estimate propensity scores, denoted by $\hat{g}(A, Z)$.
- 4) Using the propensity scores, compute "clever covariates" for each individual. A clever covariate is a covariate that is used to adjust or update the prediction $\hat{Q}(A, Z)$ based on the estimated propensity scores. This allows the incorporation of information about how covariates relate to the exposure. For a risk difference, the clever covariate is univariate and is given by:

$$h(A, Z) \equiv \left(\frac{I(A = 1)}{\hat{\pi}(A = 1|Z)} - \frac{I(A = 0)}{\hat{\pi}(A = 0|Z)} \right)$$

where $I(A = a)$ is an indicator function that equals 1 when $A = a$ and zero otherwise. For relative risk and OR, the bivariate clever covariate vector is given as the set of disjoint indicators:

$$h(A, Z) \equiv \left(\frac{I(A = 1)}{\hat{\pi}(A = 1|Z)}, \frac{I(A = 0)}{\hat{\pi}(A = 0|Z)} \right)$$

- 5) Update the probabilities $\hat{Q}(A, Z)$ through a CCW-MLE logistic regression model with $h(A, Z)$ as a supplementary variable to estimate a fluctuation parameter ϵ by regressing the observed outcome on the clever covariate(s) in an intercept-free logistic model with the initial prediction $\hat{Q}(A, Z)$ used as an offset term. The initial predictions are then updated:

$$\hat{Q}_1(A, Z) = \ell^{-1} \left(\ell \left(\hat{Q}(A, Z) \right) + h(A, Z) \epsilon \right)$$

where ℓ and ℓ^{-1} are the logistic link function and its inverse.

- 6) Compute the targeted estimate of interest using the updated probabilities by averaging over the CCW-MLE distribution of the covariates. For example, the marginal causal risk ratio can be computed as follows:

$$\hat{\psi}_{RR} = \frac{\frac{1}{n} \sum_{i=1}^n \left[I(Y_i = 1) q_0 \hat{Q}_1(1, Z_i) + I(Y_i = 0) \frac{1-q_0}{J} \hat{Q}_1(1, Z_i^j) \right]}{\frac{1}{n} \sum_{i=1}^n \left[I(Y_i = 1) q_0 \hat{Q}_1(0, Z_i) + I(Y_i = 0) \frac{1-q_0}{J} \hat{Q}_1(0, Z_i^j) \right]}$$

where $I(Y_i = y)$ is an indicator function that equals 1 if $Y_i = y$, and 0 otherwise, and the summation is over the n participants comprising cases and controls.

Standard errors can be calculated using the influence curve¹⁹ or bootstrap-based methods. Modifications to TMLE can enhance power and stability.^{9,10} An alternative to TMLE and other doubly robust methods often employed in cohort studies is to use propensity score and outcome regression-based methods to estimate the same (or similar) quantities, estimates of which can then be contrasted to assess potential impacts of model specification. Case-control studies can (and have) employed similar strategies but must take care that sampling bias is handled similarly in each approach.

Review of publications using causal inference methods with a case-control design

Published studies of English language were identified using Medical Literature Analysis and Retrieval System Online (MEDLINE) database. We included original case-control studies published from March 12, 2014, to March 11, 2024. Mendelian randomization, genome-wide association studies, and mediation analyses were excluded. The search strategy used is presented in the [Appendix S2](#), but the goal was to identify case-control studies that explicitly mentioned causal inference.

We used Covidence (Veritas Health Innovation, Melbourne, Australia), an online software for the management of systematic reviews, to screen and review the studies. Two authors (M.M. and M.X.) independently screened and reviewed each abstract and full text. Disagreements were solved by discussion to reach a consensus. We then extracted data from all included studies according to: (1) type of case-control study, (2) recruitment of cases, (3) software used, (4) type of paper, (5) parameter estimated, (6) methods used, (7) discussion of causal assumptions, and (8) adjustment for sampling bias.

Our search yielded 418 studies, 47 of which were assessed for full-text eligibility ([Figure 1](#)). Of the 23 included studies, 5 were from the United States, 5 (including 1 re-analysis) from Iran, and 5 from Europe. [Table 1](#) presents a summary of the included studies. More than half of the studies were published from 2020 onward. Twenty-six percent of the studies were nested case-control studies, 65% were matched, and 74% included incident cases. Most studies (83%) were applied, as opposed to methodological studies (17%). The OR was the most frequently reported estimated parameter (74%). Over 50% of studies performed propensity methods/IPTW, while approximately 30% used CCW-TMLE. At least 1 causal assumption (eg, exchangeability, causal consistency, positivity, or variations of those terms) was explicitly discussed in only 56% of studies, and 65% mentioned adjustment for sampling bias from the case-control design.

Discussion

In this paper, we reviewed existing methods that have been proposed for explicitly estimating marginal causal effects in case-control studies in which cases and controls are selected at the end of follow-up. Over half of the studies included in our review were published after 2020 and addressed adjustments for sampling bias. As noted above, the methods discussed here can target estimands beyond conditional ORs, including marginal effects on other scales. However, in our literature search, most studies reported only conditional ORs.

Each of the statistical methods presented has differing strengths and limitations. Intercept-adjusted logistic regression is simple to use. However, in practice, it is rarely used in our identified applications, despite a long history in the literature. Further, it alone cannot be used to derive marginal causal effects because the covariate distribution in the study sample does not represent the covariate distribution in the target population. Simulation studies have shown that this approach performs well in the presence of good model specification²⁶ but is biased under misspecified models.⁸ These findings follow a larger pattern that outcome-regression-based, causal inference method (like g-computation) is sensitive to outcome model specification because non-linearity and non-additivity must be explicitly addressed. Another outcome-regression-based approach, CCW-MLE, is appealing because it is intuitive: sampling weights address outcome and covariate sampling bias by making the weighted case-control sample resemble a cohort study (thus permitting estimation of parameters outside of ORs), and weighted analysis is possible in many software packages. Once weights are applied, it becomes possible to perform additional analyses such as mediation analysis⁴⁰⁻⁴³ that are typically not straightforward within the standard case-control framework.

In contrast, propensity score-based methods do not need to specify a model for the outcome and may be less sensitive to model misspecification but, without modification, do not address outcome sampling bias and limit the scope of available target estimands. Propensity score approaches were the most used approaches identified in our review. We intuit this is not due to an interest in marginal ORs but rather due to their ease of use and an implicit association between “propensity score” and “causal inference,” even though causal inference is made possible via assumptions and study design, rather than particular methods.⁴⁴ Nevertheless, these methods can be biased under propensity model misspecification and can be sensitive to the positivity assumption because they do not allow interpolation. Simulation studies have also highlighted a potential modification artifact between exposure and estimated propensity scores in moderate sample sizes.²⁹

CCW-TMLE incorporates CCW-MLE and propensity scores and shares the benefits of both (availability of many target estimands, efficiency of outcome-regression-based approaches, and misspecification-robustness of propensity score-based approaches). It is also doubly robust, thus reducing concerns of misspecification. It is conceptually the most difficult method in our review, but the availability of an existing R package⁴⁵ has facilitated its applications. Like CCW-MLE, CCW-TMLE allows several parameters such as the risk difference to be estimated, unlike the intercept-logistic adjustment and IPTW (without further modifications). From a public health perspective, additive scale effects like risk differences are more relevant than ORs to inform population health burdens and to assess impacts of potential exposures.^{10,46,47}

Table 1. Summary of the studies in this review.

| First author, year | Country | Study design | Types of case-control studies | Recruitment of cases | Parameter estimated | Method used | Adjusted for sampling bias | Discussion of causal assumptions |
|------------------------------------|-------------|---------------------|-------------------------------|----------------------|---------------------|--|----------------------------|----------------------------------|
| Backenroth 2016 ⁵⁶ | USA | Case-control | NA | NA | OR | Lasso logistic regression | No | No |
| Menvielle 2016 ⁵⁷ | France | Case-control | Matched | Incident cases | OR | MSM; IPTW; mediation | Yes | Yes |
| Pearl 2016 ¹⁰ | USA | Nested case-control | NA | Incident cases | RR; RD | CCW-MLE/g-estimation; CCW-TMLE | Yes | Yes |
| Bedard 2017 ⁵⁵ | France | Nested case-control | Matched | Incident cases | OR | MSM | No | Yes |
| Delcoigne 2017 ⁵⁸ | Sweden | Nested case-control | Matched | Incident cases | HR; AR | IPTW | Yes | No |
| Persson 2017 ²⁶ | Sweden | Case-control | Matched | Incident cases | OR | Intercept-adjusted; CCW-MLE/g-estimation; CCW-TMLE | Yes | Yes |
| Abdollahpour 2018 ⁵⁹ | Iran | Case-control | Matched | Incident cases | OR | IPTW | Yes | Yes |
| Abdollahpour 2018 ⁶⁰ | Iran | Case-control | Unmatched | Incident cases | OR | IPTW; MBS | Yes | Yes |
| VelasquezGarcia 2018 ⁶¹ | Canada | Case-control | Matched | Incident cases | OR | IPTW | No | Yes |
| Zhu 2019 ³¹ | USA | Case-control | Matched | NA | OR | Propensity scores | Yes | Yes |
| Dickerman 2020 ⁴⁸ | UK | Nested case-control | Unmatched | Incident cases | OR | Targeted trial | No | No |
| Figueroa 2020 ⁶² | Costa Rica | Case-control | Matched | Incident cases | OR; RR; RD | CCW-TMLE; targeted learning framework | No | No |
| Lu 2020 ⁶³ | USA | Case-cohort | Unmatched | NA | OR | Propensity scores; IPTW | Yes | Yes |
| Abdollahpour 2021 ⁶⁴ | Iran | Case-control | Unmatched | Incident cases | OR; RR; RD; PAF | CCW-TMLE | Yes | No |
| Almasi-Hashiani 2021 ⁶⁵ | Iran | Case-control | Matched | Incident cases | RR; RD; PAF | CCW-TMLE | Yes | No |
| Basnet 2021 ⁶⁶ | Nepal | Case-control | Matched | NA | OR | IPTW; | No | No |
| Boska 2021 ⁶⁷ | USA | Case-control | Matched | Incident cases | OR | CCW-MLE/g-estimation | No | No |
| Akhtiar 2022 ⁶⁸ | Kuwait | Case-control | Matched | NA | OR; RR; RD; PPF | Propensity scores; MSM; | Yes | Yes |
| Yu 2022 ⁶⁹ | South Korea | Nested case-control | Matched | NA | OR | CCW-TMLE | No | No |
| Elduma 2023 ⁷⁰ | Sudan | Case-control | NA | Incident cases | RR; RD | Propensity scores | Yes | Yes |
| Malekifar 2023 ⁷¹ | Iran | Case-control | NA | Incident cases | RR; RD; PAF | CCW-TMLE | Yes | No |
| Caubet 2024 ⁷² | Canada | Case-control | Matched | Incident cases | OR | MBS | Yes | No |
| Takeuchi 2024 ⁵³ | Japan | Nested case-control | Matched | Incident cases | HR | IPTW; mediation | Yes | Yes |

Abbreviations: USA, United States of America; N/A, information not available; OR, odds ratio; MSM, marginal structural model; IPTW, inverse probability of treatment weighting; HR, hazard ratio; RR, relative risk; RD, risk difference; CCW-MLE, case control weighted maximum likelihood estimation; CCW-TMLE, case control weighted targeted maximum likelihood estimation; AR, absolute risk; MBS, model-based standardization (surrogate of g-estimation); PAF, population-attributable fraction; PPF, population-preventable fraction.

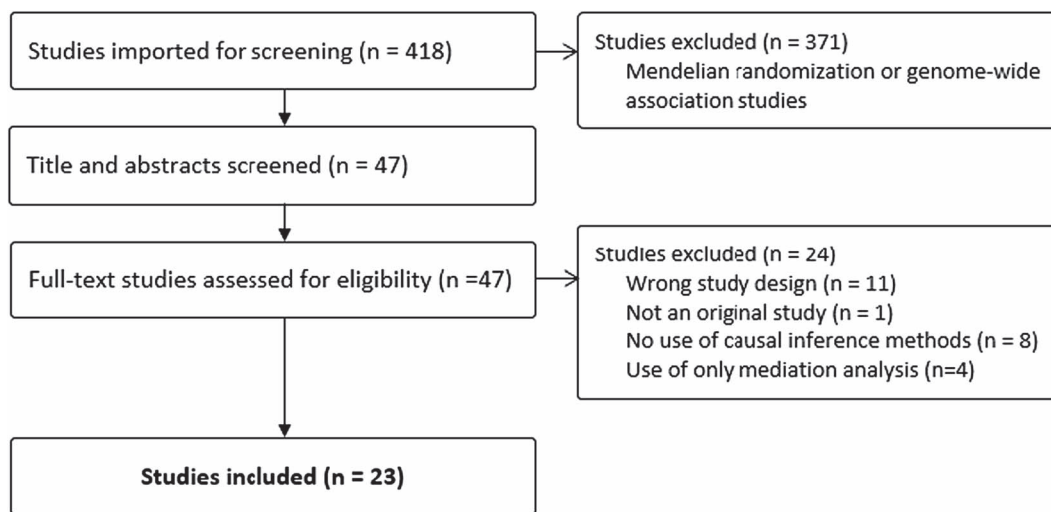


Figure 1. Preferred reporting items for systematic review and analysis (PRISMA) diagram.

We identified 1 study⁴⁸ with an expressed goal to emulate a hypothetical target trial using case-control data. Their findings indicated that appropriate case-control sampling yields estimates comparable to those derived from cohort studies. Careful consideration of the target trial framework⁴⁹ represents an exciting future direction for case-control studies, which may encourage further exploration of issues like recall bias or selection bias that are more prominent in retrospective studies.

Although the main stated purpose of many of the causal inference methods that have been developed for case-control studies is to adjust for sampling bias, more than a third of the included studies did not mention this specific purpose. In addition, causal assumptions were not discussed in around 44% of the studies. We note that many studies that used the propensity score method made no explicit reference to causal inference. Confounder control under the exchangeability assumption may have been the primary goal in these studies, where propensity-score-based methods may sometimes excel relative to alternative approaches. If one is interested in a causal OR (if it can be considered causal⁵⁰), then such methods may suffice. All methods we presented assume that the risk/prevalence of the disease in the population (and thus the sampling probabilities, marginally or within strata) is either known or is rare enough to be negligible. However, neither may be the case. Thus, if possible, sensitivity analyses should be performed using a range of plausible prevalence values,¹⁰ especially when the target parameter is a risk difference, which may be particularly sensitive to assumptions about the population risk. Most included studies were matched. Matching in case-control studies can help increase efficiency but does not remove confounding and may even introduce selection bias that needs to be suppressed by appropriate analytical methods.^{19,51,52}

Finally, we identified large gaps in existing methods that accommodate incidence density sampling. Using CCW-TMLE, van der Laan extends some methods for cumulative-incidence-sampled studies to the context of incidence density sampling.⁷ One issue raised is that such studies may be from open cohorts, which present difficulties for potential outcomes-based methods that have been developed within closed populations. A second issue is that, while CCW-TMLE methods can be applied on a time-specific basis (eg, the time a case is sampled), this works only when many cases arise simultaneously, which may not

occur in many sampling schemes (eg, sampling at exact age of diagnosis). Takeuchi et al. consider causal hazard ratios via marginal structural Cox models in a nested case-control study, where time-varying propensity score models (given time-varying confounders) are fit using inverse probability of sampling weights.^{53,54} Such approaches represent a promising avenue for more general causal inference approaches in case-control studies, and it may be beneficial to consider novel study designs that maximize efficiency when time-varying confounders may be of concern.

It is important to clarify that we focused on reviewing existing methods, rather than on a systematic review of all the papers applying these methods. In this sense, we have only considered the Medline database to give readers an idea of the application of methods in the literature. While we may have missed some relevant applications, it seems unlikely that our search strategy identified a non-representative sample of papers. Also, other methods of causal inference, such as Mendelian randomization, have been used in case-control studies but are not presented here because they rely on different sets of assumptions from the methods that we discussed herein.

In conclusion, this review aims to compile and review existing causal inference methods for case-control studies, fostering a better understanding to promote their application and spur further thinking about methodological gaps. For this purpose, we provide detailed R code for many of the reported methods (available at <https://github.com/alexpkeil1/Case-control-causal-review>). A key factor driving the growth of causal inference methods in cohort studies is the ability to address time-varying confounding for exposures that change over time. Nevertheless, our review identified only 3 studies^{48,53,55} that have confronted these challenges, highlighting a pressing need for further methodological advancements. As public health challenges become increasingly complex but no less costly to study, the ongoing development and application of causal inference methods for case-control studies will be essential for informing effective interventions and making evidence-based policy decisions.

Acknowledgments

The authors would like to thank Denis Talbot for his insightful comments on this review.

Supplementary material

Supplementary material is available at the American Journal of Epidemiology online.

Funding

None declared.

Conflict of interest

The authors declare no conflicts of interest.

References

- Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107-115. <https://doi.org/10.1093/aje/kwq084>
- Knol MJ, Vandenbroucke JP, Scott P, et al. What do case-control studies estimate? Survey of methods and assumptions in published case-control research. *Am J Epidemiol*. 2008;168(9):1073-1081. <https://doi.org/10.1093/aje/kwn217>
- Labrecque JA, Hunink MMG, Ikram MA, et al. Do case-control studies always estimate odds ratios? *Am J Epidemiol*. 2021;190(2):318-321. <https://doi.org/10.1093/aje/kwaa167>
- Miettinen O. Estimability and estimation in case-referent studies. *Am J Epidemiol*. 1976;103(2):226-235. <https://doi.org/10.1093/oxfordjournals.aje.a112220>
- Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*. 1987;125(5):761-768. <https://doi.org/10.1093/oxfordjournals.aje.a114593>
- Glass TA, Goodman SN, Hernan MA, et al. Causal inference in public health. *Annu Rev Public Health*. 2013;34(1):61-75. <https://doi.org/10.1146/annurev-publhealth-031811-124606>
- van der Laan MJ. Estimation based on case-control designs with known prevalence probability. *Int J Biostat*. 2008;4(1):Article 17. <https://doi.org/10.2202/1557-4679.1114>
- Rose S, van der Laan MJ. Simple optimal weighting of cases and controls in case-control studies. *Int J Biostat*. 2008;4(1):Article 19. <https://doi.org/10.2202/1557-4679.1115>
- Balzer L, Ahern J, Galea S, et al. Estimating effects with rare outcomes and high dimensional covariates: knowledge is power. *Epidemiol Methods*. 2016;5(1):1-18. <https://doi.org/10.1515/em-2014-0020>
- Pearl M, Balzer L, Ahern J. Targeted estimation of marginal absolute and relative associations in case-control data: an application in social epidemiology. *Epidemiology*. 2016;27(4):512-517. <https://doi.org/10.1097/EDE.0000000000000476>
- Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688-701. <https://doi.org/10.1037/h0037350>
- VanderWeele TJ, Tchetgen Tchetgen EJ. Mediation analysis with matched case-control study designs. *Am J Epidemiol*. 2016;183(9):869-870. <https://doi.org/10.1093/aje/kww038>
- Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. *Am J Epidemiol*. 2010;172(12):1339-1348. <https://doi.org/10.1093/aje/kwq332>
- Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-388. <https://doi.org/10.1097/EDE.0b013e3181d61eeb>
- Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology*. 2006;17(4):360-372. <https://doi.org/10.1097/01.ede.0000222409.00878.37>
- 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>
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41-55. <https://doi.org/10.1093/biomet/70.1.41>
- van der Laan MJ. Estimation Based on Case-Control Designs with Known Incidence Probability. UC Berkeley Division of Biostatistics Working Paper Series. 2008;Working Paper 234. <https://biostats.berpress.com/ucbbiostat/paper234>
- Rose S, Van Der Laan MJ. Why match? Investigating matched case-control study designs with causal effect estimation. *Int J Biostat*. 2009;5(1):Article 1. <https://doi.org/10.2202/1557-&break;4679.1127>
- Rose S, van der Laan M. A double robust approach to causal effects in case-control studies. *Am J Epidemiol*. 2014;179(6):663-669. <https://doi.org/10.1093/aje/kwt318>
- Cornfield J. A method of estimating comparative rates from clinical data; applications to cancer of the lung, breast, and cervix. *J Natl Cancer Inst*. 1951;11(6):1269-1275.
- Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979;66(3):403-411. <https://doi.org/10.2307/2335158>
- O'Brien KM, Lawrence KG, Keil AP. The case for case-cohort: an applied epidemiologist's guide to reframing case-cohort studies to improve usability and flexibility. *Epidemiology*. 2022;33(3):354-361. <https://doi.org/10.1097/ede.0000000000001469>
- Anderson JA. Separate sample logistic discrimination. *Biometrika*. 1972;59(1):19-35. <https://doi.org/10.2307/2334611>
- Greenland S. Multivariate estimation of exposure-specific incidence from case-control studies. *J Chronic Dis*. 1981;34(9):445-453. [https://doi.org/10.1016/0021-9681\(81\)90004-7](https://doi.org/10.1016/0021-9681(81)90004-7)
- Persson E, Waernbaum I, Lind T. Estimating marginal causal effects in a secondary analysis of case-control data. *Stat Med*. 2017;36(15):2404-2419. <https://doi.org/10.1002/sim.7277>
- Hernán MA, Robins JM. *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC; 2020.
- Robins JM. [choice as an alternative to control in observational studies]: comment. *Stat Sci*. 1999;14(3):281-293. <https://doi.org/https://doi.org/http://www.jstor.org/stable/2676761>
- Månsson R, Joffe MM, Sun W, et al. On the estimation and use of propensity scores in case-control and case-cohort studies. *Am J Epidemiol*. 2007;166(3):332-339. <https://doi.org/10.1093/aje/kwm069>
- Newman SC. Causal analysis of case-control data. *Epidemiol Perspect Innov*. 2006;3(1):1-6. <https://doi.org/10.1186/1742-5573-3-2>
- Zhu A, Zeng D, Zhang P, et al. Estimating causal log-odds ratio using the case-control sample and its application in the pharmaco-epidemiology study. *Stat Methods Med Res*. 2019;28(7):2165-2178. <https://doi.org/10.1177/0962280217750175>
- Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol*. 2017;46(2):756-762. <https://doi.org/10.1093/ije/dyw323>
- Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9-12):1393-1512. [https://doi.org/10.1016/0270-0255\(86\)90088-6](https://doi.org/10.1016/0270-0255(86)90088-6)
- Taubman SL, Robins JM, Mittleman MA, et al. Intervening on risk factors for coronary heart disease: an application of the

- parametric g-formula. *Int J Epidemiol.* 2009;38(6):1599-1611. <https://doi.org/10.1093/ije/dyp192>
35. Keil AP, Edwards JK, Richardson DB, et al. The parametric g-formula for time-to-event data: intuition and a worked example. *Epidemiology.* 2014;25(6):889-897. <https://doi.org/10.1097/ede.0000000000000160>
 36. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol.* 1981;114(4):593-603. <https://doi.org/10.1093/oxfordjournals.aje.a113225>
 37. Luque-Fernandez MA, Schomaker M, Rachet B, et al. Targeted maximum likelihood estimation for a binary treatment: a tutorial. *Stat Med.* 2018;37(16):2530-2546. <https://doi.org/10.1002/sim.7628>
 38. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. *Biometrics.* 2005;61(4):962-973. <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
 39. van der Laan MJ, Rose S. *Targeted Learning: Causal Inference for Observational and Experimental Data.* New York: Springer; 2011.
 40. Merchant AT, Pitiphat W. Total, direct, and indirect effects of paan on oral cancer. *Cancer Causes Control.* 2015;26(3):487-491. <https://doi.org/10.1007/s10552-014-0516-x>
 41. Wu D, Yang H, Winham SJ, et al. Mediation analysis of alcohol consumption, DNA methylation, and epithelial ovarian cancer. *J Hum Genet.* 2018;63(3):339-348. <https://doi.org/10.1038/s10038-017-0385-8>
 42. Dashti SG, English DR, Simpson JA, et al. Adiposity and endometrial cancer risk in postmenopausal women: a sequential causal mediation analysis. *Cancer Epidemiol Biomarkers Prev.* 2021;30(1):104-113. <https://doi.org/10.1158/1055-9965.EPI-20-0965>
 43. O'Connell MM, Ferguson JP. Pathway-specific population attributable fractions. *Int J Epidemiol.* 2022;51(6):1957-1969. <https://doi.org/10.1093/ije/dyaa079>
 44. Donald BR. For objective causal inference, design trumps analysis. *Ann Appl Stat.* 2008;2(3):808-840. <https://doi.org/10.1214/08-AOAS187>
 45. Gruber S, Laan. Laan Mvd. Tmlle: an R package for targeted maximum likelihood estimation. *J Stat Softw.* 2012;51(13):1-35. <https://doi.org/10.18637/jss.v051.i13>
 46. Rose G. Sick individuals and sick populations. *Int J Epidemiol.* 2001;30(3):427-432discussion 433-4. <https://doi.org/10.1093/ije/30.3.427>
 47. Rockhill B. Theorizing about causes at the individual level while estimating effects at the population level: implications for prevention. *Epidemiology.* 2005;16(1):124-129. <https://doi.org/10.1097/01.ede.0000147111.46244.41>
 48. Dickerman BA, Garcia-Albeniz X, Logan RW, et al. Emulating a target trial in case-control designs: an application to statins and colorectal cancer. *Int J Epidemiol.* 2020;49(5):1637-1646. <https://doi.org/10.1093/ije/dyaa144>
 49. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol.* 2016;183(8):758-764. <https://doi.org/10.1093/aje/kwv254>
 50. Greenland S. Estimation of exposure-specific rates from sparse case-control data. *J Chronic Dis.* 1987;40(12):1087-1094. [https://doi.org/10.1016/0021-9681\(87\)90075-0](https://doi.org/10.1016/0021-9681(87)90075-0)
 51. Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology.* 2007;18(6):805-835. <https://doi.org/10.1097/EDE.0b013e3181577511>
 52. Kamenetsky MEK, Keil AP. A.P. (Re-) match: adjusting for matching factors in case-control studies may be unnecessary or insufficient. *Am J Epidemiol.* 2025; kwaf116. Advance online publication. <https://doi.org/10.1093/aje/kwaf116>
 53. Takeuchi Y, Hagiwara Y, Komukai S, et al. Estimation of the causal effects of time-varying treatments in nested case-control studies using marginal structural cox models. *Biometrics.* 2024;80(1). <https://doi.org/10.1093/biomtc/ujae005>
 54. Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika.* 1997;84(2):379-394. <https://doi.org/10.1093/biomet/84.2.379>
 55. Bedard A, Serra I, Dumas O, et al. Time-dependent associations between body composition, physical activity, and current asthma in women: a marginal structural modeling analysis. *Am J Epidemiol.* 2017;186(1):21-28. <https://doi.org/10.1093/aje/kwx038>
 56. Backenroth D, Chase H, Friedman C, et al. Using rich data on comorbidities in case-control study design with electronic health record data improves control of confounding in the detection of adverse drug reactions. *PLoS One.* 2016;11(10):e0164304. <https://doi.org/10.1371/journal.pone.0164304>
 57. Menvielle G, Franck J-E, Radoi L, et al. Quantifying the mediating effects of smoking and occupational exposures in the relation between education and lung cancer: the ICARE study. *Eur J Epidemiol.* 2016;31(12):1213-1221. <https://doi.org/10.1007/s10654-016-0182-2>
 58. Delcoigne B, Colzani E, Prochazka M, et al. Breaking the matching in nested case-control data offered several advantages for risk estimation. *J Clin Epidemiol.* 2017;82:79-86. <https://doi.org/10.1016/j.jclinepi.2016.11.014>
 59. Abdollahpour I, Nedjat S, Mansournia MA, et al. Estimation of the marginal effect of regular drug use on multiple sclerosis in the Iranian population. *PLoS One.* 2018;13(4):e0196244. <https://doi.org/10.1371/journal.pone.0196244>
 60. Abdollahpour I, Nedjat S, Mansournia MA, et al. Estimating the marginal causal effect of fish consumption during adolescence on multiple sclerosis: a population-based incident case-control study. *Neuroepidemiology.* 2018;50(3-4):111-118. <https://doi.org/10.1159/000487640>
 61. Velasquez Garcia HA, Sobolev BG, Gotay CC, et al. Mammographic non-dense area and breast cancer risk in postmenopausal women: a causal inference approach in a case-control study. *Breast Cancer Res Treat.* 2018;170(1):159-168. <https://doi.org/10.1007/s10549-018-4737-7>
 62. Figueroa SC, Kennedy CJ, Wesseling C, et al. Early immune stimulation and childhood acute lymphoblastic leukemia in Costa Rica: a comparison of statistical approaches. *Environ Res.* 2020;182(ei2, 0147621):109023. <https://doi.org/10.1016/j.envres.2019.109023>
 63. Lu D, Guo F, Li F. Evaluating the causal effects of cellphone distraction on crash risk using propensity score methods. *Accid Anal Prev.* 2020;143(acs, 1254476):105579. <https://doi.org/10.1016/j.aap.2020.105579>
 64. Abdollahpour I, Nedjat S, Almasi-Hashiani A, et al. Estimating the marginal causal effect and potential impact of Waterpipe smoking on risk of multiple sclerosis using the targeted maximum likelihood estimation method: a large, population-based incident case-control study. *Am J Epidemiol.* 2021;190(7):1332-1340. <https://doi.org/10.1093/aje/kwab036>
 65. Almasi-Hashiani A, Nedjat S, Ghiasvand R, et al. The causal effect and impact of reproductive factors on breast cancer using super learner and targeted maximum likelihood estimation: a case-control study in Fars Province, Iran. *BMC Public Health.* 2021;21(1):1219. <https://doi.org/10.1186/s12889-021-11307-5>
 66. Basnet TB, Srijana GC, Basnet R, et al. Causal effects of dietary calcium, zinc and iron intakes on coronary artery disease in

- men: G-estimation and inverse probability of treatment weighting (IPTW) analyses. *Clinical nutrition ESPEN*. 2021;42(101654592):73-81. <https://doi.org/10.1016/j.clnesp.2020.12.030>
67. Boska RL, Bishop TM, Ashrafioun L. Pain conditions and suicide attempts in military veterans: a case-control design. *Pain Med*. 2021;22(12):2846-2850. <https://doi.org/10.1093/pm/pnab287>
68. Akhtar S, El-Muzaini H, Alroughani R. Recombinant hepatitis B vaccine uptake and multiple sclerosis risk: a marginal structural modeling approach. *Mult Scler Relat Disord*. 2022;58(101580247):103487. <https://doi.org/10.1016/j.msard.2022.103487>
69. Yu Y, Choi J, Lee MH, et al. Maternal disease factors associated with neonatal jaundice: a case-control study. *BMC Pregnancy Childbirth*. 2022;22(1):247. <https://doi.org/10.1186/s12884-022-04566-6>
70. Elduma AH, Holakouie-Naieni K, Almasi-Hashiani A, et al. The targeted maximum likelihood estimation to estimate the causal effects of the previous tuberculosis treatment in multidrug-resistant tuberculosis in Sudan. *PLoS One*. 2023;18(1):e0279976. <https://doi.org/10.1371/journal.pone.0279976>
71. Malekifar P, Nedjat S, Abdollahpour I, et al. Impact of alcohol consumption on multiple sclerosis using model-based standardization and misclassification adjustment via probabilistic bias analysis. *Arch Iran Med*. 2023;26(10):567-574. <https://dx.doi.org/10.34172/aim.2023.83>
72. Caubet M, L'Espérance K, Koushik A, et al. An empirical evaluation of approximate and exact regression-based causal mediation approaches for a binary outcome and a continuous or a binary mediator for case-control study designs. *BMC Med Res Methodol*. 2024;24(1):72. <https://doi.org/10.1186/s12874-024-02156-y>