

RESEARCH

Open Access



Multiply robust estimator for the difference in survival functions using pseudo-observations

Ce Wang^{1†}, Kecheng Wei^{1†}, Chen Huang¹, Yongfu Yu^{1,2*} and Guoyou Qin^{1,2*}

Abstract

Background When estimating the causal effect on survival outcomes in observational studies, it is necessary to adjust confounding factors due to unbalanced covariates between treatment and control groups. There is no study on multiple robust method for estimating the difference in survival functions. In this study, we propose a multiply robust (MR) estimator, allowing multiple propensity score models and outcome regression models, to provide multiple protection.

Method Based on the previous MR estimator (Han 2014) and pseudo-observation approach, we proposed a new MR estimator for estimating the difference in survival functions. The proposed MR estimator based on the pseudo-observation approach has several advantages. First, the proposed estimator has a small bias when any PS and OR models were correctly specified. Second, the proposed estimator considers the advantage of the pseudo-observation approach, which avoids proportional hazards assumption. A Monte Carlo simulation study was performed to evaluate the performance of the proposed estimator. And the proposed estimator was used to estimate the effect of chemotherapy on triple-negative breast cancer (TNBC) in real data.

Results The simulation studies showed that the bias of the proposed estimator was small, and the coverage rate was close to 95% when any model for propensity score or outcome regression is correctly specified regardless of whether the proportional hazard assumption holds, finite sample size and censoring rate. And the simulation results also showed that even though the propensity score models are misspecified, the bias of the proposed estimator was still small when there is a correct model in candidate outcome regression models. And we applied the proposed estimator in real data, finding that chemotherapy could improve the prognosis of TNBC.

Conclusions The proposed estimator, allowing multiple propensity score and outcome regression models, provides multiple protection for estimating the difference in survival functions.

The proposed estimator provided a new choice when researchers have a "difficult time" choosing only one model for their studies.

Keywords Survival function, Survival outcome, Empirical likelihood, Propensity score, Multiply robust

[†]Ce Wang and Kecheng Wei contributed equally to this work.

*Correspondence:

Yongfu Yu

yu@fudan.edu.cn

Guoyou Qin

gyqin@fudan.edu.cn

Full list of author information is available at the end of the article



Background

Inferring the causal effect between specific treatment (exposure or intervention) and the outcome is the primary goal of much-applied research. Survival outcomes are common in epidemiologic studies and require special handling due to being subject to the right censoring during follow-up. In randomized controlled trials, the difference in survival functions based on Kaplan–Meier (KM) estimator between treatment and control groups can be used to assess the causal effect of treatment on survival outcomes [1, 2]. However, when treatment assignment is not randomized in observational studies, there are some unbalanced covariates between treatment and control groups, which leads to a biased estimation of the difference in survival functions [3]. There are two general approaches to eliminating or reducing confounding biases. The first approach is direct confounding adjustment via an outcome regression (OR) model such as Cox model [4]. The survival function in the control or treatment group is estimated by averaging the fitted values from the outcome regression model for each group, but the estimation may be biased when the proportional hazards assumption fails. Another approach is combining KM estimator with inverse probability weighting (IPW), where each subject is weighted by the inverse of the probability of receiving the treatment, known as propensity score (PS), thereby balancing the distribution of covariates between control and treatment groups [5, 6]. IPW method relies on the PS model reflecting the relationship of confounding factors with treatment [7]. The two methods above are unbiased only if the OR and PS models are correctly specified, which are usually unknown in practice.

There is growing interest in developing doubly robust (DR) methods for estimating the average causal effect (ACE) of survival outcomes to protect against potential misspecification of OR or PS model [4, 8–10]. DR methods are based on combining the IPW and direct confounding adjustment methods and provide double protection from model misspecification, as DR methods are valid when either one of the OR or PS models is specified correctly. However, DR methods do not provide sufficient protection for estimation consistency, as they allow only one model for IPW and one for direct confounding adjustment. Consistency of DR methods is still lost when PS and OR models are both misspecified [11, 12].

Compared to DR methods, multiply robust approaches allow multiple PS or OR models, increasing the likelihood of including the correct model. And multiply robust approaches are consistent when the candidate models contain correctly specified models [13]. Although multiply robust approaches are used for causal inference [14] or estimating a population mean with missing values [11, 12], they are proposed in the

context of a non-survival outcome. Recently, Shu et al. [13] proposed a multiply robust approach for estimating marginal hazard ratio, yet the method only includes PS models, not OR models.

In this study, we proposed a multiply robust (MR) estimator, including multiple PS and OR models, for estimating the difference in survival functions, which has not been studied in the literature.

Methods

Notation

For each subject $i (i = 1, 2, \dots, n)$, we observe a p -dimensional vector of covariates X_i , treatment status Z_i (i.e., it could be treatment status: $Z_i = 1$ if treated and $Z_i = 0$ if untreated), observed survival time T_i (the smaller value of true survival time \tilde{T}_i and censoring time C_i) and censoring indicator $\delta_i (\delta_i = 1$ when $T_i = \tilde{T}_i$ and $\delta_i = 0$ when $T_i = C_i$). We assume that true survival time \tilde{T}_i and censoring time are independent, given X_i and Z_i .

Note that (T^1, T^0) represent a pair of potential survival times and cannot be observed simultaneously for a subject, where T^1 represents the one that would occur if every subject were treated and T^0 represents the one that would happen if every subject were untreated. We relate the observed survival time to the potential survival times according to the consistency assumption, that is, $T = T^Z = ZT^1 + (1 - Z)T^0$. The difference in survival functions between treatment and control groups is defined to be

$$\Delta(t) = E\left[\mathbf{I}(T^1 > t)\right] - E\left[\mathbf{I}(T^0 > t)\right] = S^1(t) - S^0(t) \quad (1)$$

where t is the time point at which the difference in survival functions is estimated. where $S^1(t)$ is the survival function of the treatment group and $S^0(t)$ is the survival function of the control group.

Pseudo-observations for survival outcomes

Although survival functions for every individual can not be observed, they can be estimated by the pseudo-observation method proposed by Andersen et al. [15] and used to estimate the survival functions at a single time point [16]. Further, this approach was applied to estimate ACE of survival outcomes without considering assumptions of proportional hazards [17]. The pseudo-observation for i th individual survival function at a fixed time point t is defined as

$$\hat{S}_i(t) = n\hat{S}(t) - (n - 1)\hat{S}^{-i}(t) \quad (2)$$

where $\hat{S}(t)$ and $\hat{S}^{-i}(t)$ are the KM estimator using all samples and leaving out i th individual, respectively. Graw et al. [18] have examined the asymptotic properties of

the pseudo observations for competing risk models, also applying to $\widehat{S}_i(t)$.

Pseudo-observation-based IPW

In observational studies, direct comparison of survival functions between two groups may lead to a biased causal effect estimation due to unbalanced covariates between two groups. The IPW method based on the PS $\pi(X)$ is commonly used to adjust for confounding bias. We assume that there is no unmeasured confounding and positivity ($0 < \pi(X) < 1$) [3]. The ACE can be estimated by IPW estimator in the context of a non-survival outcome as follows

$$\widehat{\Delta}_{ipw} = \frac{1}{n} \sum_{i=1}^n \left[\frac{Z_i Y_i}{\widehat{\pi}_i} - \frac{(1 - Z_i) Y_i}{1 - \widehat{\pi}_i} \right] \tag{3}$$

where Y_i is the i th individual outcome (continuous or discrete outcomes), and $\widehat{\pi}$ is the estimated values of $\pi(X)$, for example, derived from the logistic regression models.

For time-to-event outcomes, adapting (3) for estimating $\Delta(t)$ is difficult as we do not observe individual survival functions; yet, the individual survival functions can be obtained according to (2). Once the individual survival functions are obtained, it is easy to estimate the $\Delta(t)$ for a fixed time point t with (3), and the IPW estimator based on the pseudo-observation method is defined as [4, 17],

$$\widehat{\Delta}(t)_{ipw} = \frac{1}{n} \sum_{i=1}^n \left[\frac{Z_i \widehat{S}_i(t)}{\widehat{\pi}_i} - \frac{(1 - Z_i) \widehat{S}_i(t)}{1 - \widehat{\pi}_i} \right] \tag{4}$$

Pseudo-observation-based direct confounding adjustment method

The direct confounding adjustment method for eliminating confounding bias in an OR model is another method to estimate ACE in observational studies. For time-to-event outcomes, Cox model is the most commonly used to be as an outcome regression model

$$h(t|X_i, Z_i) = h_0(t) \exp(\beta^T X_i + \gamma Z_i) \tag{5}$$

where $h_0(t)$ is the baseline hazard function, β is a set of parameters of covariates and γ is the parameter of treatment variable Z . According to (5), we can estimate individual survival function $S(t|X_i, Z_i)$ for the i th subject given X_i and Z_i

$$S(t|X_i, Z_i) = S_0(t) \exp(\beta^T X_i + \gamma Z_i) \tag{6}$$

where $S_0(t)$ can be estimated by fitting the Cox model. The survival functions in treatment and control are estimated by averaging individual survival

functions estimated from (6), respectively. And $\Delta(t)$ can be estimated by directly comparing the survival functions between two groups. However, this method requires the proportional hazards assumption.

Pseudo observations estimated from (2) can be used to predict individual survival functions with covariates and treatment variable for a fixed time point t , which avoids the proportional hazards assumption [4, 17]. Specifically, let $\mathbf{t} = \{t_1, \dots, t_h\}$ as the set of distinct times and define the pseudo-observation for subject i at time t_h as $\widehat{S}_i(t_h)$, where $h = 1, \dots, H$ and $i = 1, \dots, n$. Then we can convert the Cox model to a generalized linear model (GLM) with a suitable link function $g(x) = \log(-\log(x))$ [15],

$$g(S_i(t_h)) = \xi^{t_h} + \gamma Z_i + \beta^T X_i \tag{7}$$

Where ξ^{t_h} is the intercept term for time t_h , and γ and β are regression parameters. Note that, although ξ^{t_h} are time specific, γ and β are shared between different time points $\mathbf{t} = \{t_1, \dots, t_h\}$. Estimating the unknown parameters γ , β and $\beta_H = \{\xi^1, \dots, \xi^{t_h}\}$ could be solved by the following generalized estimating equation (GEE)

$$\sum_{i=1}^n \frac{\partial g^{-1}(t, \beta^*; Z_i, X_i)}{\partial \beta} V_i^{-1} \{ \widehat{S}_i(t) - g^{-1}(t, \beta^*; Z_i, X_i) \} = U(\beta^*) \tag{8}$$

where $\beta^* = (\beta_H, \gamma, \beta)$, $\widehat{S}_i(\mathbf{t}) = (\widehat{S}_i(t_1), \dots, \widehat{S}_i(t_h))^T$ and V_i^{-1} is a working covariance matrix which may account for the correlation structure inherent to the pseudo-observations [17]. The difference in survival functions with direct confounding adjustment based on the pseudo-observation method under model (7) can be estimated as

$$\widehat{\Delta}(t)_{OR} = \frac{1}{n} \sum_{i=1}^n \{ g^{-1}(\widehat{\xi}^t + \widehat{\gamma} + \widehat{\beta}^T X_i) - g^{-1}(\widehat{\xi}^t + \widehat{\beta}^T X_i) \} \tag{9}$$

If the outcome regression model (5) is correctly specified, the $\widehat{\Delta}(t)_{OR}$ is consistent.

Multiply robust estimator for the difference in survival functions

We combined the advantages of previous MR (Han, 2014) and PO method to proposed the proposed MR method. The proposed method could consider a set of OR models and a sets of PS models and also avoid the proportional hazards assumption. The step of estimating the difference in survival functions is similar to the previous MR (Han, 2014). First, specifying multiple PS models and multiple OR models. Second, estimating weights of individual in the treatment and control groups, respectively. Finally, estimating the difference in survival functions between treatment and control groups.

Specifically, to achieve multiple robustness, we specify multiple PS models and multiple OR models, $\mathcal{P} = \{\pi^l(X), l = 1, 2, 3, \dots, L\}$ for the propensity score and $\mathcal{M} = \{m_Z^k(X), k = 1, 2, 3, \dots, K\}$ for outcome regression, where $m_Z^k(X) = m^k(X, Z)$. Without loss of generality, let $\mathbb{I} = 1, \dots, n_1$ represent the indexes for treated subjects and $\mathbb{J} = 1, \dots, n_0$ the indexes for untreated subjects. n_1 and $n_0 = n - n_1$ are the size of treatment and control groups, respectively.

$$\hat{w}_i = \frac{1}{n_1} \frac{1}{1 + \hat{\rho}_1^T \hat{\mathbf{g}}_1(X_i)} \quad (i \in \mathbb{I}) \tag{18}$$

$$\hat{w}_j = \frac{1}{n_0} \frac{1}{1 + \hat{\rho}_0^T \hat{\mathbf{g}}_0(X_j)} \quad (j \in \mathbb{J}) \tag{19}$$

where

$$\hat{\mathbf{g}}_1(X)^T = \{\hat{\pi}^1(X) - \hat{\theta}_1^1, \dots, \hat{\pi}^L(X) - \hat{\theta}_1^L, \hat{m}_1^1(X) - \hat{\eta}_1^1, \dots, \hat{m}_1^K(X) - \hat{\eta}_1^K\}$$

$$\hat{\mathbf{g}}_0(X)^T = \{(1 - \hat{\pi}^1(X)) - \hat{\theta}_0^1, \dots, (1 - \hat{\pi}^L(X)) - \hat{\theta}_0^L, \hat{m}_0^1(X) - \hat{\eta}_0^1, \dots, \hat{m}_0^K(X) - \hat{\eta}_0^K\}$$

Second, to recover survival function from treated subjects for a time point t , we impose $w_i (i \in \mathbb{I})$ as the weights of individual pseudo-observation $S_i(t)$ from the treated subjects, which can be estimated by maximizing $\prod_{i \in \mathbb{I}} w_i$ according to the following constraints:

$$w_i \geq 0 (i \in \mathbb{I}) \tag{10}$$

$$\sum_{i \in \mathbb{I}} w_i = 1 \tag{11}$$

$$\sum_{i \in \mathbb{I}} w_i \hat{\pi}^l(X_i) = \hat{\theta}_1^l (l = 1, 2, 3, \dots, L) \tag{12}$$

$$\sum_{i \in \mathbb{I}} w_i \hat{m}_1^k(X_i) = \hat{\eta}_1^k (k = 1, 2, 3, \dots, K) \tag{13}$$

where $\hat{\theta}_1^l = n^{-1} \sum_{i=1}^n \pi^l(X_i)$ and $\hat{\eta}_1^k = n^{-1} \sum_{i=1}^n m_1^k(X_i)$. Similarly, we also impose $w_j (j \in \mathbb{J})$ as the weights of individual pseudo-observation $S_j(t)$ from the untreated subjects, which can be estimated by maximizing $\prod_{j \in \mathbb{J}} w_j$ subject to the following constraints:

$$w_j \geq 0 (j \in \mathbb{J}) \tag{14}$$

$$\sum_{j \in \mathbb{J}} w_j = 1 \tag{15}$$

$$\sum_{j \in \mathbb{J}} w_j (1 - \hat{\pi}^l(X_j)) = \hat{\theta}_0^l (l = 1, 2, 3, \dots, L) \tag{16}$$

$$\sum_{j \in \mathbb{J}} w_j \hat{m}_0^k(X_j) = \hat{\eta}_0^k (k = 1, 2, 3, \dots, K) \tag{17}$$

where $\hat{\theta}_0^l = n^{-1} \sum_{j=1}^n (1 - \pi^l(X_j))$ and $\hat{\eta}_0^k = n^{-1} \sum_{j=1}^n m_0^k(X_j)$. According to Lagrange multiplier method [19], it is easy to show that w_i and w_j maximizing the $\prod_{i \in \mathbb{I}} w_i$ and $\prod_{j \in \mathbb{J}} w_j$ are given by

and $\hat{\rho}_t^T = (\hat{\rho}_{t1}, \hat{\rho}_{t2}, \dots, \hat{\rho}_{tS})$ is $S = J + K$ -dimensional vector satisfying the equations

$$\frac{1}{n_1} \sum_{i \in \mathbb{I}} \frac{\hat{\mathbf{g}}_1(X_i)}{1 + \hat{\rho}_1^T \hat{\mathbf{g}}_1(X_i)} = 0 \tag{20}$$

$$\frac{1}{n_0} \sum_{j \in \mathbb{J}} \frac{\hat{\mathbf{g}}_0(X_j)}{1 + \hat{\rho}_0^T \hat{\mathbf{g}}_0(X_j)} = 0 \tag{21}$$

Due to the non-negativity of w_i and w_j , $\hat{\rho}_t^T$ must satisfy that $1 + \hat{\rho}_1^T \hat{\mathbf{g}}_1(X_i) > 0$ and $1 + \hat{\rho}_0^T \hat{\mathbf{g}}_0(X_j) > 0$. The estimation of \hat{w}_i and \hat{w}_j can be solved by the Newton–Raphson algorithm [19].

Finally, the MR estimator of the difference in survival functions between treatment and control groups for the fixed time t is defined as

$$\hat{\Delta}(t)_{mr} = \sum_{i \in \mathbb{I}} \hat{w}_i \hat{S}_i(t) - \sum_{j \in \mathbb{J}} \hat{w}_j \hat{S}_j(t) \tag{22}$$

$\hat{S}_i(t)$ and $\hat{S}_j(t)$ are the individual survival functions estimated with PO method.

Bootstrap confidence interval

We perform the bootstrap method to estimate the confidence interval of $\hat{\Delta}(t)_{mr}$. Specifically, n subjects are resampled from the original data with the replacement for B times to obtain B bootstrap samples, where B is a user-specified number. For $b = 1, \dots, B$, let $\hat{\Delta}_{mr}^b(t)$ be the estimated difference of survival functions from the b -th bootstrap sample. Then the bootstrap variance estimator for $\hat{\Delta}_{mr}^b(t)$ is defined as

$$\widehat{var}(\hat{\Delta}_{mr}^b(t)) = \frac{1}{B-1} \sum_{b=1}^B (\hat{\Delta}_{mr}^b(t) - \frac{1}{B} \sum_{b=1}^B \hat{\Delta}_{mr}^b(t))^2 \tag{23}$$

A normality-based 95% confidence interval for $\Delta(t)_{mr}$ is $\hat{\Delta}_{mr}^b(t) \pm 1.96 \sqrt{\widehat{var}(\hat{\Delta}_{mr}^b(t))}$.

Simulation studies

Simulation design

We conducted comprehensive simulation studies to evaluate the performance of the proposed method. There were three covariates X_1, X_2, X_3 distributed as standard normal with mean zero and unit variance, where $\text{corr}(X_1, X_3) = 0.2$ and X_2 was independent of the X_1 and X_3 . The treatment indicator Z was simulated from the Bernoulli distribution according to the following propensity score

$$\text{logit}[\pi(X)] = (X_1 + X_2 + X_3)/3$$

The survival times were simulated from a Cox-Weibull model with $\tilde{T}_1 = \left[\frac{-\log(u)}{\lambda_1 \exp(L-\gamma)} \right]^{\frac{1}{\nu_1}}$ for treatment group and $\tilde{T}_0 = \left[\frac{-\log(u)}{\lambda_0 \exp(L)} \right]^{\frac{1}{\nu_0}}$ for control group, where $L = -3 + 0.5 * (X_1 + X_2)$, $\gamma = 1$, u was simulated from $\text{Uniform}(0,1)$. And we defined the proportional hazard setting via setting the shape and scale parameters [$\nu_1 = 0.005, \lambda_1 = 3$], [$\nu_0 = 0.005, \lambda_0 = 3$] in the

treatment and control groups (Figure S1). An independent censoring time was simulated from an exponential distribution with rates, and we set the different rates [$\exp(-4)$ or $\exp(-3)$] to obtain censoring rates (approximately 25% or 50%).

In this simulation, we specified two models

$$\mathbb{A} = \left\{ \begin{aligned} \text{logit} \left[\pi^1(X; \beta^1) \right] &= (X_1, X_2, X_3) \beta^1 \\ \text{logit} \left[\pi^2(X; \beta^2) \right] &= (X_1^2, X_2^2, X_3^2) \beta^2 \end{aligned} \right\}$$

for propensity score, and two models

$$\mathbb{B} = \left\{ \begin{aligned} m^1(X, Z; \gamma^1) &= (1, X_1, X_2, Z) \gamma^1 \\ m^2(X, Z; \gamma^2) &= (1, X_1^2, X_2^2, Z) \gamma^2 \end{aligned} \right\}$$

for outcome regression. According to the data-generating process, $\pi^1(X; \beta^1)$ and $m^1(X, Z; \gamma^1)$ were correctly specified, while $\pi^2(X; \beta^2)$ and $m^2(X, Z; \gamma^2)$ were incorrectly specified.

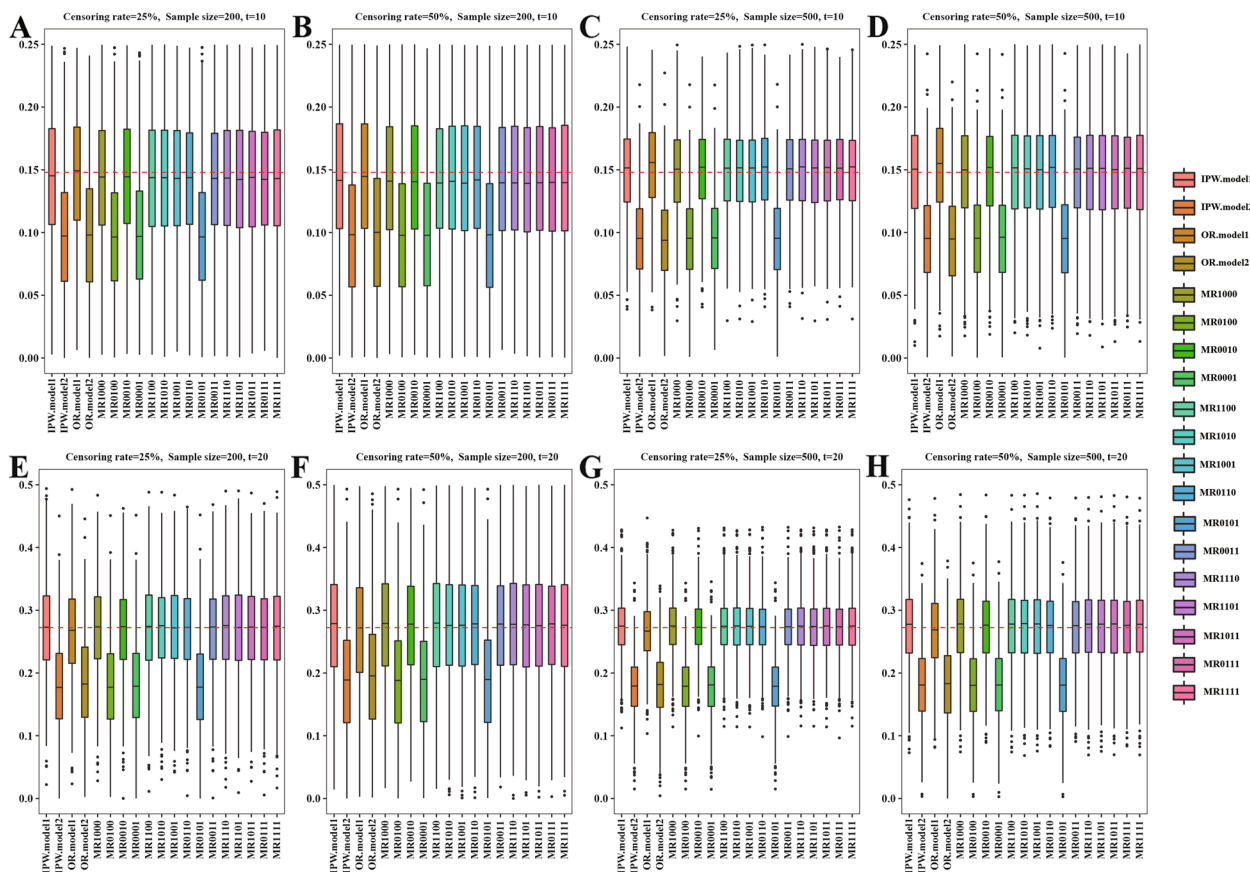


Fig. 1 Simulation results with different sample sizes = 200 or 500 and different censoring rates = 25% or 50% in the scenario where proportional hazard assumption holds based on 1000 replication. The red dotted line represents the true value, the true values of $\Delta(t = 10)$ or $\Delta(t = 20)$ are 0.1480 and 0.2725, respectively. IPW: inverse probability weighting; OR: outcome regression; MR, multiply robust; MR estimators are denoted as “MR-0000”, where each digit of the four numbers, from left to right, indicates if $\pi^1(X; \beta^1), \pi^2(X; \beta^2), m^1(X, Z; \gamma^1)$ or $m^2(X, Z; \gamma^2)$ is included in the estimator (“1” means yes and “0” means no)

We were interested in estimating the difference in survival functions $\Delta(t^*)$ at $t^* = 10$ and 20 . We performed IPW, direct confounding adjustment, and multiply robust (MR) methods to estimate the $\Delta(t)$. To distinguish these estimation methods, two IPW-based estimators were denoted as “IPW.model1” and “IPW.model2”; two direct-confounding-adjustment-based estimators were denoted as “OR.model1” and “OR.model2”; for the MR estimators, each estimator is denoted as “MR-0000”, where each digit of the four numbers, from left to right, indicates if $\pi^1(X; \beta^1)$, $\pi^2(X; \beta^2)$, $m^1(X, Z; \gamma^1)$ or $m^2(X, Z; \gamma^2)$ is included in the estimator (“1” means yes and “0” means no).

In addition, we defined the non-proportional hazard setting via setting the shape scale parameters [$(\nu_1 = 0.005, \lambda_1 = 3)$, $(\nu_0 = 0.005, \lambda_0 = 3.2)$] in the treatment and control groups, respectively (Figure S1); and we also specified two models incorrectly

$$\mathbb{A} = \left\{ \begin{array}{l} \text{logit}[\pi^1(X; \beta^1)] = (X_1^2, X_2^2)\beta^1 \\ \text{logit}[\pi^2(X; \beta^2)] = (X_1^2, X_2^2, X_3^2)\beta^1 \end{array} \right\}$$

for propensity score. In the study, we investigated a sample size of $n = 200$ and $n = 500$ with 1000 replications. We used three evaluation criteria: mean relative bias, root mean square error (RMSE), and coverage rate. The boxplots of mean relative bias, RMSE and coverage rate were shown in the main article (Figs. 1, 2, 3 and 4) and the table of estimation results were shown in the supplementary material (Table S1- S4).

Simulation results

Figure 1 and Table S1 showed the simulation results of the difference between treatment and control groups in the proportional hazard setting with the different censoring rates (25% or 50%) and different sample sizes (200 or 500). We could draw a few conclusions from the simulation results.

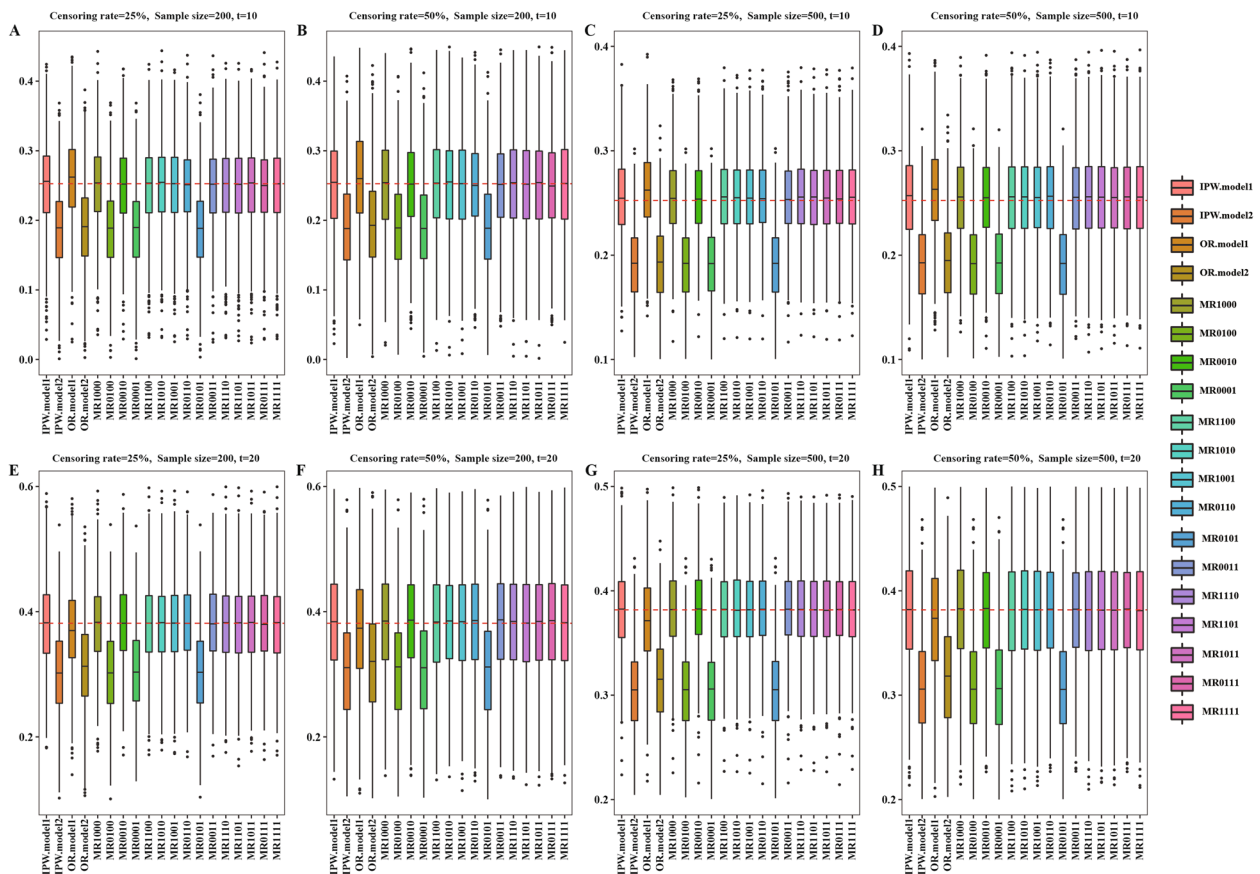


Fig. 2 Simulation results with different sample sizes = 200 or 500 and different censoring rates = 25% or 50% in the scenario where proportional hazard assumption does not hold based on 1000 replications. The red dotted line represents the true value, the true values of $\Delta(t = 10)$ or $\Delta(t = 20)$ are 0.2524 and 0.3816, respectively. IPW: inverse probability weighting; OR: outcome regression; MR, multiply robust; MR estimators are denoted as “MR-0000”, where each digit of the four numbers, from left to right, indicates if $\pi^1(X; \beta^1)$, $\pi^2(X; \beta^2)$, $m^1(X, Z; \gamma^1)$ or $m^2(X, Z; \gamma^2)$ is included in the estimator (“1” means yes and “0” means no)

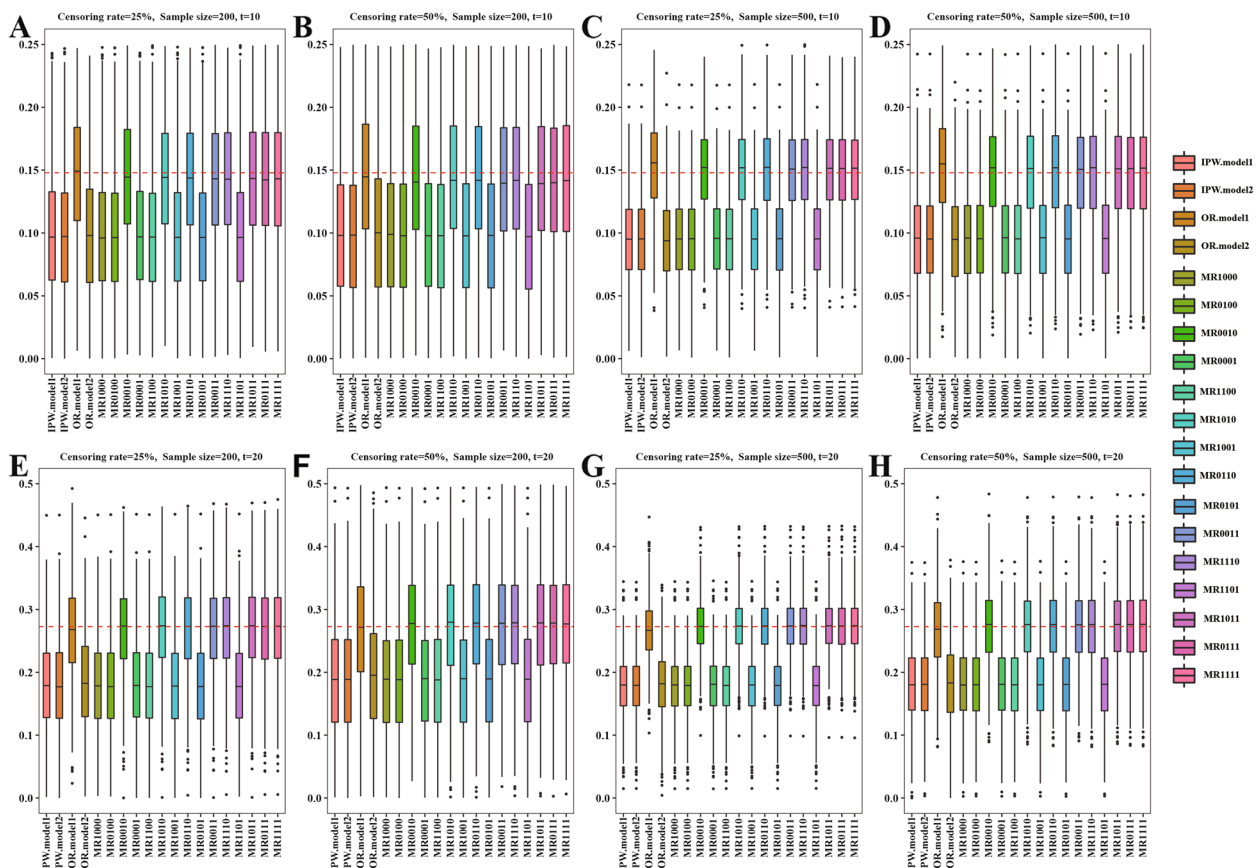


Fig. 3 Simulation results with different sample sizes = 200 or 500 and different censoring rates = 25% or 50% in the scenario where proportional hazard assumption holds based on 1000 replications when PS models misspecified. The red dotted line represents the true value, the true values of $\Delta(t = 10)$ or $\Delta(t = 20)$ are 0.1480 and 0.2725, respectively. IPW: inverse probability weighting; OR: outcome regression; MR, multiply robust; MR estimators are denoted as “MR-0000”, where each digit of the four numbers, from left to right, indicates if $\pi^1(\mathbf{X}; \boldsymbol{\beta}^1)$, $\pi^2(\mathbf{X}; \boldsymbol{\beta}^2)$, $m^1(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma}^1)$ or $m^2(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma}^2)$ is included in the estimator (“1” means yes and “0” means no)

1. When specifying a propensity score model $\pi(\mathbf{X}; \boldsymbol{\beta})$, or an outcome regression model $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$: the biases of IPW and MR were negligible if $\pi(\mathbf{X}; \boldsymbol{\beta})$ was correctly specified (IPW.model1, MR1000). The biases of OR and MR were small if $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$ was correctly modeled (OR.model1, MR0010), and MR0010 had the smallest RMSE among all estimators. The biases and RMSEs were large, and the coverage rates were much less than 95% if $\pi(\mathbf{X}; \boldsymbol{\beta})$ or $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$ were incorrectly specified (IPW.model2, OR.model2, MR0100, MR0001).
2. When specifying a propensity score model $\pi(\mathbf{X}; \boldsymbol{\beta})$, and an outcome regression model $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$: the biases MR were ignorable when either $\pi(\mathbf{X}; \boldsymbol{\beta})$ or $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$ was correctly specified (MR1010, MR1001, MR0110). The biases and RMSEs were severe large, and the coverage rates were well below 95% when the $\pi(\mathbf{X}; \boldsymbol{\beta})$ and $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$ were both incorrectly specified (MR0101). MR1010, MR1001

- and MR0110 had almost the same RMSEs, which were smaller than IPW.model1 and OR.model1.
3. When specifying multiply propensity score models $\pi(\mathbf{X}; \boldsymbol{\beta})$, and multiply outcome regression model $m(\mathbf{X}, \mathbf{Z}; \boldsymbol{\gamma})$: the small biases of MR1100, MR0011, MR1110, MR1101, MR1011, MR0111, and MR1111 well suggested the multiple robustness of our proposed methods. Based on our simulation results, the RMSEs of those estimators were almost all smaller than IPW.model1 and OR.model1, suggesting that adding more models did not lead to a significant increase in RMSE.

In addition, Fig. 2 and Table S2 showed the estimation results in the non-proportional hazard setting. Figures 3 and 4 and Tables S3 and S4 showed the simulation results in proportional and non-proportional hazard settings with two PS models specified incorrectly. Similar patterns were observed in Figs. 2, 3 and 4 and Tables S2-S4.

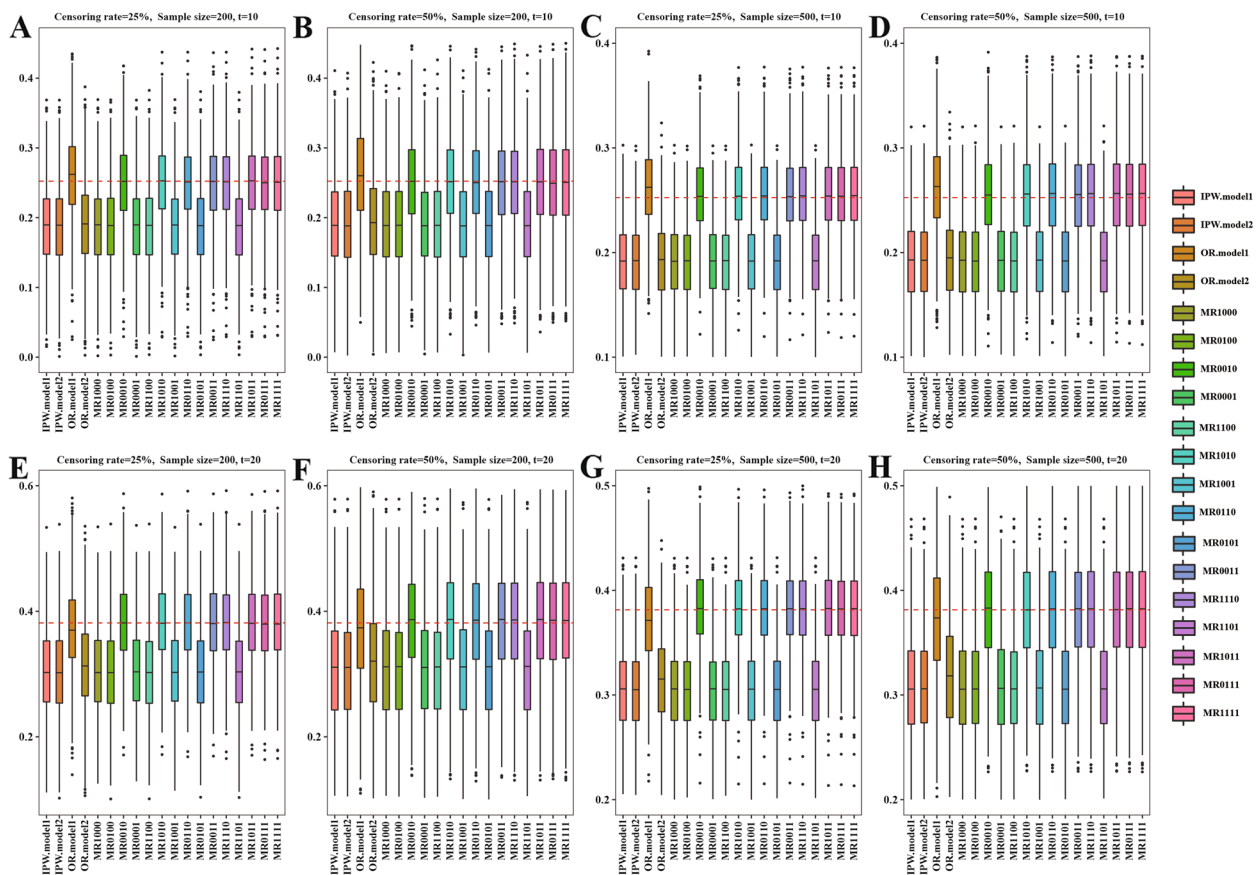


Fig. 4 Simulation results with different sample sizes = 200 or 500 and different censoring rates = 25% or 50% in the scenario where proportional hazard assumption dose not hold based on 1000 replications when PS models misspecified. The red dotted line represents the true value, the true values of $\Delta(t = 10)$ or $\Delta(t = 20)$ are 0.2524 and 0.3816, respectively. IPW: inverse probability weighting; OR: outcome regression; MR, multiply robust; MR estimators are denoted as “MR-0000”, where each digit of the four numbers, from left to right, indicates if $\pi^1(X; \beta^1)$, $\pi^2(X; \beta^2)$, $m^1(X, Z; \gamma^1)$ or $m^2(X, Z; \gamma^2)$ is included in the estimator (“1” means yes and “0” means no)

In conclusion, the simulation results showed that our MR method could provide more protection against model misspecification even when PS models are misspecified in the proportional or non-proportional hazard settings.

Empirical study

To illustrate the proposed method, we analyzed the real-world data on triple-negative breast cancer (TNBC) from Surveillance, Epidemiology, and End Results (SEER) [20]. TNBC, characterized by an absence of estrogen, progesterone receptors, and human epidermal growth factor receptor 2, has high invasiveness, high metastatic potential, proneness to relapse, and poor prognosis [21]. Chemotherapy remains the primary treatment for TNBC and improves the prognosis of TNBC patients [22]. The objective of the study was to estimate the effect of chemotherapy on TNBC. Baseline information included marital status, race, laterality, grade, The American Joint

Committee on Cancer (AJCC) stage, distant metastasis, age, surgery, and chemotherapy. Inclusion criteria are: (1) patients diagnosed between 2010 and 2012, (2) patients aged between 20 and 79, (3) one primary tumor only, (4) complete baseline and survival information. A total of 10,613 patients were included in the analysis, with 79.9% opting for chemotherapy and a censoring rate of 80%. The baseline information was summarized in Table 1. We found that the chemotherapy group had a higher proportion of patients with III/IV stage, ≥ 50 years old, receiving radiotherapy and surgery compared to the non-chemotherapy group.

To identify the candidate models for MR method, we explored the association of chemotherapy with all covariates through the multivariate logistic regression, and the association of survival with all covariates through the multivariate Cox model. Two sets of covariates for each model were identified using the significant level at 0.05 and 0.1. Two sets of covariates in $[\{\pi^1(X; \beta^1), \pi^2(X; \beta^2)\}]$

Table 1 Baseline characteristics between chemotherapy and non-chemotherapy groups

	Non-chemotherapy ^a	Chemotherapy
Survival (months)	44.61 ± 18.20	45.40 ± 16.82
Death		
Yes	463 (21.7)	1876 (22.1)
No	1673 (78.3)	6601 (77.9)
Marital Status		
Yes ^b	1190 (55.7)	5176 (61.1)
No ^c	946 (44.3)	3301 (38.9)
Laterality		
Left	1105 (51.7)	4333 (51.1)
Right	1031 (48.3)	4144 (48.9)
Race		
Black	409 (19.1)	1798 (21.2)
Non-black ^d	1727 (80.9)	6679 (78.8)
Grade		
Well/Moderately	1545 (72.3)	7302 (86.1)
Poorly/Undifferentiated	591 (27.7)	1175 (13.9)
Stage		
I/II	1926 (90.2)	6551 (77.3)
III/IV	210 (9.8)	1926 (22.7)
Distant metastasis		
Yes	101 (4.7)	727 (8.6)
No	2035 (95.3)	7750 (91.4)
Age		
< 50	382 (17.9)	3054 (36.0)
≥ 50	1754 (82.1)	5423 (64.0)
Radiotherapy		
Yes	783 (36.7)	5022 (59.2)
No	1353 (63.3)	3455 (40.8)
Surgery		
Yes	1994 (93.4)	8097 (95.5)
No	142 (6.6)	380 (4.5)

^a including not receiving chemotherapy and unknown status

^b including divorced, separated, single (never married), unmarried and widowed

^c including married and domestic partner

^d including White, American Indian/Alaskan native, Asian/Pacific Islander, and others-unspecified

are: (i) marital status, grade, stage, age, radiotherapy, and surgery; (ii) marital status, race, grade, stage, age, radiotherapy, and surgery. Two sets of covariates in $[m^1(X, Z; \gamma^1), m^2(X, Z; \gamma^2)]$ are: (i) marital status, race, grade, stage, distant metastasis, age, radiotherapy, surgery, and chemotherapy; (ii) marital status, race, laterality, grade, stage, distant metastasis, age, radiotherapy, surgery, and chemotherapy. We applied the IPW, direct confounding adjustment, and MR methods to estimate the $\Delta(t)$ at the time point $t^* = \text{year } 1, 2, 3, 4, 5$ and 6. The results were shown in Fig. 5 and Table S5.

From Fig. 5 and Table S5, all estimators indicated that chemotherapy could improve the prognosis of TNBC. There were some differences among the values of different estimators. For example, the estimates of MR1000, MR0100 and MR1100 are larger than the other MR estimates at $t^* = \text{year } 1, 2, 3, 4, 5, 6$; and the estimates of MR0010, MR0001 and MR0000 seem to be smaller than the other MR estimates at $t^* = \text{year } 1, 2, 3$ year. Yet, the proposed method could include multiple PS models and OR models, which were more robust to the model specification to a certain extent and might yield more accurate estimates. And the estimated chemotherapy effects using MR1111 are 0.0455(0.0278–0.0633), 0.0379(0.0173–0.0586), 0.0337(0.0116–0.0559), 0.0364(0.0127–0.0601), 0.0340(0.0061–0.0620) and 0.0423(0.0101–0.0745) at $t^* = \text{year } 1, 2, 3, 4, 5, 6$, respectively. The results suggests that the chemotherapy group had a survival rate of 4.55%, 3.79%, 3.37%, 3.64%, 3.40% and 4.23% higher than the controls at $t^* = \text{year } 1, 2, 3, 4, 5, 6$, respectively; and the effects were all statistically significant at the 95% level.

Discussion

In this study, we considered estimating the difference in survival functions between control and treatment groups in observational studies. The proposed MR method allowed multiple PS and OR models, and the bias of the proposed method is small when any model for propensity score or outcome regression is correctly specified regardless of in proportional or non-proportional hazard settings. Although we mainly focused on estimating the difference in survival functions in methodology, our proposed method can be easily extended to estimate other time-to-event parameters. For example, the restricted average survival causal effect $\int_0^{t^*} S^1(t) - \int_0^{t^*} S^0(t)$ compares the survival time at t^* , and the survival quantile effect $(S^1)^{-1}(1-q) - (S^0)^{-1}(1-q)$ compares the q -quantile of the survival time distribution between two groups [2].

Wang [4] combined the advantages of augmented inverse propensity weighted and PO method to proposed a simple, doubly robust method for estimating the causal effect of survival outcome [4]. Although the DR method also can avoid proportional hazards assumption, the DR method may lead to biased estimates when OR model and PS model are both misspecified [4]. By contrast, our proposed method could allow multiple PS models and multiple OR models, which provides extra protection from model misspecification. Our simulation studies showed that the IPW and direct confounding adjustment methods could lead to large bias and low coverage rate when misspecifying the propensity score and outcome regression models. By contrast, our proposed method

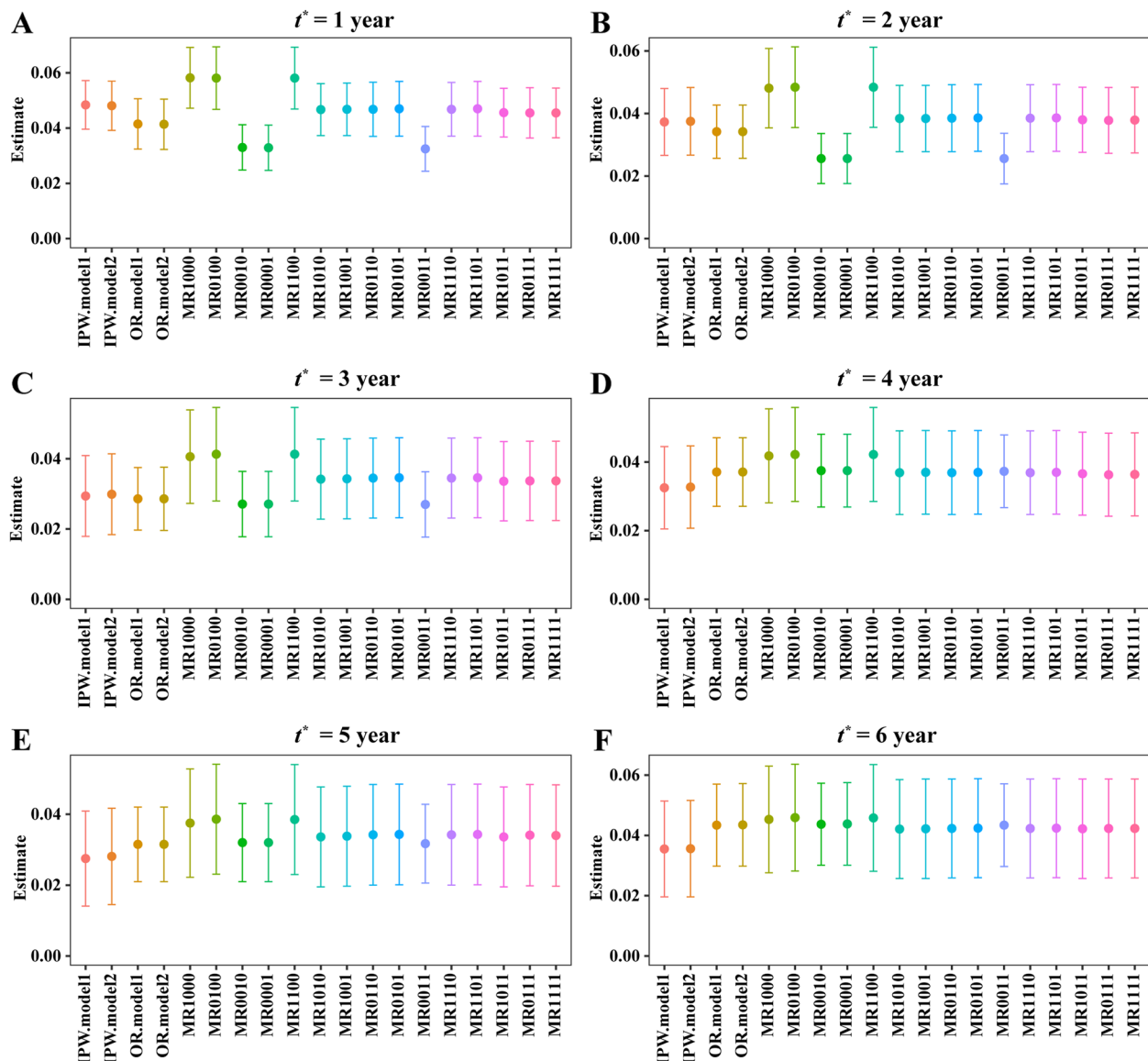


Fig. 5 The triple-negative breast cancer data analysis: estimated $\Delta(t^*)$ for $t^* = 1, 2, 3, 4, 5$ and 6 years using different methods

showed satisfactory performance with finite sample sizes. For the proposed method, the bias was ignorable and the coverage rate was close to 95% when correctly specifying any PS and OR models.

There are some limitations in the study. First, we assume that there is no unmeasured confounding; However, there are often confounding variables that cannot be measured in observational studies, which leads to a biased estimation [23, 24]. Second, our proposed method could not deal with competing risks, which are also common in survival data [25, 26]. Third, we identify the candidate models based on the significance level of 0.05 and 0.1 in the empirical study, while the model specification may need clinical knowledge and existing literature, or

statistical methods [27, 28]. Fourth, although the proposed method could multiple PS models and multiple OR models, which provides multiple protection against model misspecification, there may a biased estimated when the specified models do not include a correct model.

Conclusions

Under the positivity, no unmeasured confounding (conditional exchangeability), consistency assumptions, the proposed MR method, pseudo-observation-based IPW and pseudo-observation-based direct confounding adjustment methods could be used to estimate the difference in survival functions. However, Pseudo-observation-based

IPW or pseudo-observation-based direct confounding adjustment may lead to a biased estimate when the PS model or OR models is misspecified. By contrast, the proposed MR method based on the pseudo-observation approach has several advantages. First, the proposed method could allow a set of PS models and a set of OR models, and the proposed method has a small bias when any PS and OR models were correctly specified. Second, The proposed method considered the advantage of pseudo-observation approach, which avoids proportional hazards assumption. In actual research, the proposed estimator provided a new choice when researchers have a difficult time choosing only one model for their studies.

Abbreviations

ACE	Average causal effect
AJCC	American Joint Committee on Cancer
GEE	Generalized estimating equation
GLM	Generalized linear model
IPW	Inverse probability weighting
KM	Kaplan–Meier
MR	Multiply robust
OR	Outcome regression
PS	Propensity score
RMSE	Root mean square error
TNBC	Triple-negative breast cancer

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-023-02065-6>.

Additional file 1: Table S1. Simulation results with different sample sizes =200 or 500 and different censoring rates =25% or 50% in the scenario where proportional hazard assumption holds based on 1000 replication. **Table S2.** Simulation results with different sample sizes =200 or 500 and different censoring rates =25% or 50% in the scenario where proportional hazard assumption dose not hold based on 1000 replication. **Table S3.** Simulation results with different sample sizes =200 or 500 and different censoring rates =25% or 50% in the scenario where proportional hazard assumption holds based on 1000 replication when PS models misspecified. **Table S4.** Simulation results with different sample sizes =200 or 500 and different censoring rates =25% or 50% in the scenario where proportional hazard assumption dose not hold based on 1000 replication when PS models misspecified. **Table S5.** The triple-negative breast cancer data analysis: estimated $\Delta(t^*)$ for $t^* =$ year 1, 2, 3, 4, 5 and 6 years using different methods. **Figure S1.** Weibull survival functions defining simulation scenario. Figure A shows proportional hazards assumption holds; and Figure B shows that proportional hazards assumption dose not hold.

Acknowledgements

Not applicable.

Authors' contributions

GYQ and YFY designed the study. CW and KCW wrote the manuscript. CW carried out the simulations and data analysis. CW, CH, and KCW revised the manuscript. All authors have read and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China (No. 82273730 and 82173612), Shanghai Rising-Star Program (21QA1401300), Shanghai Municipal Natural Science Foundation (22ZR1414900) and Shanghai Municipal Science and Technology Major Project (ZD2021CY001). The sponsors

had no role in study design, data collection, data analysis, data interpretation, or writing of this report.

Availability of data and materials

Publicly available datasets were analyzed in this study. These data can be downloaded from <https://seer.cancer.gov/>. The simulation code can be downloaded from the github website <https://github.com/Ce-Wang/PO-MR.git>.

Declarations

Ethics approval and consent to participate

Since the simulated datasets did not involve any human data, ethics approval was not applicable; and the real data is publicly available, thus ethics approval was not required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biostatistics, Key Laboratory for Health Technology Assessment, National Commission of Health, Key Laboratory of Public Health Safety of Ministry of Education, School of Public Health, Fudan University, Shanghai, China. ²Shanghai Institute of Infectious Disease and Biosecurity, Shanghai, China.

Received: 1 February 2023 Accepted: 11 October 2023

Published online: 23 October 2023

References

- Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53(282):457–81.
- Mao H, Li L, Yang W, Shen Y. On the propensity score weighting analysis with survival outcome: Estimands, estimation, and inference. *Stat Med.* 2018;37(26):3745–63.
- Hernán MA, Robins JM: Causal inference. In.: CRC Boca Raton, FL; 2010.
- Wang J. A simple, doubly robust, efficient estimator for survival functions using pseudo observations. *Pharm Stat.* 2018;17(1):38–48.
- Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics.* 2000;56(3):779–88.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med.* 2005;24(20):3089–110.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70(1):41–55.
- Choi S, Choi T, Lee HY, Han SW, Bandyopadhyay D: Doubly-robust methods for differences in restricted mean lifetimes using pseudo-observations. *Pharm Stat* 2022.
- Zhang M, Schaubel DE. Double-robust semiparametric estimator for differences in restricted mean lifetimes in observational studies. *Biometrics.* 2012;68(4):999–1009.
- Zhang M, Schaubel DE. Contrasting treatment-specific survival using double-robust estimators. *Stat Med.* 2012;31(30):4255–68.
- Han P, Wang L. Estimation with missing data: beyond double robustness. *Biometrika.* 2013;100(2):417–30.
- Han P. A further study of the multiply robust estimator in missing data analysis. *J Stat Plan Inference.* 2014;148:101–10.
- Shu D, Han P, Wang R, Toh S. Estimating the marginal hazard ratio by simultaneously using a set of propensity score models: A multiply robust approach. *Stat Med.* 2021;40(5):1224–42.
- Wang L. Multiple robustness estimation in causal inference. *Commun Stat Theory Methods.* 2019;48(23):5701–18.
- Andersen PK, Pohar Perme M. Pseudo-observations in survival analysis. *Stat Methods Med Res.* 2010;19(1):71–99.

16. Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. *Stat Med.* 2007;26(24):4505–19.
17. Andersen PK, Syriopoulou E, Parner ET. Causal inference in survival analysis using pseudo-observations. *Stat Med.* 2017;36(17):2669–81.
18. Graw F, Gerds TA, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* 2009;15(2):241–55.
19. Han P. Multiply robust estimation in regression analysis with missing data. *J Am Stat Assoc.* 2014;109(507):1159–73.
20. Hayat MJ, Howlader N, Reichman ME, Edwards BK. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) Program. *Oncologist.* 2007;12(1):20–37.
21. Yin L, Duan J-J, Bian X-W, Yu S-c. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res.* 2020;22(1):1–13.
22. Isakoff SJ. Triple negative breast cancer: role of specific chemotherapy agents. *Cancer J (Sudbury, Mass).* 2010;16(1):53.
23. Shi X, Miao W, Nelson JC, Tchetgen Tchetgen EJ. Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding. *J R Stat Soc Ser B Stat Methodol.* 2020;82(2):521–40.
24. Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med.* 2014;33(13):2297–340.
25. Ozenne BMH, Scheike TH, Stærk L, Gerds TA. On the estimation of average treatment effects with right-censored time to event outcome and competing risks. *Biom J.* 2020;62(3):751–63.
26. Lin J, Trinquart L. Doubly-robust estimator of the difference in restricted mean times lost with competing risks data. *Stat Methods Med Res.* 2022;31(10):1881–903.
27. Shortreed SM, Ertefaie A. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics.* 2017;73(4):1111–22.
28. Ertefaie A, Asgharian M, Stephens DA. Variable selection in causal inference using a simultaneous penalization method. *Journal of Causal Inference.* 2018;6(1):20170010.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

