Communications for Statistical Applications and Methods
2024, Vol. 31, No. 1, 155–178

https://doi.org/10.29220/CSAM.2024.31.1.155
Print ISSN 2287-7843 / Online ISSN 2383-4757

Regression discontinuity for survival data

Youngjoo Cho[1,a]

[a]Department of Applied Statistics, Konkuk University, Korea

Abstract

Regression discontinuity (RD) design is one of the most widely used methods in causal inference for estimation of treatment effect when the treatment is created by a cutpoint from the covariate of interest. There has been little attention to RD design, although it provides a very useful tool for analysis of treatment effect for censored data. In this paper, we define the causal effect for survival function in RD design when the treatment is assigned deterministically by the covariate of interest. We propose estimators of this causal effect for survival data by using transformation, which leads unbiased estimator of the survival function with local linear regression. Simulation studies show the validity of our approach. We also illustrate our proposed method using the prostate, lung, colorectal and ovarian (PLCO) dataset.

Keywords: survival analysis, causal inference, local linear regression, regression discontinuity, doubly robust

1. Introduction

Regression discontinuity (RD) design is one of the most widely used methods in causal inference for the estimation of treatment effect by creating discontinuity with a covariate of interest. In this design, the treatment assignment is decided by covariate of interest deterministically or probabilistically, a so-called running variable with a pre-determined threshold. When the treatment assignment is a determinis- tic function of running variable, it is called sharp RD design. If the treatment assignment is a function of running variable with randomness, it is called fuzzy RD design. Due to its characteristic, in sharp RD, the treatment assignment is random in the threshold of running variable. In other words, in this sharp RD, although our study is an observational study, we have the same environment as ran- domized assignment of treatment in the threshold of running variable. This property enables us to avoid unmeasured confounders assumption, which is a fundamental one in causal inference but may be unreasonable in practice.

The RD design was first proposed by Thistlethwaite and Campbell (1960), and it has received much attention in social science and economics. For example, Ludwig and Miller (2007) study the effect of funding in education. Lee (2008) uses RD design to study the effect of party affiliation probability of Democrats winning in the next election. For theoretical work, Hahn *et al*. (1999) and Hahn *et al*. (2001) prove nonparametric identification of treatment effect and asymptotic results of the estimator of the treatment effect on RD design. To estimate the treatment effect, they use local linear regression. This local linear or polynomial regression is a widely used method on the RD design because they effectively handle discontinuity and are theoretically well supported.

This work was supported by Konkuk University (2021-A019-0164).

[1]Department of Applied Statistics, Konkuk University, 120 Neungdong-ro, Gwangjin-gu, Seoul 07341, Korea. E-mail: yvc5154@konkuk.ac.kr

Published 31 January 2024 / journal homepage: http://csam.or.kr

Since choosing bandwidth in the local regression is a crucial issue, many researchers have proposed several methods. Ludwig and Miller (2007) propose method for choosing bandwidth from the local linear regression by cross-validation. Imbens and Kalyanaraman (2012) propose choosing bandwidth by optimizing mean squared error. Recently, Calonico *et al*. (2014) propose method which corrects bias from local regression and provides robust confidence interval. Calonico *et al*. (2020) propose bandwidth choice to achieve smaller coverage rate than one in optimizing mean squared error.

However, the aforementioned works are focused on uncensored data. The RD design in the censored data can provide a useful answer to the clinician's research question. As discussed by Shoag *et al*. (2015) and Cho *et al*. (2021), in prostate cancer research, the usefulness of patient screening based on prostate-specific antigen (PSA) for survival or prostate-cancer specific incidence is still an open question. In this setup, the outcome of interest is time to death or prostate cancer incidence, and it is subject to censoring, i.e., patients may not experience death or prostate cancer on their observed time. In practice, a patient is considered high risk if his PSA level is greater than or equal to 4.0mg/nl. People in the screening group with a PSA of 4.0 ng/ml at any time were suggested to receive further checkup and biopsy. This is a clearly sharp RD setting: Shoag *et al*. (2015) use RD design with a binary outcome to answer this question using prostate, lung, colorectal, and ovarian (PLCO) data with prostate cancer. The objective of this PLCO trial is to investigate whether screenings for prostate, lung, colorectal, ovarian cancers are effective to reduce mortality. It is a multi-center, two-armed, and randomized trial (Prorok *et al*., 2000). Men and women people with age between 55 to 74 were enrolled at 10 centers in the United States from 1993 to 2001. Patients with prostate cancer in the screening group received PSA testing for 6 years and digital rectal examination for 4 years annually (Andriole *et al*., 2009).

There is little research on using RD design in censored data. Recently, Bor *et al*. (2014) and Moscoe *et al*. (2015) use RD design to answer a research question on the effect of early versus late treatment initiation for the survival of HIV patients. Cho *et al*. (2021) extend Steingrimsson *et al*. (2019) to RD design and employ local linear regression by Fan and Gijbels (1996) to estimate treatment effects under RD design. They focus on survival time and analyze the aforementioned PLCO data. However, there is no research to discuss modeling survival probability, an essential quantity in the censored data, in the RD design. Moreover, our proposed method allows different treatment effects at each time point, opposite to Cho *et al*. (2021)'s method.

In this paper, we discuss modeling survival probability in the RD design. Our approach employs the approach of Steingrimsson *et al*. (2019) and Cho *et al*. (2021). Our method enjoys doubly robustness, as Steingrimsson *et al*. (2019) and Cho *et al*. (2021). Moreover, we also investigate the application of the pseudo-value approach (Anderson *et al*., 2003) to RD design, which provides an unbiased estimator of the survival function.

The paper is organized as follows. In the next section, we discuss the estimation of treatment effect with regard to survival probability in the RD design for uncensored and censored data. We propose asymptotic theory and statistical inference for the proposed method, and the extension with the pseudo-value approach from Anderson *et al*. (2003) in Section 3. We demonstrate simulation studies with various settings in Section 4. Real data analysis is shown in Section 5. Conclusion and future research are discussed in Section 6.

2. Regression discontinuity with regards to survival probability

From framework of Rubin (1974), we define the potential outcome for survival data. Let $T^{(1)}$ and $T^{(0)}$ be the potential time to the event of treatment group and control group, respectively. Let W be a running variable. In the RD design, there are two designs: Sharp and fuzzy design. In this paper, we only discuss the sharp design case. In the sharp design, the treatment is assigned deterministically by the threshold. Let w_0 be threshold on W. Our treatment variable Z is defined by

$$Z = I(W \geq w_0).$$

We are interested in the average treatment effect (ATE) with respect to survival function.

$$\text{ATE} = P\left(T^{(1)} > t\right) - P\left(T^{(0)} > t\right).$$

From potential outcome structure, our observable outcome is $T = ZT^{(1)} + (1 - Z)T^{(0)}$. From the structure of the sharp RD, the ATE can be expressed by the difference of survival functions in the neighborhood of w_0. In the neighborhood of w_0, the limits of two survival functions above and below w_0 are equal to $P(T^{(1)} > t)$ and $P(T^{(0)} > t)$ with difference of treatment status and their actual survival functions $P(T > t)$ (Zajonc, 2012). To identify the causal effect, the assumptions similar to Cho *et al.* (2021) are required.

C.1 Participants do not have perfect manipulation on the cutoff.

C.2 $P(T^{(1)} > t|W = w)$ and $P(T^{(0)} > t|W = w)$ are continuous on $W = w_0$ for all t.

Condition C.1 is required to create a "randomized environment" in the cutoff. Condition C.2 shows that the potential survival function for treatment and control given W is smooth under $W = w_0$. With these conditions, we can identify the average causal effect of the survival function

$$\begin{aligned}
\tau(t) &= P\left(T^{(1)} > t \mid W = w_0\right) - P\left(T^{(0)} > t \mid W = w_0\right) \\
&= \lim_{w \downarrow w_0} P\left(T^{(1)} > t \mid W = w\right) - \lim_{w \uparrow w_0} P\left(T^{(0)} > t \mid W = w\right) \\
&= \lim_{w \downarrow w_0} P\left(T^{(1)} > t \mid W = w, Z = 1\right) - \lim_{w \uparrow w_0} P\left(T^{(0)} > t \mid W = w, Z = 0\right) \\
&= \lim_{w \downarrow w_0} P\left(T > t \mid W = w\right) - \lim_{w \uparrow w_0} P\left(T > t \mid W = w\right) \\
&= \lim_{w \downarrow w_0} E\{I\left(T > t \mid W = w\right)\} - \lim_{w \uparrow w_0} E\{I\left(T > t \mid W = w\right)\}.
\end{aligned}$$

The full data is $(T_i, W_i, Z_i)_{i=1}^n$, the i.i.d copies of (T, W, Z). As Hahn *et al.* (1999) point out, boundary observations leads poor numerical results for computing estimator and perform inference. To address this issue, local linear regression method (Fan and Gijbels, 1996) is widely used in RD design. We apply the Brier score (Brier, 1950), a squared error loss for probability to local linear regression. Let $V(t) = I(T > t)$, $K(\cdot)$ be kernel function and h be bandwidth. Define $\{\alpha_R(t), \beta_R(t)\}$ and $\{\alpha_L(t), \beta_L(t)\}$ as regression parameters correponding to local linear regression of right and left limits. Then the proposed loss functions are

$$U_R(\alpha_R(t), \beta_R(t)) = \sum_{i=1}^{n} I(W_i \geq w_0) \left\{V_i(t) - \alpha_R(t) - \beta_R(t)(W_i - w_0)\right\}^2 K\left(\frac{W_i - w_0}{h}\right),$$

$$U_L(\alpha_L(t), \beta_L(t)) = \sum_{i=1}^{n} I(W_i < w_0) \left\{V_i(t) - \alpha_L(t) - \beta_L(t)(W_i - w_0)\right\}^2 K\left(\frac{W_i - w_0}{h}\right),$$

where $V_i(t) = I(T_i > t)$. Then by minimization of U_R and U_L, we obtain $\{\hat{\alpha}_R(t), \hat{\beta}_R(t), \hat{\alpha}_L(t), \hat{\beta}_L(t)\}$. Then the corresponding estimator for SRD is

$$\hat{\tau}(t) = \hat{\alpha}_R(t) - \hat{\alpha}_L(t).$$

For this approach, we can use standard methods of RD methodology proposed by the aforementioned literatures. Our outcome is $I(T > t)$, and we apply local linear regression.

Now we introduce censoring in our data. Let C be time to the censoring and $a \wedge b$ be minimum a and b where $a, b \in \mathbb{R}$. The observed data is i.i.d $(\tilde{T}_i, \Delta_i, W_i, Z_i)$ where $\tilde{T}_i = T_i \wedge C_i$ and $\Delta_i = I(T_i \leq C_i)$. For censored data case, it is very difficult to use response directly because $P(\tilde{T} > t) \neq P(T > t)$. By adapting idea from Cho *et al.* (2021), our goal is to find time-dependent transformation q_t such that $E\{q_t(T|W)\} = E(I(T > t)|W) = P(T > t|W)$. The transformation q_t is also censoring unbiased transformation (Fan and Gijbels, 1994; Rubin and van der Laan, 2007; Cho *et al.*, 2021). The widely used function q_t is inverse probability censoring weighted (IPCW) method.

$$E\left[\frac{\Delta \cdot I(T > t)}{G(T \mid W)}\right] = \left[\frac{P(C \geq T \mid W)}{G(T \mid W)}\right] E\{I(T > t \mid X)\} = P(T > t \mid W). \tag{2.1}$$

Hence we obtain unbiased estimator of $P(T > t|W)$. The transformation (2.1) is

$$V_{IPCW_1}(t) = \frac{\Delta V(t)}{G(T \mid W)}. \tag{2.2}$$

However, this approach only uses data with uncensored observation (i.e., $\Delta = 1$). To use more information than $V_{IPCW_1}(t)$, Gref *et al.* (1999) and Lostritto *et al.* (2012) use truncation on T and Δ. We consider $T(t) = T \wedge t$, $\tilde{T}(t) = T(t) \wedge C$ and $\Delta(t) = I(T(t) \leq C)$. Throughout the truncation, $\Delta = 1$ implies $\Delta(t) = 1$, but $\Delta = 0$ may lead $\Delta(t) = 1$. From the idea of Gref *et al.* (1999) and Lostritto *et al.* (2012), we propose

$$V_{IPCW_2}(t) = \frac{\Delta(t)V(t)}{G(T(t) \mid W)}. \tag{2.3}$$

Due to the aforementioned reasoning, $V_{IPCW_2}(t)$ has smaller variance than $V_{IPCW_1}(t)$ because it uses more information than $V_{IPCW_1}(t)$. These approaches are easy to implement and supported by asymptotic theory (e.g., Strawderman, 2000). For example, $V_{IPCW_2}(t)$ is uniform consistent with regard to mean square error (Gerds and Schumacher, 2006) and has good performance compared to $V_{IPCW_1}(t)$ (Cho *et al.*, 2022).

However, these approaches require that censoring distribution is correct and these approaches yield inefficient estimators because we only use $\Delta = 1$ or $\Delta(t) = 1$. Motivated by semiparametric efficiency theory (Robins *et al.*, 1994; Tsiatis, 2007), we obtain transformation

$$\begin{aligned} V_{DR}(t) &= \frac{\Delta V(t)}{G(T \mid W)} + \int_0^t \frac{Q(t, u \mid W)}{G(u \mid W)} dM_G(u \mid W) \\ &= \frac{\Delta(t)V(t)}{G(T(t) \mid W)} + \int_0^{\tilde{T}(t)} \frac{Q(t, u \mid W)}{G(u \mid W)} dM_G(u \mid W), \end{aligned} \tag{2.4}$$

where $\lambda_G(s|W)$ is the true hazard function of G given W and

$$\begin{aligned} M_G(u \mid W) &= I\left(\tilde{T} \leq u, \Delta = 0\right) - \int_0^u I\left(\tilde{T} \geq s\right) \lambda_G(s \mid W) ds \\ Q(t, u \mid W) &= \frac{P(T \geq t \mid W)}{P(T \geq u \mid W)}, \quad t \geq u. \end{aligned}$$

The equality of last two terms in (2.4) is proved in Cho *et al.* (2020). Note that Q involves the modeling survival function. The (2.4) is doubly robust because it is required to be either model for $G(\cdot)$ or survival model correct, but not necessarily both. It guarantees that $E(V_{DR}(t)|W = w) = P(T > t|W = w)$ for any given w, so it is censoring unbiased transformation. This doubly robust transformation is a combination of the IPCW term and the mean zero martingale transformation term. This martingale transformation term utilizes censored data, which yields more efficiency than $V_{IPCW_1}(t)$ and $V_{IPCW_2}(t)$, but it has fairly similar performance or slightly better efficiency than $V_{IPCW_2}(t)$ (Cho *et al.*, 2022).

3. Proposed method

3.1. Causal effect estimation and inference with IPCW and DR estimators

Let $\hat{V}_{IPCW_1,i}(t)$, $\hat{V}_{IPCW_2,i}(t)$ and $\hat{V}_{DR,i}(t)$ be the transformed response for i^{th} observation using (2.2), (2.3) and (2.4). Denote $N_{i,G}(u) = I(\tilde{T}_i \le u, \Delta_i = 0)$. Then we can express $\hat{V}_{IPCW_1,i}(t)$, $\hat{V}_{IPCW_2,i}(t)$ and $\hat{V}_{DR,i}(t)$ by

$$\hat{V}_{IPCW_1,i}(t) = \frac{\Delta_i I\left(\tilde{T}_i \ge t\right)}{\hat{G}\left(\tilde{T}_i\right)},$$

$$\hat{V}_{IPCW_2,i}(t) = \frac{\Delta_i(t) I\left(\tilde{T}_i \ge t\right)}{\hat{G}\left(\tilde{T}_i(t)\right)},$$

$$\hat{V}_{DR,i}(t) = \frac{\Delta_i(t) I\left(\tilde{T}_i \ge t\right)}{\hat{G}\left(\tilde{T}_i(t)\right)} + \int_0^{\tilde{T}_i(t)} \frac{\hat{Q}(t, u \mid W_i)}{\hat{G}(u)} d\hat{M}_{i,G}(u),$$

where

$$M_{i,G}(u) = N_{i,G}(u) - \int_0^u I\left(\tilde{T}_i \ge s\right) d\hat{\Lambda}_G(s) ds,$$

and $\hat{\Lambda}_G(s)$ is the estimated cumulative hazard function with respect to C, and $\hat{Q}(t, u|W_i)$ is the estimator of $Q(t, u|W_i)$. For the calculation of $\hat{G}$, due to the random censoring assumption, we use the Kaplan-Meier estimator. Hence we compute $\hat{\Lambda}_G$ by Nelson-Aalen estimator with respect to C. We use various survival models for $\hat{Q}$ such as the Cox model (Cox, 1972) and the parametric accelerated failure time (AFT) model.

After the calculation of $\hat{V}_{IPCW_1,i}(t)$, $\hat{V}_{IPCW_2,i}(t)$ or $\hat{V}_{DR,i}(t)$, we use a local linear apporach from Fan and Gijbels (1996) on transformed response as Cho *et al.* (2021). Let $\hat{V}_{CUT,i}$ be one of $\hat{V}_{IPCW_1,i}(t)$, $\hat{V}_{IPCW_2,i}(t)$ or $\hat{V}_{DR,i}(t)$ and $K(\cdot)$ be a kernel function and h be bandwidth. Suppose that the h is given. We will explain bandwidth estimation in the next section. As Imbens and Lemieux (2008), we build two loss functions

$$U_R\left(\alpha_R(t), \beta_R(t); \hat{G}, \hat{S}\right) = \sum_{i=1}^{n} I\left(W_i \ge w_0\right) \left\{\hat{V}_{CUT,i}(t) - \alpha_R(t) - \beta_R(t)\left(W_i - w_0\right)\right\}^2 K\left(\frac{W_i - w_0}{h}\right) \quad (3.1)$$

$$U_L\left(\alpha_L(t), \beta_L(t); \hat{G}, \hat{S}\right) = \sum_{i=1}^{n} I\left(W_i < w_0\right) \left\{\hat{V}_{CUT,i}(t) - \alpha_L(t) - \beta_L(t)(W_i - w_0)\right\}^2 K\left(\frac{W_i - w_0}{h}\right). \quad (3.2)$$

Then we calculate $\{\alpha_R(t), \beta_R(t), \alpha_L(t), \beta_L(t)\}$ which minimize $\{U_R(\alpha_R(t), \beta_R(t); \hat{G}, \hat{S}), U_L(\alpha_L(t), \beta_L(t); \hat{G}, \hat{S})\}$ by weighted least squares. Let $\hat{\alpha}_{R,DR}(t), \hat{\beta}_{R,DR}(t), \hat{\alpha}_{L,DR}(t), \hat{\beta}_{L,DR}(t)$ be minimizer of (3.1) with

$\hat{V}_{CUT,i}(t) = \hat{V}_{DR,i}(t)$. We can similarly define

$$\left\{\hat{\alpha}_{R,IPCW_1}(t), \hat{\beta}_{R,IPCW_1}(t), \hat{\alpha}_{L,IPCW_1}(t), \hat{\beta}_{L,IPCW_1}(t)\right\}$$
$$\left\{\hat{\alpha}_{R,IPCW_2}(t), \hat{\beta}_{R,IPCW_2}(t), \hat{\alpha}_{L,IPCW_2}(t), \hat{\beta}_{L,IPCW_2}(t)\right\}.$$

In this case, for simplicity, we suppress notations $\hat{G}$ and $\hat{S}$. Then our sharp RD estimator is obtained by

$$\begin{aligned}\hat{\tau}_{IPCW_1}(t) &= \hat{\alpha}_{R,IPCW_1}(t) - \hat{\alpha}_{L,IPCW_1}(t),\\ \hat{\tau}_{IPCW_2}(t) &= \hat{\alpha}_{R,IPCW_2}(t) - \hat{\alpha}_{L,IPCW_2}(t),\\ \hat{\tau}_{DR}(t) &= \hat{\alpha}_{R,DR}(t) - \hat{\alpha}_{L,DR}(t).\end{aligned}$$

We adapt the method from Cho *et al.* (2021) for survival functions. Our $\hat{V}_{CUT,i}$ is the new response for the regression and we adapt the idea of the Brier score for estimation. When $\hat{Q} = 0$, $\hat{V}_{DR,i}(t)$ reduces either $\hat{V}_{ICPW_1,i}(t)$ or $\hat{V}_{ICPW_2,i}(t)$. Furthermore, when there is no censoring, $\hat{V}_{ICPW_1,i}(t)$ and $\hat{V}_{ICPW_2,i}$ reduce to $V_i(t)$. This is an advantage of our method: Our method is not only restricted to uncensored data but also can be applied to censored data. Our method bridges between censored and uncensored data, which does not happen in the typical survival analysis modeling.

Our approach is different from Cho *et al.* (2021) in several ways. First, Cho *et al.* (2021) use $\log T$ to estimate causal effects. This estimation is useful, but interpretation with logarithm may be practically not easy. We express causal effects in terms of survival function, and it is more relevant in practice. Moreover, since our causal effect depends on time t, we allow different causal effect in each time, while Cho *et al.* (2021)'s method does not allow it. Hence our approach is more flexible than Cho *et al.* (2021)'s method.

Remark. We observe the same phenomenon in Cho *et al.* (2021); in their paper, when $E(\log(T)|T \geq u, W)$ is 0, their DR-transformed outcome reduces to an IPCW-transformed outcome, and when additionally no censoring exists, the IPCW-transformed outcome reduces to the usual continuous outcome.

Due to the nature of our estimation procedure, we can adapt the result of asymptotic results from the mean of the logarithm of survival time in Cho *et al.* (2021). We show that our estimators are asymptotically normal in Appendix A (see Theorem 1 and proof in the Appendix A.).

Now we discuss the estimation of bandwidth for $\hat{\tau}(t)$. First, we apply the method from Ludwig and Miller (2007), which is also used in Cho *et al.* (2021). The criterion proposed by Ludwig and Miller (2007) uses empirical distributions W_i with $W_i < w_0$ and W_i with $W_i \geq w_0$. Then with values of W_i from these empirical distributions, we compute squared error from transformed response and corresponding our DR estimator with some range of W.

This method is simple and does not depend on the variance of the transformed response. First, define $\hat{a}_L(\xi)$ to be the ξ quantile of the empirical distribution of W using observations $W_i < w_0$ and let $\hat{a}_R(1-\xi)$ be the $1-\xi$ quantile of the empirical distribution of W using observations $W_i \geq w_0$. Then we compute the following quantity:

$$CV_{DR,LM}\left(h; \hat{G}, \hat{S}\right) = \frac{1}{n} \sum_{\hat{a}_L(\xi) \leq W_i \leq \hat{a}_R(1-\xi)} \left(\hat{V}_{DR,i}(t) - \hat{\gamma}_{DR}\left(h, t; W_i\right)\right)^2, \tag{3.3}$$

where

$$\hat{\gamma}_{DR}(h, t \mid w) = \begin{cases} \hat{\alpha}_{L,DR}(h, t; w), & \text{if } w < w_0, \\ \hat{\alpha}_{R,DR}(h, t; w), & \text{if } w \geq w_0. \end{cases} \tag{3.4}$$

In other words, for given h, we first compute the estimator of $\hat{\alpha}_{L,DR}(t)$ and $\hat{\alpha}_{R,DR}(t)$ with some truncated range of W_i, and secondly calculate the average of squared error with regards to $\hat{\alpha}_{L,DR}(h,t;w)$ and $\hat{\alpha}_{R,DR}(h,t;w)$, respectively. We then choose

$$\hat{h}_{DR}\left(t,\hat{G},\hat{S}\right) = \underset{h}{\operatorname{argmin}}\, CV_{DR,LM}\left(h,t;\hat{G},\hat{S}\right).$$

We can derive a similar quantity for $V_{IPCW_1,i}(\hat{G}), V_{IPCW_2,i}(\hat{G}), i = 1,\ldots,n$. The second quantity is the mean squared error criterion for the RD estimator by Imbens and Kalyanaraman (2012). Since local linear (or polynomial) regression gives a bias to the estimator, it is desirable to select bandwidth that minimizes the mean squared error. In this case, the target quantity is the mean squared error of the sharp RD estimator. In the uncensored data, for sharp RD estimator $\hat{\tau}_h$, which is $\hat{\tau}$ with given h and true value τ_0, we want to find h to minimize

$$E\left\{(\hat{\tau}_h - \tau_0)^2\right\}. \tag{3.5}$$

Imbens and Kalyanaraman (2012) provide asymptotic results for the expansion of (3.5). With uncensored data, let Y be an outcome, $\sigma^2_+(w_0)$ and $\sigma^2_-(w_0)$ be the right and left limits of $\mathrm{Var}(Y|W=w)$ on the threshold w_0 and $m^{(2)}_+$ and $m^{(2)}_-$ be the right and left hand limits of the second derivative of $E(Y|W=w)$ on threshold w_0. They propose the bandwidth selection approach by

$$h_{MSE,IK} = C_K \left\{ \frac{\sigma^2_+(w_0) + \sigma^2_-(w_0)}{g(w_0)\left(m^{(2)}_+(w_0) - m^{(2)}_-(w_0)\right)^2} \right\} n^{-\frac{1}{5}},$$

where C_K is constant from the function of kernel function and g is the density function of W. This approach provides an estimator whose the mean square error is asymptotically optimal and it is widely used in the practice (Calonico *et al.*, 2020). However, this method requires estimation of g, σ^2_+ and σ^2_-, which is not preferred. Calonico *et al.* (2014) provide an alternative expansion of (3.5)

$$E\left\{\hat{\tau}_h - \tau_0\right\}^2 = h^4\left(\mathcal{B}^2 + o_p(1)\right) + \frac{1}{nh}\left(\mathcal{V} + o_p(1)\right), \tag{3.6}$$

where $\mathcal{V}$ and $\mathcal{B}$ is variance and bias of $\hat{\tau}_h$. They suggest that optimal bandwidth based on the mean square error in (3.6) by

$$h_{MSE,C} = \left(\frac{\mathcal{V}}{4\mathcal{B}^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \tag{3.7}$$

Then we can compute this $h_{MSE,C}$ by following steps:

Step 1: Take initial bandwidths to compute $\mathcal{B}$ and $\mathcal{V}$. For this, one can use Silverman's rule of thumb (Silverman, 1986).

Step 2: By using (3.6) and (3.7), compute the final bandwidth $h_{MSE,C}$.

We use IPCW and DR estimators in the place of $\hat{\tau}_h$ and compute $h_{MSE,C}$, and use them in the inference.

Now we want to perform inference for the proposed sharp RD estimators. To compute variance, we adapt the approaches of Cho *et al.* (2021). We only propose the method based on DR transformation; the derivation of IPCW_1 and IPCW_2 are similar. Suppose that h is computed by Ludwig and Miller (2007) or the mean square error criterion by Calonico *et al.* (2014).

Table 1: Simulation results with data generated from Cox model with 30% censoring rate

			Bias		ESD		SE				Cover			
			LM	MSE	LM	MSE	LM		MSE		LM		MSE	
							NN	HC0	NN	HC0	NN	HC0	NN	HC0
$n = 500$	t_1	$IPCW_1$	0.001	−0.004	0.181	0.334	0.173	0.171	0.308	0.300	0.946	0.940	0.924	0.930
		$IPCW_2$	0.003	0.008	0.105	0.178	0.100	0.099	0.179	0.174	0.928	0.924	0.944	0.940
		DR(Cox)	0.003	0.006	0.096	0.160	0.087	0.086	0.156	0.151	0.920	0.916	0.940	0.928
		DR(lognorm)	0.003	0.006	0.096	0.160	0.087	0.086	0.156	0.151	0.916	0.916	0.940	0.926
		DR(loglog)	0.003	0.006	0.096	0.160	0.087	0.086	0.156	0.151	0.920	0.916	0.940	0.928
		Pseudo	0.002	0.008	0.096	0.156	0.087	0.086	0.156	0.152	0.916	0.916	0.944	0.928
	t_2	$IPCW_1$	0.008	0.001	0.176	0.323	0.173	0.172	0.306	0.299	0.950	0.950	0.922	0.928
		$IPCW_2$	0.004	0.013	0.124	0.224	0.119	0.118	0.213	0.208	0.950	0.936	0.928	0.924
		DR(Cox)	0.004	0.015	0.105	0.191	0.102	0.101	0.182	0.177	0.952	0.946	0.938	0.930
		DR(lognorm)	0.004	0.015	0.105	0.191	0.102	0.101	0.182	0.178	0.956	0.952	0.938	0.932
		DR(loglog)	0.004	0.015	0.105	0.191	0.102	0.101	0.182	0.178	0.954	0.946	0.940	0.930
		Pseudo	0.004	0.013	0.105	0.186	0.102	0.101	0.183	0.178	0.954	0.944	0.946	0.936
	t_3	$IPCW_1$	−0.007	−0.020	0.163	0.297	0.151	0.150	0.263	0.257	0.936	0.932	0.884	0.884
		$IPCW_2$	−0.007	−0.015	0.129	0.239	0.122	0.120	0.214	0.208	0.930	0.938	0.894	0.892
		DR(Cox)	−0.006	−0.008	0.104	0.199	0.097	0.097	0.174	0.168	0.938	0.930	0.918	0.900
		DR(lognorm)	−0.006	−0.008	0.105	0.200	0.098	0.097	0.174	0.169	0.932	0.930	0.920	0.900
		DR(loglog)	−0.006	−0.008	0.105	0.199	0.098	0.097	0.174	0.169	0.936	0.932	0.918	0.902
		Pseudo	−0.005	−0.011	0.105	0.191	0.098	0.097	0.175	0.170	0.934	0.932	0.930	0.916
$n = 1000$	t_1	$IPCW_1$	−0.002	0.006	0.126	0.238	0.122	0.121	0.217	0.215	0.940	0.944	0.934	0.932
		$IPCW_2$	−0.004	−0.000	0.073	0.130	0.070	0.070	0.124	0.123	0.944	0.948	0.940	0.940
		DR(Cox)	−0.001	−0.002	0.069	0.120	0.061	0.061	0.109	0.108	0.924	0.926	0.932	0.934
		DR(lognorm)	−0.001	−0.001	0.069	0.120	0.061	0.061	0.109	0.108	0.922	0.926	0.936	0.936
		DR(loglog)	−0.001	−0.001	0.069	0.120	0.061	0.061	0.109	0.108	0.922	0.926	0.934	0.936
		Pseudo	−0.001	−0.002	0.069	0.120	0.061	0.061	0.109	0.108	0.924	0.924	0.932	0.934
	t_2	$IPCW_1$	0.000	0.003	0.126	0.250	0.122	0.121	0.219	0.216	0.948	0.944	0.910	0.908
		$IPCW_2$	0.002	0.001	0.082	0.161	0.084	0.083	0.150	0.148	0.966	0.960	0.938	0.942
		DR(Cox)	0.001	−0.002	0.075	0.136	0.071	0.071	0.127	0.125	0.940	0.940	0.940	0.940
		DR(lognorm)	0.001	−0.002	0.075	0.136	0.071	0.071	0.127	0.126	0.940	0.938	0.940	0.940
		DR(loglog)	0.001	−0.002	0.074	0.136	0.071	0.071	0.127	0.126	0.940	0.938	0.942	0.940
		Pseudo	0.001	−0.002	0.074	0.136	0.071	0.071	0.127	0.126	0.940	0.938	0.942	0.940
	t_3	$IPCW_1$	0.003	0.004	0.115	0.216	0.108	0.108	0.190	0.188	0.942	0.946	0.932	0.920
		$IPCW_2$	0.001	0.002	0.086	0.157	0.086	0.085	0.151	0.150	0.952	0.944	0.950	0.938
		DR(Cox)	0.000	−0.001	0.072	0.128	0.068	0.068	0.121	0.120	0.950	0.950	0.936	0.938
		DR(lognorm)	0.001	−0.001	0.072	0.129	0.069	0.068	0.122	0.121	0.944	0.946	0.938	0.942
		DR(loglog)	0.001	−0.001	0.072	0.129	0.068	0.068	0.121	0.121	0.950	0.946	0.938	0.938
		Pseudo	0.001	−0.001	0.072	0.129	0.069	0.068	0.122	0.121	0.944	0.950	0.938	0.934

$IPCW_1$: (2.2), $IPCW_2$: (2.3), DR(Cox): (2.4) with calculating Q by Cox model, DR(lognorm): (2.4) with calculating Q with AFT model from lognormal distribution, DR(loglog): (2.4) with calculating Q with AFT model from log-logistic distribution, Pseudo : pseudo-value approach by Anderson *et al.* (2003), HC0: plug-in, NN: nearest neighbor, LM : bandwidth selection by Ludwig and Miller (2007), MSE : bandwidth selection by MSE optimization in Calonico *et al.* (2014).

For variance estimation, we propose an estimation adapted by Cho *et al.* (2021). By applying asymptotic result in the Appendix A (Cho *et al.*, 2021), given h, the asymptotic variance of $\hat{\tau}_{DR}(t)$ is

$$\Sigma^{DR}_{SRD}(G_0, S^*)(t) = \frac{1}{n} e_1^T \left(\mathbf{\Gamma}^{-1}_{h+} \boldsymbol{\phi}_{VV+,DR}(t) \mathbf{\Gamma}^{-1}_{h+} + \mathbf{\Gamma}^{-1}_{h-} \boldsymbol{\phi}_{VV-,DR}(t) \mathbf{\Gamma}^{-1}_{h-} \right) e_1,$$

where $\mathbf{\Gamma}_{h+}$ and $\mathbf{\Gamma}_{h-}$, $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$ are defined in Appendix, and $e_1 = (1, 0)^T$.

It is crucial to estimate $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$ for variance estimation. As discussed in Cho *et al.* (2021), we use two estimation methods for $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$: Plug-in and nearest neighbor (NN) methods (Calonico *et al.*, 2014). In the plug-in approach, we define usual residuals from DR

Table 2: Simulation results with data generated from Cox model with 60% censoring rate

			Bias		ESD		SE				Cover			
			LM	MSE	LM	MSE	LM		MSE		LM		MSE	
							NN	HC0	NN	HC0	NN	HC0	NN	HC0
$n = 500$	t_1	$IPCW_1$	0.002	−0.008	0.371	0.698	0.358	0.355	0.622	0.601	0.938	0.936	0.936	0.924
		$IPCW_2$	0.002	0.016	0.141	0.245	0.131	0.129	0.234	0.227	0.934	0.934	0.934	0.932
		DR(Cox)	0.005	0.010	0.098	0.172	0.092	0.091	0.165	0.160	0.922	0.926	0.932	0.930
		DR(lognorm)	0.005	0.010	0.098	0.172	0.092	0.091	0.165	0.160	0.924	0.928	0.932	0.930
		DR(loglog)	0.005	0.010	0.098	0.172	0.092	0.091	0.165	0.160	0.924	0.928	0.932	0.932
		Pseudo	0.005	0.010	0.098	0.172	0.092	0.091	0.165	0.160	0.922	0.926	0.934	0.926
	t_2	$IPCW_1$	0.001	−0.007	0.400	0.683	0.351	0.347	0.595	0.576	0.948	0.948	0.910	0.904
		$IPCW_2$	−0.005	0.008	0.217	0.411	0.202	0.201	0.358	0.349	0.946	0.936	0.946	0.944
		DR(Cox)	−0.001	0.008	0.128	0.238	0.123	0.122	0.220	0.214	0.934	0.932	0.914	0.914
		DR(lognorm)	−0.000	0.009	0.128	0.238	0.123	0.122	0.221	0.214	0.936	0.934	0.918	0.916
		DR(loglog)	−0.001	0.009	0.128	0.238	0.123	0.122	0.220	0.214	0.938	0.932	0.910	0.916
		Pseudo	0.000	0.009	0.130	0.238	0.124	0.123	0.221	0.215	0.934	0.932	0.922	0.920
$n = 1000$	t_1	$IPCW_1$	−0.004	−0.000	0.278	0.491	0.257	0.256	0.449	0.445	0.940	0.938	0.934	0.928
		$IPCW_2$	−0.001	0.005	0.095	0.176	0.092	0.091	0.163	0.161	0.938	0.940	0.944	0.942
		DR(Cox)	0.000	0.002	0.070	0.123	0.064	0.064	0.114	0.113	0.940	0.932	0.948	0.946
		DR(lognorm)	0.000	0.002	0.070	0.123	0.064	0.064	0.114	0.113	0.942	0.932	0.948	0.950
		DR(loglog)	0.000	0.002	0.070	0.123	0.064	0.064	0.114	0.113	0.940	0.932	0.950	0.948
		Pseudo	0.001	0.002	0.070	0.123	0.064	0.064	0.114	0.113	0.938	0.932	0.948	0.948
	t_2	$IPCW_1$	−0.004	−0.003	0.265	0.475	0.251	0.249	0.437	0.432	0.958	0.960	0.936	0.936
		$IPCW_2$	−0.004	−0.003	0.149	0.281	0.144	0.143	0.255	0.252	0.948	0.948	0.924	0.928
		DR(Cox)	−0.001	−0.004	0.087	0.164	0.087	0.086	0.155	0.153	0.950	0.950	0.946	0.942
		DR(lognorm)	0.000	−0.003	0.087	0.164	0.087	0.087	0.155	0.153	0.948	0.946	0.944	0.942
		DR(loglog)	−0.000	−0.003	0.087	0.164	0.087	0.086	0.155	0.153	0.948	0.950	0.946	0.942
		Pseudo	−0.000	−0.004	0.087	0.165	0.087	0.087	0.156	0.153	0.950	0.950	0.944	0.942

$IPCW_1$: (2.2), $IPCW_2$: (2.3), DR(Cox): (2.4) with calculating Q by Cox model, DR(lognorm): (2.4) with calculating Q with AFT model from lognormal distribution, DR(loglog): (2.4) with calculating Q with AFT model from log-logistic distribution, Pseudo : pseudo-value approach by Anderson *et al.* (2003), HC0: plug-in, NN: nearest neighbor, LM : bandwidth selection by Ludwig and Miller (2007), MSE : bandwidth selection by MSE optimization in Calonico *et al.* (2014).

transformed outcome and $\{\hat{\alpha}_{R,DR}(t), \hat{\alpha}_{L,DR}(t)\}$. Then by using these residuals to estimate $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$ and compute empirical version of $\Sigma^{DR}_{SRD}(G_0, S^*)(t)$. NN method uses residuals by distance in each observation i. To reduce the influence of outliers in variance estimation, we define the closest values on each individual transformed DR outcome. Then we compute residuals based on the average of the closest values and the DR outcome, and then estimate $\Sigma^{DR}_{SRD}(G_0, S^*)(t)$. We can also apply these two approaches to $IPCW_1$ and $IPCW_2$ estimators. Details of the derivation of the variance are shown in the Appendix B.

3.2. Causal effect estimation and inference with the pseudo-values

Now we extend this idea to another type of unbiased estimator of $P(T > t|W = w)$ with local linear regression. One of the methods in survival data to directly model survival quantities is the pseudo-value approach (Anderson *et al.*, 2003) for our method. Let θ be a scalar parameter. Let $X_1, \ldots, X_n$ be independent and identically distributed data, and f be a function such that $E\{f(X_i)\} = \theta$. Suppose that we have $A_1, \ldots, A_n$ independent and identically distributed covariates. We define conditional expectation $\theta_i = E\{f(X_i)|A_i\}$, which gives θ when we integrate θ_i with respect to A_i. Then we define pseudo-value for i^{th} observation by

$$\hat{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}^{-i}.$$

Table 3: Simulation results with data generated from additive hazard model with 30% censoring rate

			Bias		ESD		SE				Cover			
			LM	MSE	LM	MSE	LM		MSE		LM		MSE	
							NN	HC0	NN	HC0	NN	HC0	NN	HC0
$n = 500$	t_1	$IPCW_1$	0.002	−0.001	0.175	0.330	0.170	0.169	0.303	0.295	0.948	0.948	0.922	0.922
		$IPCW_2$	0.004	0.005	0.107	0.179	0.102	0.101	0.183	0.177	0.934	0.932	0.952	0.948
		DR(Cox)	0.003	0.003	0.097	0.159	0.089	0.088	0.159	0.154	0.936	0.930	0.950	0.942
		DR(lognorm)	0.003	0.004	0.097	0.158	0.089	0.088	0.159	0.154	0.936	0.930	0.950	0.944
		DR(loglog)	0.003	0.004	0.097	0.158	0.089	0.088	0.159	0.154	0.936	0.930	0.950	0.942
		Pseudo	0.003	0.004	0.097	0.158	0.089	0.088	0.159	0.154	0.936	0.930	0.950	0.940
	t_2	$IPCW_1$	0.004	0.010	0.182	0.324	0.171	0.170	0.304	0.297	0.948	0.938	0.924	0.924
		$IPCW_2$	0.005	0.017	0.132	0.231	0.122	0.121	0.219	0.213	0.930	0.930	0.930	0.926
		DR(Cox)	0.003	0.016	0.117	0.196	0.105	0.105	0.188	0.183	0.926	0.918	0.944	0.942
		DR(lognorm)	0.003	0.016	0.117	0.196	0.105	0.104	0.188	0.183	0.926	0.918	0.944	0.940
		DR(loglog)	0.003	0.016	0.117	0.196	0.105	0.105	0.188	0.183	0.926	0.918	0.942	0.942
		Pseudo	0.003	0.017	0.117	0.196	0.106	0.105	0.188	0.183	0.922	0.916	0.946	0.944
	t_3	$IPCW_1$	−0.004	−0.009	0.162	0.292	0.148	0.147	0.259	0.253	0.940	0.948	0.898	0.898
		$IPCW_2$	−0.006	−0.010	0.129	0.242	0.121	0.120	0.216	0.210	0.928	0.936	0.914	0.910
		DR(Cox)	−0.003	−0.006	0.106	0.201	0.100	0.099	0.178	0.173	0.952	0.950	0.926	0.918
		DR(lognorm)	−0.003	−0.006	0.107	0.202	0.100	0.099	0.179	0.174	0.950	0.948	0.916	0.920
		DR(loglog)	−0.002	−0.006	0.107	0.201	0.100	0.099	0.178	0.173	0.952	0.948	0.918	0.918
		Pseudo	−0.002	−0.005	0.107	0.201	0.100	0.099	0.178	0.173	0.952	0.954	0.922	0.920
$n = 1000$	t_1	$IPCW_1$	−0.004	−0.004	0.124	0.238	0.121	0.120	0.215	0.212	0.956	0.948	0.948	0.938
		$IPCW_2$	−0.005	−0.005	0.075	0.134	0.071	0.071	0.127	0.125	0.924	0.924	0.950	0.946
		DR(Cox)	−0.003	−0.006	0.070	0.121	0.062	0.062	0.111	0.110	0.926	0.926	0.934	0.932
		DR(lognorm)	−0.003	−0.006	0.070	0.122	0.062	0.062	0.111	0.110	0.928	0.922	0.938	0.936
		DR(loglog)	−0.003	−0.006	0.070	0.122	0.062	0.062	0.111	0.110	0.926	0.922	0.932	0.934
		Pseudo	−0.003	−0.006	0.070	0.121	0.062	0.062	0.111	0.110	0.926	0.926	0.934	0.936
	t_2	$IPCW_1$	−0.000	−0.005	0.129	0.243	0.122	0.121	0.218	0.215	0.942	0.940	0.926	0.920
		$IPCW_2$	0.002	−0.001	0.087	0.164	0.086	0.086	0.155	0.153	0.962	0.956	0.940	0.942
		DR(Cox)	0.000	−0.006	0.076	0.139	0.074	0.073	0.132	0.130	0.944	0.942	0.946	0.946
		DR(lognorm)	0.000	−0.006	0.076	0.139	0.074	0.073	0.132	0.130	0.940	0.942	0.950	0.946
		DR(loglog)	0.000	−0.006	0.076	0.139	0.074	0.073	0.132	0.130	0.944	0.942	0.948	0.948
		Pseudo	0.000	−0.006	0.076	0.139	0.074	0.074	0.132	0.130	0.942	0.942	0.948	0.944
	t_3	$IPCW_1$	−0.001	−0.008	0.112	0.206	0.106	0.106	0.187	0.184	0.942	0.936	0.916	0.918
		$IPCW_2$	−0.001	−0.006	0.090	0.161	0.086	0.086	0.152	0.150	0.938	0.942	0.960	0.950
		DR(Cox)	−0.001	−0.006	0.074	0.131	0.070	0.070	0.124	0.123	0.942	0.946	0.940	0.940
		DR(lognorm)	−0.001	−0.006	0.076	0.132	0.071	0.070	0.125	0.124	0.934	0.938	0.942	0.934
		DR(loglog)	−0.001	−0.006	0.075	0.131	0.070	0.070	0.125	0.124	0.938	0.940	0.938	0.936
		Pseudo	−0.001	−0.006	0.075	0.132	0.070	0.070	0.125	0.124	0.940	0.946	0.936	0.938

$IPCW_1$: (2.2), $IPCW_2$: (2.3), DR(Cox): (2.4) with calculating Q by Cox model, DR(lognorm): (2.4) with calculating Q with AFT model from lognormal distribution, DR(loglog): (2.4) with calculating Q with AFT model from log-logistic distribution, Pseudo : pseudo-value approach by Anderson *et al.* (2003), HC0: plug-in, NN: nearest neighbor, LM : bandwidth selection by Ludwig and Miller (2007), MSE : bandwidth selection by MSE optimization in Calonico *et al.* (2014).

For the survival function, $f(t) = I(T > t)$. Then our pseudo-values for fixed time t are

$$\hat{\theta}_i(t) = n\hat{\theta}(t) - (n-1)\hat{\theta}^{-i}(t),$$

where $\hat{\theta}(t)$ is the Kaplan-Meier estimator at time t and $\hat{\theta}^{-i}(t)$ is the leave-one-out Kaplan-Meier estimator.

Since the pseudo-value approach is based on conditional expectation θ_i, it is sensible to apply the approach in our RD design. In other words, we compute the pseudo-values and use them as outcomes for each individual. Next, we do local linear regression by plugging in $\hat{\theta}_i(t)$ in the place of $\hat{V}_{CUT,i}(t)$ as

Table 4: Simulation results with data generated from additive hazard model with 50% censoring rate

			Bias		ESD		SE				Cover			
			LM	MSE	LM	MSE	LM		MSE		LM		MSE	
							NN	HC0	NN	HC0	NN	HC0	NN	HC0
$n = 500$	t_1	$IPCW_1$	−0.008	−0.032	0.301	0.535	0.282	0.279	0.493	0.477	0.940	0.938	0.936	0.922
		$IPCW_2$	−0.001	0.004	0.128	0.215	0.119	0.118	0.213	0.206	0.920	0.922	0.954	0.942
		DR(Cox)	0.004	0.005	0.098	0.164	0.091	0.090	0.163	0.158	0.924	0.928	0.938	0.932
		DR(lognorm)	0.005	0.005	0.097	0.163	0.091	0.090	0.163	0.158	0.930	0.930	0.940	0.932
		DR(loglog)	0.005	0.005	0.098	0.163	0.091	0.090	0.163	0.158	0.928	0.928	0.936	0.932
		Pseudo	0.005	0.005	0.098	0.164	0.091	0.090	0.163	0.158	0.924	0.930	0.938	0.932
	t_2	$IPCW_1$	−0.011	−0.022	0.294	0.522	0.276	0.274	0.480	0.466	0.950	0.944	0.914	0.916
		$IPCW_2$	0.003	0.011	0.174	0.311	0.159	0.158	0.282	0.274	0.924	0.920	0.922	0.928
		DR(Cox)	0.002	0.017	0.127	0.213	0.115	0.114	0.205	0.200	0.924	0.918	0.946	0.942
		DR(lognorm)	0.002	0.016	0.127	0.213	0.115	0.114	0.205	0.200	0.918	0.916	0.948	0.942
		DR(loglog)	0.002	0.017	0.127	0.213	0.115	0.114	0.205	0.200	0.922	0.920	0.948	0.946
		Pseudo	0.002	0.017	0.127	0.213	0.116	0.115	0.205	0.200	0.920	0.916	0.946	0.944
$n = 1000$	t_1	$IPCW_1$	−0.004	0.001	0.211	0.381	0.200	0.199	0.354	0.352	0.932	0.936	0.934	0.938
		$IPCW_2$	−0.004	−0.005	0.087	0.162	0.084	0.083	0.149	0.147	0.944	0.940	0.936	0.938
		DR(Cox)	−0.003	−0.005	0.070	0.123	0.064	0.064	0.114	0.112	0.926	0.932	0.926	0.934
		DR(lognorm)	−0.003	−0.005	0.070	0.123	0.064	0.064	0.114	0.112	0.930	0.930	0.926	0.934
		DR(loglog)	−0.003	−0.005	0.070	0.123	0.064	0.064	0.114	0.112	0.926	0.930	0.926	0.934
		Pseudo	−0.003	−0.005	0.070	0.123	0.064	0.064	0.114	0.112	0.928	0.932	0.928	0.934
	t_2	$IPCW_1$	−0.001	−0.002	0.210	0.388	0.197	0.196	0.347	0.345	0.952	0.954	0.926	0.926
		$IPCW_2$	−0.002	−0.004	0.117	0.220	0.112	0.112	0.201	0.199	0.946	0.948	0.940	0.936
		DR(Cox)	−0.001	−0.007	0.081	0.152	0.081	0.081	0.145	0.143	0.950	0.950	0.942	0.942
		DR(lognorm)	−0.000	−0.007	0.081	0.152	0.081	0.081	0.145	0.143	0.946	0.948	0.944	0.940
		DR(loglog)	−0.000	−0.007	0.081	0.152	0.081	0.081	0.145	0.143	0.946	0.950	0.940	0.940
		Pseudo	−0.000	−0.007	0.081	0.152	0.081	0.081	0.145	0.143	0.950	0.946	0.944	0.942

$IPCW_1$: (2.2), $IPCW_2$: (2.3), DR(Cox): (2.4) with calculating Q by Cox model, DR(lognorm): (2.4) with calculating Q with AFT model from lognormal distribution, DR(loglog): (2.4) with calculating Q with AFT model from log-logistic distribution, Pseudo : pseudo-value approach by Anderson *et al.* (2003), HC0: plug-in, NN: nearest neighbor, LM : bandwidth selection by Ludwig and Miller (2007), MSE : bandwidth selection by MSE optimization in Calonico *et al.* (2014).

shown in (3.1) and (3.2), along with bandwidth selection methods in (3.8) and (3.9). Then we obtain $\alpha_R(t)$ and $\alpha_L(t)$, say $\hat{\alpha}_{R,Pseudo,\hat{h}}(t)$ and $\hat{\alpha}_{L,Pseudo,\hat{h}}(t)$, and our estimator is

$$\hat{\tau}_{Pseudo}(t) = \hat{\alpha}_{R,Pseudo,\hat{h}}(t) - \hat{\alpha}_{L,Pseudo,\hat{h}}(t).$$

Statistical inference corresponding to $\hat{\tau}_{Pseudo}(t)$ can be performed similarly to IPCW and DR estimators.

4. Simulation

We do various Monte Carlo simulations to evaluate the numerical performance of our proposed method with the finite sample. We generate $W \sim \text{Unif}(0, 1)$. Our first model is the Cox proportional hazard model (Cox, 1972), which is

$$\lambda(t \mid W) = e^{\beta I(W \geq 0.5)},$$

where $\beta = -1$. Censoring variable C follows $\text{Unif}(0, b)$ where b is taken to achieve 30% censoring rate. Sample sizes are 500 and 1000. We compute the 25^{th}, 50^{th} and 75^{th} percentile of failure times by using Monte Carlo simulations. We implement and compare the performance of five methods:

Two inverse probability censoring weighted (IPCW) methods and doubly robust methods proposed in Section 2 with 3 outcome regression models (Cox model, AFT model with lognormal distribution and AFT model with log-logistic distribution). We denote using transformation in (2.2) and (2.3) as IPCW_1 and IPCW_2. Moreover, we call doubly robust transformation in (2.4) by calculating Q by Cox model, AFT models from lognormal distribution and log-logistic distributions by DR(Cox), DR(lognorm) and DR(loglog), respectively. We also apply the pseudo-value approach (Pseudo) by Anderson *et al*. (2003) as described in Section 4. In each method, for bandwidth estimation, we use approaches by Ludwig and Miller (2007) (denoted as LM) and MSE optimization in Calonico *et al*. (2014) (denoted as MSE). For the computation of standard error, we use the plug-in (HC0) and nearest neighbor (NN) methods mentioned in the previous section. We compute bias (Bias), empirical standard deviation (ESD), mean of standard error (SEE) and 95% coverage rate (Cover). We use the `rdrobust` package Calonico *et al*. (2015) in `R` to compute estimators and standard errors.

The simulation result from these three conditional expectations is shown in Table 1. All estimators are nearly unbiased. DR estimators are more efficient than IPCW_1 and IPCW_2 estimators. The efficiency gain of DR estimators compared to IPCW_1 is larger than that of IPCW_2. It is interesting to note that the pseudo-value approach has the similar performance with DR estimators. All methods have satisfactory large sample properties in general although all estimators in t_3 for $n = 500$ have a lower coverage rate compared to the nominal 95% level. However, when the sample size increases, all methods achieve the nominal coverage rate. It is interesting to notice that both ESD and SEE from the LM method are smaller than the MSE method although biases from the two bandwidth choices are not very large. The ways to estimate $\sigma^2_{DR,+}(t; W_i, G_0, S^*)$ and $\sigma^2_{DR,-}(t; W_i, G_0, S^*)$ do not influence the estimation of variance; both NN and HC0 provide similar values to the SEE.

As shown in the data analysis section, since our data has a high censoring rate, we run another simulation with 60% censoring rate. Table 2 shows the result. In this simulation, we only consider the 25^{th} and 50^{th} percentile of the failure time due to a high censoring rate. ESD and SEE are higher than the 30% censoring rate, which is sensible due to a higher loss of information than the 30% censoring case. The general trend of simulation result is similar to one in the 30% censoring rate.

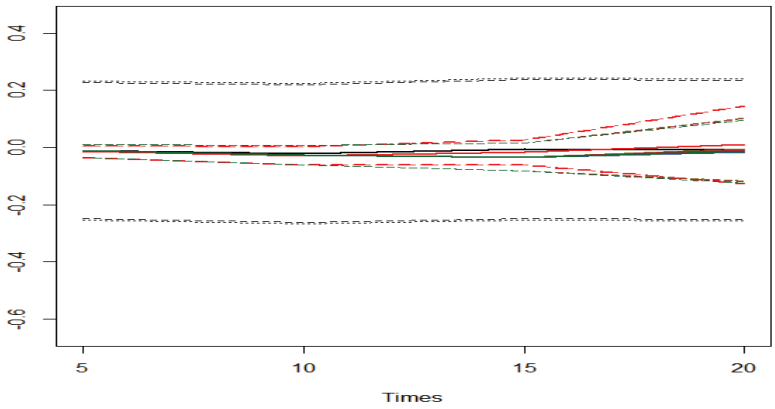
Now we consider another simulation setting. Let hazard function $\lambda(t|W)$ be

$$\lambda(t \mid W) = 2 + \beta I\left(W \geq 0.5\right),$$

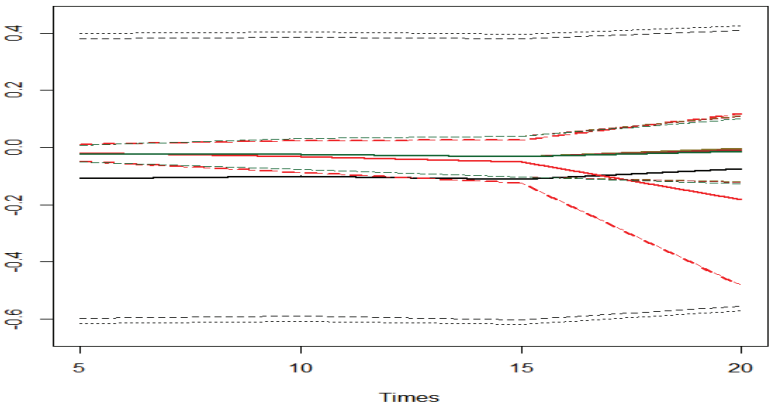
where $\beta = -1$. This model is the additive hazard model (Lin and Ying, 1994), which is widely used in causal inference for survival analysis. Similar to the Cox model, we generate censoring variable C following $\text{Unif}(0, b)$, where b is chosen to achieve 30% censoring rate. As same procedure we did in the Cox model, we calculate the 25^{th}, 50^{th} and 75^{th} percentiles of failure time. The simulation result is shown in Table 3. We can see a similar trend; all methods have good large sample properties. We also run simulation with 50% censoring rate. Similar to Cox model, we only consider 25^{th} and 50^{th} percentiles. Table 4 shows the simulation result. The result is similar to one in 30% censoring rate, except containing a higher ESD and SSE.

5. Real data analysis

We apply our method to the PLCO dataset focusing on prostate cancer to evaluate whether a PSA-based screening strategy can be a meaningful tool to diagnose survival. In this PLCO trial, from 1993 to 2001, 76,678 men were randomized to receive annual PSA screening for 6 years or no PSA screening at all. The previous study of the PLCO trial has shown that there had been no significant decrease of mortality or prostate cancer-specific incidence although detection of prostate cancer had



(a)



(b)

Figure 1: *Estimated treatment effects (solid lines) and their 95% confidence interval (nearest neighbor method: Dased lines, plug-in method: Dotted lines). Left and right plots are methods with LM and MSE, respectively. In each plot, the colors imply as follows - Black : IPCW$_1$, red : IPCW$_2$, blue : DR(Cox), gold : DR(lognorm), brown : DR(loglog), darkgreen : Pseudo.*

been increased (Andriole *et al*., 2009). For further workup in this trial, the golden rule is PSA level 4.0ng/ml, which implies that clinicians would expect to find the difference in mortality or cancer incidence on the threshold 4.0ng/ml. The question is "Is there any difference of survival probability at the PSA level greater or equal to 4.0mg/nl and less than 4.0mg/nl?" If the difference exists, then the PSA level 4.0mg/nl can be used as threshold for further checkup and biopsy (Cho *et al*., 2021) As explained in Cho *et al*. (2021) and the Introduction section, the motivation and this trial setup naturally create the sharp RD design. Cho *et al*. (2021) focus on the difference in mean of logarithm of time for the occurrence of cancer, whichever comes first (first cancer occurrence) or for prostate

Table 5: Sharp RD design analysis result for mortality in the PLCO dataset with the screening group only

		Est		95% CI			
		LM	MSE	LM		MSE	
				NN	HC0	NN	HC0
t_1	$IPCW_1$	−0.010	−0.108	(−0.249, 0.228)	(−0.254, 0.234)	(−0.596, 0.381)	(−0.616, 0.400)
	$IPCW_2$	−0.013	−0.019	(−0.033, 0.007)	(−0.033, 0.007)	(−0.049, 0.011)	(−0.049, 0.010)
	DR(Cox)	−0.012	−0.021	(−0.033, 0.010)	(−0.033, 0.010)	(−0.051, 0.009)	(−0.051, 0.008)
	DR(lognorm)	−0.012	−0.021	(−0.033, 0.010)	(−0.033, 0.010)	(−0.051, 0.009)	(−0.051, 0.008)
	DR(loglog)	−0.012	−0.021	(−0.033, 0.010)	(−0.033, 0.010)	(−0.051, 0.009)	(−0.051, 0.008)
	Pseudo	−0.012	−0.021	(−0.033, 0.010)	(−0.033, 0.010)	(−0.051, 0.009)	(−0.051, 0.008)
t_2	$IPCW_1$	−0.020	−0.101	(−0.261, 0.220)	(−0.267, 0.226)	(−0.590, 0.387)	(−0.609, 0.406)
	$IPCW_2$	−0.029	−0.032	(−0.060, 0.003)	(−0.060, 0.003)	(−0.086, 0.023)	(−0.086, 0.023)
	DR(Cox)	−0.025	−0.022	(−0.060, 0.009)	(−0.060, 0.009)	(−0.077, 0.032)	(−0.076, 0.031)
	DR(lognorm)	−0.025	−0.022	(−0.060, 0.009)	(−0.060, 0.009)	(−0.077, 0.032)	(−0.076, 0.032)
	DR(loglog)	−0.025	−0.022	(−0.060, 0.009)	(−0.060, 0.009)	(−0.077, 0.032)	(−0.076, 0.032)
	Pseudo	−0.025	−0.022	(−0.060, 0.009)	(−0.060, 0.009)	(−0.077, 0.032)	(−0.076, 0.031)
t_3	$IPCW_1$	−0.005	−0.111	(−0.249, 0.238)	(−0.253, 0.243)	(−0.602, 0.380)	(−0.617, 0.396)
	$IPCW_2$	−0.016	−0.049	(−0.060, 0.028)	(−0.060, 0.028)	(−0.124, 0.026)	(−0.124, 0.025)
	DR(Cox)	−0.033	−0.031	(−0.083, 0.017)	(−0.082, 0.016)	(−0.103, 0.040)	(−0.102, 0.039)
	DR(lognorm)	−0.033	−0.031	(−0.083, 0.017)	(−0.082, 0.016)	(−0.103, 0.040)	(−0.102, 0.039)
	DR(loglog)	−0.033	−0.031	(−0.083, 0.016)	(−0.082, 0.016)	(−0.103, 0.040)	(−0.102, 0.039)
	Pseudo	−0.033	−0.031	(−0.083, 0.017)	(−0.082, 0.016)	(−0.103, 0.040)	(−0.102, 0.039)
t_4	$IPCW_1$	−0.008	−0.073	(−0.251, 0.235)	(−0.256, 0.240)	(−0.555, 0.410)	(−0.571, 0.426)
	$IPCW_2$	0.010	−0.182	(−0.125, 0.145)	(−0.126, 0.145)	(−0.482, 0.117)	(−0.484, 0.119)
	DR(Cox)	−0.014	−0.013	(−0.123, 0.094)	(−0.123, 0.094)	(−0.127, 0.101)	(−0.127, 0.101)
	DR(lognorm)	−0.006	−0.003	(−0.116, 0.104)	(−0.116, 0.104)	(−0.118, 0.112)	(−0.118, 0.112)
	DR(loglog)	−0.009	−0.007	(−0.118, 0.100)	(−0.118, 0.100)	(−0.121, 0.108)	(−0.121, 0.108)
	Pseudo	−0.014	−0.013	(−0.123, 0.095)	(−0.123, 0.095)	(−0.127, 0.101)	(−0.127, 0.101)

$IPCW_1$: (2.2), $IPCW_2$: (2.3), DR(Cox): (2.4) with calculating Q by Cox model, DR(lognorm): (2.4) with calculating Q with AFT model from lognormal distribution, DR(loglog): (2.4) with calculating Q with AFT model from lognormal distribution log-logistic distributions, HC0: plug-in, NN: nearest neighbor, LM : bandwidth selection by Ludwig and Miller (2007), MSE : bandwidth selection by MSE optimization in Calonico *et al.* (2014).

cancer (prostate cancer occurrence). In this analysis, we want to find whether there exists a difference between disease-free survival at the PSA level 4.0mg/nl.

In this data analysis, we focus on the screening group with baseline PSA level which was measured before PSA screening. The censoring rate is 66%, so it is not possible to compute higher percentile times by Kaplan-Meier. Hence we use observed times $t_1 = 5, t_2 = 10, t_3 = 15$ and $t_4 = 20$.

We calculate the causal effect and the associated 95% confidence interval. The results are shown in Table 5 and Figure 1. The causal effect is similar between the LM and MSE methods, but the 95% confidence interval from the MSE method is wider than the LM method, as observed in the simulation studies. From the 95% confidence interval, we recognize that the $IPCW_1$ estimator shows wide variability as simulation studies. All DR estimators and the pseudo-value based estimator have smaller variances than $IPCW_1$ and $IPCW_2$. All methods show that the intervals in the later time points are wider than the ones in the earlier. This is sensible due to the high rate of censoring. It is interesting to note that the 95% confidence interval $IPCW_2$ estimator from the MSE method is much wider in the t_4 than earlier time points. Since all 95% confidence intervals include 0, we can conclude that there is no significant effect of survival difference using PSA level 4.0ng/ml.

6. Conclusion

In this paper, we propose the estimation of treatment effects in the survival data for the survival function. This methodology is practically more meaningful than the one in Cho *et al.* (2021) in that the survival function is more relevant application than using restricted mean survival time.

We use IPCW and DR robust transformation to account for censoring. However, the survival function is between 0 and 1, so directly applying local linear regression may cause an issue because the predicted values do not belong to 0 and 1. It is an interesting future research to the survival model directly without this restriction.

As discussed in Section 3, there is a bias term involving the second derivative of $\lim_{w\downarrow w_0} P(T > t|W = w)$ and $\lim_{w\uparrow w_0} P(T > t|W = w)$. Let $\Lambda_0(u|W)$ be the covariate-dependent cumulative hazard function of T. If limit and differentiation are interchanged, since $P(T > t|W = w) = e^{-\Lambda_0(u|w)}$, the second derivative of $\lim_{w\to w_0} P(T > t|W = w)$ can be expressed

$$\begin{aligned}
&\lim_{w\downarrow w_0}\left[\left(\frac{\partial}{\partial t}\lambda_0(t\mid w)\right)+\{\lambda_0(t\mid w)\}^2\right]e^{-\Lambda_0(t|w)}\\
&\lim_{w\uparrow w_0}\left[\left(\frac{\partial}{\partial t}\lambda_0(t\mid w)\right)+\{\lambda_0(t\mid w)\}^2\right]e^{-\Lambda_0(t|w)}.
\end{aligned}\tag{6.1}$$

When (6.1) goes 0 in given t and when $h \propto n^{-a}$ where $1/5 < a < 2/5$ (Imbens and Lemieux, 2008), i.e., when we do undersmoothing, the bias term will be negligible. However, undersmoothing may cause increase in variance due to bias-variance tradeoff. Calonico *et al.* (2014) propose bias-corrected estimator for uncensored data. It is an interesting future research to reduce bias in the estimation of RD design for censored data.

In practice, one would like to include covariates in the modeling. In a randomized study, including covariates leads to the treatment effect being more powerful by reducing the variance of the treatment effect. In RD design, the inclusion of covariates is discussed in uncensored data (Calonico *et al.*, 2019), but it has not been studied deeply in censored data. It would be an interesting research to adjust the effect of covariates in RD design.

In the simulation study, we have seen that the pseudo-value approach has similar performance to DR method. Moreover, in the PLCO dataset analysis, the pseudo-value method has almost identical estimated value and confidence interval to DR(Cox). It will be an interesting future research why the pseudo-value approach shows the similar performance. In the PLCO dataset, one may be interested in the analysis of prostate cancer incidence given that other cancers exist. This is competing risks setup. Our methodology can be extended to competing risks data, but more assumptions are required to identify treatment effects. One may consider using cause-specific hazard. However, in the view of practical and identifiable quantity, cumulative incidence function will be a reasonable one for identification of treatment effect. This can also be an interesting future research.

Acknowledgment

This paper was supported by Konkuk University in 2021 (2021-A019-0164).

Appendix A: Proof of Theorem 1

Let $O = (\tilde{T}, \Delta, W)$ and $\mathcal{O} = (\tilde{T}_i, \Delta_i, W_i), i = 1, \ldots, n$. Define

$$\begin{aligned}
\mu^+(t; w) &= \lim_{w \downarrow w_0} P(T > t \mid W = w), \quad \mu^-(t; w) = \lim_{w \uparrow w_0} P(T > t \mid W = w), \\
V_{DR}(t; O, G, S) &= \frac{\Delta(t) I(T > t)}{G(T(t))} + \int_0^{\tilde{T}} \frac{Q(t, u \mid W, S)}{G(u)} dM_G(u), \\
V_{DR^*}(t; O, G, S) &= V_{DR}(t; O, G, S) - \mu^+(t; w_0) - \mu'^+(t; w_0)(W - w_0), \\
L_{ih}^+ &= I(W_i \geq w_0) K\Big(\frac{W_i - w_0}{h}\Big), \quad L_{ih}^- = I(W_i < w_0) K\Big(\frac{W_i - w_0}{h}\Big), \\
\sigma^2_{DR}(t; w, G, S) &= \mathrm{Var}(V_{DR}(t; O, G, S) \mid W = w), \\
\sigma^{2+}_{DR}(t; w_0, G, S) &= \lim_{\epsilon \downarrow w_0} \mathrm{Var}(V_{DR}(t; O, G, S) \mid W = w), \\
\sigma^{2-}_{DR}(t; w_0, G, S) &= \lim_{\epsilon \uparrow w_0} \mathrm{Var}(V_{DR}(t; O, G, S) \mid W = w).
\end{aligned}$$

We further define

$$\mathbf{X}_h = \begin{pmatrix} 1 & \frac{W_1 - w_0}{h} \\ 1 & \frac{W_2 - w_0}{h} \\ \vdots & \vdots \\ 1 & \frac{W_n - w_0}{h} \end{pmatrix}$$

$$\mathbf{W}_h^+ = \begin{pmatrix} I(W_1 \geq w_0) K\Big(\frac{W_1 - w_0}{h}\Big) & 0 & 0 & \ldots & 0 \\ 0 & I(W_2 \geq w_0) K\Big(\frac{W_i - w_0}{h}\Big) & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & I(W_n \geq w_0) K\Big(\frac{W_i - w_0}{h}\Big) \end{pmatrix}$$

$$\mathbf{W}_h^- = \begin{pmatrix} I(W_1 < w_0) K\Big(\frac{W_1 - w_0}{h}\Big) & 0 & 0 & \ldots & 0 \\ 0 & I(W_2 < w_0) K\Big(\frac{W_i - w_0}{h}\Big) & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & I(W_n < w_0) K\Big(\frac{W_i - w_0}{h}\Big) \end{pmatrix}.$$

Then $E\{V_{DR}(O; G_0, S^*)\} = P(T > t|W) \equiv \mu(t; W)$. Similar to Cho *et al.* (2021), let h be given and we assume following conditions:

(C1) For $W \neq w_0$, let $\mu(t; w)$ be twice continuously differentiable functions. Let $\mu'(t; w)$ and $\mu''(t; w)$ be the first and second derivatives of $\mu(t; w)$. Let $\mu'^+(t; w)$ and $\mu''^+(t; w)$ be the first and second derivatives of $\mu^+(t; w)$. We define similarly to $\mu'^-(t; w)$ and $\mu''^-(t; w)$.

(C2) There exists $a > 0$ such that $|\mu^+(t; w)|, |\mu'^+(t; w)|, |\mu''^+(t; w)|$ are uniformly bounded on $(w_0, w_0 + a]$. Similarly, $|\mu^-(t; w)|, |\mu'^-(t; w)|, |\mu''^-(t; w)|$ are uniformly bounded on $[w_0 - a, w_0)$.

(C3) $\mu^+(t;w_0), \mu'^+(t;w_0), \mu''^+(t;w_0), \mu^-(t;w_0).\mu'^-(t;w_0)$ and $\mu''^-(t;w_0)$ are finite.

(C4) Let $g(w)$ be the common density of W_i. Assume that $g(w)$ is continuous and bounded away from zero in a neighborhood of w_0.

(C5) $\sigma^2_{DR}(w;G_0,S^*)$ are uniformly bounded in a neighborhood of w_0, and $\sigma^{2+}_{DR}(w_0;G_0,S^*)$, $\sigma^{2-}_{DR}(w_0;G_0,S^*)$ and are finite.

(C6) $\lim_{\substack{W_i\uparrow w_0\\ W_i\downarrow w_0}} E[|V_{DR,i}(t;O_i,G_0,S^*)-\mu(t;W_i)|^r|W_i], r=1,2,3$ are finite. We assume similarly when $W_i \downarrow w_0$.

(C7) K is continuous and symmetric. Moreover, support of K is compact and for any u, $K(u)\geq 0$.

(C8) The bandwidth satisfies $h\sim n^{-1/5}$ where $\sim$ indicates "asymptotically equivalent".

(C9) Let $H_n = o_p(1)$. Then

$$E\left[\left(\frac{W_i-w_0}{h}\right)^{j_1}\left(L^+_{ih}\right)^{j_2}H_n\right]=O(1),\quad j_1=0,\ldots,6,\ \ j_2=1,2,3.$$

(C10) $\hat{G}$ is uniformly consistent to G_0.

(C11) $\hat{S}$ is uniformly consistent to S^* where S^* is possibly incorrect model of S.

Let G_0 be the survival function of censoring from the true model and S^* be the survival function of failure time, which is possibly incorrect. Let

$$\rho^+=\frac{\left(\int_0^\infty u^2K(u)du\right)^2-\left(\int_0^\infty u^3K(u)du\right)\left(\int_0^\infty uK(u)du\right)}{2\left\{\left(\int_0^\infty u^2K(u)du\right)\left(\int_0^\infty K(u)du\right)-\left(\int_0^\infty uK(u)du\right)^2\right\}},$$

$$\rho^-=\frac{\left(\int_{-\infty}^0 u^2K(u)du\right)^2-\left(\int_{-\infty}^0 u^3K(u)du\right)\left(\int_{-\infty}^0 uK(u)du\right)}{2\left\{\left(\int_{-\infty}^0 u^2K(u)du\right)\left(\int_{-\infty}^0 K(u)du\right)-\left(\int_{-\infty}^0 uK(u)du\right)^2\right\}},$$

$$\upsilon^+=\frac{\int_0^\infty\left\{\left(\int_0^\infty s^2K(s)ds\right)-\left(\int_0^\infty sK(s)ds\right)u\right\}^2\{K(u)\}^2du}{g(w_0)\left\{\left(\int_0^\infty u^2K(u)du\right)\left(\int_0^\infty K(u)du\right)-\left(\int_0^\infty uK(u)du\right)^2\right\}^2},$$

$$\upsilon^-=\frac{\int_{-\infty}^0\left\{\left(\int_{-\infty}^0 s^2K(s)ds\right)-\left(\int_{-\infty}^0 sK(s)ds\right)u\right\}^2\{K(u)\}^2du}{g(w_0)\left\{\left(\int_{-\infty}^0 u^2K(u)du\right)\left(\int_{-\infty}^0 K(u)du\right)-\left(\int_{-\infty}^0 uK(u)du\right)^2\right\}^2}.$$

Moreover, since $\tau(t) = \lim_{w \downarrow w_0} P(T > t|W = w) - \lim_{w \uparrow w} P(T > t|W = w_0)$, from our definition, $\tau(t) = \mu^+(t; w_0) - \mu^-(t; w_0)$. Define

$$\begin{aligned}
\varphi(t) &= \rho^+ \mu''^+(t; w_0) - \rho^- \mu''^-(t; w_0), \\
\Sigma_{IPCW_1}(t; G) &= \upsilon^+ \sigma^{2+}_{IPCW_1}(t; w_0, G) + \upsilon^- \sigma^{2-}_{IPCW_1}(t; w_0, G), \\
\Sigma_{IPCW_2}(t; G) &= \upsilon^+ \sigma^{2+}_{IPCW_2}(t; w_0, G) + \upsilon^- \sigma^{2-}_{IPCW_2}(t; w_0, G), \\
\Sigma_{DR}(t; G, S) &= \upsilon^+ \sigma^{2+}_{DR}(t; w_0, G, S) + \upsilon^- \sigma^{2-}_{DR}(t; w_0, G, S).
\end{aligned}$$

We propose the following theorem.

Theorem 1. *Suppose that conditions (C1)–(C11) in the Appendix hold. By extending results from Cho et al. (2021),*

$$\begin{aligned}
\sqrt{nh}\{\hat{\tau}_{IPCW_1}(t) - \tau(t) - \varphi(t)\} &\xrightarrow{d} N(0, \Sigma_{IPCW_1}(t; G_0)), \\
\sqrt{nh}\{\hat{\tau}_{IPCW_2}(t) - \tau(t) - \varphi(t)\} &\xrightarrow{d} N(0, \Sigma_{IPCW_2}(t; G_0)) \\
\sqrt{nh}\{\hat{\tau}_{DR}(t) - \tau(t) - \varphi(t)\} &\xrightarrow{d} N(0, \Sigma_{DR}(t; G_0, S^*)).
\end{aligned} \tag{A.1}$$

In this proof, we only show the result for the DR estimators; Results are similar to the IPCW estimators. Let $W^r_{ih} = ((W_i - w_0)/h)^r$, $r = 0, 1$. Consider

$$\mathbf{A}^+_{DR^*,i,h}(G, S) = \begin{pmatrix} W^0_{ih} V_{DR^*,i}(t; O_i, G, S) L^+_{ih} \\ W^1_{ih} V_{DR^*,i}(t; O_i, G, S) L^+_{ih} \end{pmatrix}, \quad \mathbf{A}^-_{DR^*,i,h}(G, S) = \begin{pmatrix} W^0_{ih} V_{DR^*,i}(t; O_i, G, S) L^-_{ih} \\ W^1_{ih} V_{DR^*,i}(t; O_i, G, S) L^-_{ih} \end{pmatrix}.$$

Then by changing response in Cho *et al.* (2021) with $V_{DR}(t; O, G, S)$, the following lemmas hold.
Lemma 1. Let

$$\begin{aligned}
\mathbf{A}^+_{DR^*,h}(t; W_i, G, S) &= E(\mathbf{A}^+_{DR^*,i,h}(t; G, S) \mid W_i), \\
\mathbf{A}^-_{DR^*,h}(t; W_i, G, S) &= E(\mathbf{A}^-_{DR^*,i,h}(t; G, S) \mid W_i).
\end{aligned}$$

Then

$$\begin{aligned}
\frac{1}{nh} \sum_{i=1}^n \mathbf{A}^+_{DR^*,h}\left(W_i; \hat{G}, \hat{S}\right) &= E\left\{\frac{1}{nh} \sum_{i=1}^n \mathbf{A}^+_{DR^*,i,h}(G_0, S^*)\right\} + o_p\left(h^2\right), \\
\frac{1}{nh} \sum_{i=1}^n \mathbf{A}^-_{DR^*,h}\left(t; W_i, \hat{G}, \hat{S}\right) &= E\left\{\frac{1}{nh} \sum_{i=1}^n \mathbf{A}^-_{DR^*,i,h}(G_0, S^*)\right\} + o_p\left(h^2\right).
\end{aligned}$$

Lemma 2. Let $d_q = \int_0^\infty u^q \{K(u)\}^2 du$, $q = 0, 1, 2$ and

$$\begin{aligned}
\bar{\mathbf{A}}^+_{DR^*,h}\left(t; O, \hat{G}, \hat{S}\right) &= \frac{1}{nh} \sum_{i=1}^n \left\{\mathbf{A}^+_{DR^*,i,h}\left(t; O_i, \hat{G}, \hat{S}\right) - \mathbf{A}^+_{DR^*,h}\left(t; W_i, \hat{G}, \hat{S}\right)\right\}, \\
\bar{\mathbf{A}}^-_{DR^*,h}\left(t; O, \hat{G}, \hat{S}\right) &= \frac{1}{nh} \sum_{i=1}^n \left\{\mathbf{A}^-_{DR^*,i,h}\left(t; O_i, \hat{G}, \hat{S}\right) - \mathbf{A}^-_{DR^*,h}\left(t; W_i, \hat{G}, \hat{S}\right)\right\}.
\end{aligned}$$

Then

$$\mathrm{Var}\left\{\bar{\mathbf{A}}^{+}_{DR^*,h}(t;O,G_0,S)\right\} = \frac{1}{nh}\sigma^{2+}_{DR}(t;w_0;G_0,S)g(w_0)\mathcal{D},$$
$$\mathrm{Var}\left\{\bar{\mathbf{A}}^{-}_{DR^*,h}(t;O,G_0,S)\right\} = \frac{1}{nh}\sigma^{2-}_{DR}(t;w_0,G_0,S)g(w_0)\mathcal{D},$$

where

$$\mathcal{D} = \begin{pmatrix} d_0 + o(1) & d_1 + o(1) \\ d_1 + o(1) & d_2 + o(1) \end{pmatrix}.$$

Lemma 3.

$$\sqrt{nh}\bar{\mathbf{A}}^{+}_{DR^*,h}(t;O,G_0,S) \xrightarrow{d} \{g(w_0)\}N\left(0,\sigma^{2+}_{DR}(t;w_0,G_0,S)\mathcal{D}\right),$$
$$\sqrt{nh}\bar{\mathbf{A}}^{-}_{DR^*,h}(t;O,G_0,S) \xrightarrow{d} \{g(w_0)\}N\left(0,\sigma^{2-}_{DR}(t;w_0,G_0,S)\mathcal{D}\right).$$

Lemma 4.

$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\mathbf{A}^{+}_{DR^*,i,h}\left(t;O_i,\hat{G},\hat{S}\right) - \frac{1}{2}n^{1/2}h^{5/2}g(w_0)\mu''^{+}(t;w_0)\boldsymbol{\delta} \xrightarrow{d} \{g(w_0)\}^{\frac{1}{2}}N\left(0,\sigma^{2+}_{DR}(t;w_0,G_0,S^*)\mathcal{D}\right),$$
$$\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\mathbf{A}^{-}_{DR^*,i,h}\left(t;O_i,\hat{G},\hat{S}\right) - \frac{1}{2}n^{1/2}h^{5/2}g(w_0)\mu''^{-}(t;w_0)\boldsymbol{\delta} \xrightarrow{d} \{g(w_0)\}^{\frac{1}{2}}N\left(0,\sigma^{2-}_{DR}(t;w_0,G_0,S^*)\mathcal{D}\right),$$

where

$$\boldsymbol{\delta} = \begin{pmatrix} \int_0^\infty u^2K(u)du \\ \int_0^\infty u^3K(u)du \end{pmatrix}.$$

Now we prove main theorem. Define $\mathbf{V}_{DR}(t;O,\hat{G},\hat{S}) = \{V_{DR,i}(t;O_i,\hat{G},\hat{S})\}_{i=1}^n$. Then by using matrices, we can express $\mathbf{A}^{+}_{DR^*,i,h}(t;O_i,\hat{G},\hat{S})$ and $\mathbf{A}^{-}_{DR^*,i,h}(t;O_i,\hat{G},\hat{S})$ by

$$\sum_{i=1}^{n}\mathbf{A}^{+}_{DR^*,i,h}\left(t;O_i,\hat{G},\hat{S}\right) = \mathbf{X}_h^T\mathbf{W}_h^{+}\mathbf{V}_{DR}\left(t;O,\hat{G},\hat{S}\right),$$
$$\sum_{i=1}^{n}\mathbf{A}^{-}_{DR^*,i,h}\left(t;O_i,\hat{G},\hat{S}\right) = \mathbf{X}_h^T\mathbf{W}_h^{-}\mathbf{V}_{DR}\left(t;O,\hat{G},\hat{S}\right).$$

Define

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \int_0^\infty K(u)du & \int_0^\infty uK(u)du \\ \int_0^\infty uK(u)du & \int_0^\infty u^2K(u)du \end{pmatrix}.$$

Then

$$\sqrt{nh}\begin{pmatrix} \hat{\alpha}_{R,DR}(t) - \mu^{+}(t;w_0) \\ \hat{\beta}_{R,DR}(t) - \mu'^{+}(t;w_0) \end{pmatrix} = \boldsymbol{\kappa}\left(\left(\mathbf{X}_h^T\mathbf{W}_h^{+}\mathbf{X}_h\right)^{-1}\mathbf{X}_h^T\mathbf{W}_h^{+}\mathbf{V}_{DR}\left(t;O,\hat{G},\hat{S}\right)\right),$$

where $\boldsymbol{\kappa} = \begin{pmatrix} 1 & 0 \\ 0 & h^{-1} \end{pmatrix}$. Then

$$\sqrt{nh}\begin{pmatrix} \hat{\alpha}_{R,DR}(t) - \mu^{+}(t; w_0) \\ \hat{\beta}_{R,DR}(t) - \mu'^{+}(t; w_0) \end{pmatrix} - \frac{1}{2}\boldsymbol{\kappa}^{-1}\boldsymbol{\Upsilon}^{-1}\mu''^{+}(t; w_0)\boldsymbol{\delta} \xrightarrow{d}$$
$$N\Big(0, \sigma^{2+}_{DR}(w_0; G_0, S^*)\, g(w_0)^{-1}\, \boldsymbol{\kappa}^{-1}\boldsymbol{\Upsilon}^{-1}\boldsymbol{\mathcal{V}}\boldsymbol{\Upsilon}^{-1}\boldsymbol{\kappa}^{-1}\Big).$$

Then

$$\sqrt{nh}\Big(\hat{\alpha}_{R,DR}(t) - \mu^{+}(t; w_0) - \rho^{+}\mu''^{+}(t; w_0)\Big) \xrightarrow{d} N\Big(0, \upsilon^{+}\sigma^{2+}_{DR}(t; w_0, G_0, S^*)\Big).$$

By applying similar arguments to $\hat{\alpha}_{L,DR}(t)$,

$$\sqrt{nh}\Big(\hat{\alpha}_{L,DR}(t) - \mu^{-}(t; w_0) - \rho^{-}\mu''^{-}(t; w_0)\Big) \xrightarrow{d} N\Big(0, \upsilon^{-}\sigma^{2-}_{DR}(t; w_0, G_0, S^*)\Big).$$

Then due to independence of $\hat{\alpha}_{R,DR}$ and $\hat{\alpha}_{L,DR}$, it is easy to see that

$$\sqrt{nh}\left(\hat{\tau}_{DR}(t) - \tau(t) - \varphi(t)\right) \xrightarrow{d} N\left(0, \Sigma_{DR}(t; G_0, S^*)\right).$$

Hence we obtain the result. By defining $\Sigma_{IPCW_1}(t; G_0) = \upsilon^{+}\sigma^{2+}_{IPCW_1}(t; w_0, G_0) + \upsilon^{+}\sigma^{2-}_{IPCW_1}(t; w_0, G_0)$ and $\Sigma_{IPCW_2}(t; G_0) = \upsilon^{+}\sigma^{2+}_{IPCW_2}(t; w_0, G_0) + \upsilon^{+}\sigma^{2-}_{IPCW_2}(t; w_0, G_0)$, the proof for $\hat{\tau}_{IPCW_1}$ and $\hat{\tau}_{IPCW_2}$ are similar.

Appendix B: Details for estimation of variance of the proposed estimators

As previous section, let

$$\boldsymbol{X}_h = \begin{pmatrix} 1 & \dfrac{W_1 - w_0}{h} \\ \vdots & \vdots \\ 1 & \dfrac{W_n - w_0}{h} \end{pmatrix},$$

and let $\boldsymbol{W}_{h+}$ and $\boldsymbol{W}_{h-}$ be $n \times n$ diagonal matrices with diagnoal elements being $I(W_i \geq w_0)K((W_i - w_0)/h)$, $i = 1, \ldots, n$ and $I(W_i < w_0)K((W_i - w_0)/h), i = 1, \ldots, n$. Let $\boldsymbol{\Gamma}_{h+} = \boldsymbol{X}_h^T\boldsymbol{W}_{h+}\boldsymbol{X}_h$ and $\boldsymbol{\Gamma}_{h-} = \boldsymbol{X}_h^T\boldsymbol{W}_{h-}\boldsymbol{X}_h$. Moreover, define

$$\sigma^2_{DR,+}(t, w; G_0, S^*) = \mathrm{Var}\left\{\frac{\Delta I\left(T^{(1)} > t\right)}{G_0(T)} + \int_0^{T(t)} \frac{Q_{T^{(1)}}(u, W; S^*)}{G_0(u)} dM_G(u)\middle| W = w\right\},$$
$$\sigma^2_{DR,-}(t, w; G_0, S^*) = \mathrm{Var}\left\{\frac{\Delta I\left(T^{(0)} > t\right)}{G_0(T)} + \int_0^{T(t)} \frac{Q_{T^{(0)}}(u, W; S^*)}{G_0(u)} dM_G(u)\middle| W = w\right\},$$

and $Q_{T^{(k)}}(u, W; S^*) = P_{S^*}(T^{(k)} \geq t | T \geq u, W), k = 0, 1$. Then by applying theory in Cho *et al.* (2021), given h, the asymptotic variance of $\hat{\tau}_{DR}(t)$ is

$$\Sigma^{DR}_{SRD}(G_0, S^*)(t) = \frac{1}{n} e_1^T\left(\boldsymbol{\Gamma}^{-1}_{h+}\boldsymbol{\phi}_{VV+,DR}(t)\boldsymbol{\Gamma}^{-1}_{h+} + \boldsymbol{\Gamma}^{-1}_{h-}\boldsymbol{\phi}_{VV-,DR}(t)\boldsymbol{\Gamma}^{-1}_{h-}\right)e_1,$$

where

$$\boldsymbol{\phi}_{VV+,DR}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i \geq w_0)K\Big(\frac{W_i - w_0}{h}\Big)K\Big(\frac{W_i - w_0}{h}\Big)b_i b_i^T \sigma^2_{DR,+}(t; W_i, G_0, S^*) \tag{B.1}$$

$$\boldsymbol{\phi}_{VV-,DR}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i < w_0)K\Big(\frac{W_i - w_0}{h}\Big)K\Big(\frac{W_i - w_0}{h}\Big)b_i b_i^T \sigma^2_{DR,-}(t; W_i, G_0, S^*), \tag{B.2}$$

with $b_i = (1, (W_i - w_0)/h)^T$. To estimate the variance, it is necessary to estimate $\sigma^2_{DR,+}(t; W_i, G_0, S^*)$ and $\sigma^2_{DR,-}(t; W_i, G_0, S^*)$. By adapting approaches from Calonico *et al*. (2014) and Cho *et al*. (2021), we compute residuals from transposed response and RD estimator. Let $\hat{h}$ be estimated bandwidth by Ludwig and Miller (2007) or Calonico *et al*. (2014).

$$\hat{e}_{i,DR+,\hat{h}}(t) = \hat{V}_{DR,i,\hat{h}}(t) - \hat{\alpha}_{R,DR,\hat{h}}(t),$$
$$\hat{e}_{i,DR-,\hat{h}}(t) = \hat{V}_{DR,i,\hat{h}}(t) - \hat{\alpha}_{L,DR,\hat{h}}(t),$$

where $\hat{V}_{DR,i,\hat{h}}(t)$ and $\hat{\alpha}_{L,DR,\hat{h}}(t)$ are $\hat{V}_{DR,i}(t)$ and $\hat{\alpha}_{L,DR}(t)$ with estimated bandwidth, respectively. As can be seen, $\hat{e}_{i,DR+,\hat{h}}(t)$ and $\hat{e}_{i,DR-,\hat{h}}(t)$ are residuals with respect to $\hat{V}_{DR,i}(t)$. By plugging in these two quantities to (B.1), we obtain estimators of $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$.

$$\hat{\boldsymbol{\phi}}^{pir}_{VV+,DR,\hat{h}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i \geq w_0)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)b_i b_i^T \hat{e}_{i,DR+,\hat{h}}(t)^2,$$

$$\hat{\boldsymbol{\phi}}^{pir}_{VV-,DR,\hat{h}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i < w_0)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)b_i b_i^T \hat{e}_{i,DR-,\hat{h}}(t)^2.$$

We call this as plug-in approach. Another approach is a nonparametric nearest neighbor (NN) based variance estimator by Calonico *et al*. (2014). We compute residuals based on "closeness" to each observation i and use them to compute variance. This approach is more robust to outliers than the plug-in approach. We define "neighbors" in each i for $\hat{V}_{DR,i,\hat{h}}(t)$. Let $\hat{V}_{DR,l_{+,k}(i),\hat{h}}(t)$ be the k^{th} closest unit to unit i among $\{W_i : W_i \geq w_0\}$ and $\hat{V}_{DR,l_{-,k},\hat{h}}(t)$ be the k^{th} closest unit to unit i among $\{W_i : W_i < w_0\}$, respectively. Then we compute residuals from the defined neighbors and estimate $\sigma^2_{VV+,DR}(t; W_i, G_0, S^*)$ and $\sigma^2_{VV-,DR}(t; W_i, G_0, S^*)$ by

$$\hat{\sigma}^2_{VV+,DR,\hat{h}}(t; W_i) = I(W_i \geq w_0)\frac{K}{K+1}\Big(\hat{V}_{DR,i,\hat{h}}(t) - \frac{1}{K}\sum_{k=1}^{K}\hat{V}_{l_{+,k}(i),DR,\hat{h}}(t)\Big)^2,$$

$$\hat{\sigma}^2_{VV-,DR,\hat{h}}(t; W_i) = I(W_i < w_0)\frac{K}{K+1}\Big(\hat{V}_{DR,i,\hat{h}}(t) - \frac{1}{K}\sum_{k=1}^{K}\hat{V}_{l_{-,k}(i),DR,\hat{h}}(t)\Big)^2.$$

From these estimators, the second method to estimate $\boldsymbol{\phi}_{VV+,DR}(t)$ and $\boldsymbol{\phi}_{VV-,DR}(t)$ is

$$\hat{\boldsymbol{\phi}}^{nn}_{VV+,DR,\hat{h}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i \geq w_0)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)b_i b_i^T \hat{\sigma}^2_{VV+,DR,\hat{h}}(t; W_i),$$

$$\hat{\boldsymbol{\phi}}^{nn}_{YY-,DR,\hat{h}}(t) = \frac{1}{n}\sum_{i=1}^{n} I(W_i < w_0)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)K\Big(\frac{W_i - w_0}{\hat{h}}\Big)b_i b_i^T \hat{\sigma}^2_{VV-,DR,\hat{h}}(t; W_i).$$

We call this approach as the nearest neighbor method. From these two methods, we can finally derive the variance of our RD estimator. Let $\mathbf{X}_{\hat{h}}$, $\mathbf{W}_{\hat{h}+}$, and $\mathbf{W}_{\hat{h}-}$ be $\mathbf{X}_h$, $\mathbf{W}_{h+}$, and $\mathbf{W}_{h-}$ with estimated bandwidth. Then, the variance estimator is

$$\frac{1}{n} e_1^T \left(\hat{\mathbf{\Gamma}}_{\hat{h}+}^{-1} \hat{\boldsymbol{\phi}}_{VV+,DR,\hat{h}}(t) \hat{\mathbf{\Gamma}}_{\hat{h}+}^{-1} + \hat{\mathbf{\Gamma}}_{\hat{h}-}^{-1} \hat{\boldsymbol{\phi}}_{VV-,DR,\hat{h}}(t) \hat{\mathbf{\Gamma}}_{\hat{h}-}^{-1} \right) e_1,$$

where

$$\hat{\mathbf{\Gamma}}_{\hat{h}+} = \mathbf{X}_{\hat{h}}^T \mathbf{W}_{\hat{h}+} \mathbf{X}_{\hat{h}}, \quad \hat{\mathbf{\Gamma}}_{\hat{h}-} = \mathbf{X}_{\hat{h}}^T \mathbf{W}_{\hat{h}-} \mathbf{X}_{\hat{h}},$$

and $\hat{\boldsymbol{\phi}}_{VV+,DR,\hat{h}}(t)$ is either $\hat{\boldsymbol{\phi}}^{pir}_{VV+,DR,\hat{h}}(t)$ or $\hat{\boldsymbol{\phi}}^{nn}_{VV+,DR,\hat{h}}(t)$. Definition of $\hat{\boldsymbol{\phi}}_{VV-,DR,\hat{h}}(t)$ is similar. By using the same process, we can also estimate variance of IPCW_1 and IPCW_2 estimators.

References

Andersen PK, Klein JP, and Rosthøj S (2003). Generalised linear models for correlated pseudo-observations, with applications to multi-state models, *Biometrika*, **90**, 15–27.

Andriole GL, Crawford ED, Grubb III RL *et al.* (2009). Mortality results from a randomized prostate-cancer screening trial, *New England Journal of Medicine*, **360**, 1310–1319.

Bor J, Moscoe E, Mutevedzi P, Newell ML, and Bärnighausen T (2014). Regression discontinuity designs in epidemiology: Causal inference without randomized trials, *Epidemiology*, **25**, 729–737.

Brier GW (1950). Verification of forecasts expressed in terms of probability, *Monthly Weather Review*, **78**, 1–3.

Calonico S, Cattaneo MD, and Farrell MH (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs, *The Econometrics Journal*, **23**, 192–210.

Calonico S, Cattaneo MD, and Titiunik R (2014). Robust nonparametric confidence intervals for regression-discontinuity designs, *Econometrica*, **82**, 2295–2326.

Calonico S, Cattaneo MD, and Titiunik R (2015). `rdrobust`: An R package for robust nonparametric inference in regression-discontinuity designs, *R Journal*, **7**, 38–51.

Calonico S, Cattaneo MD, Farrell MH, and Titiunik R (2019). Regression discontinuity designs using covariates, *Review of Economics and Statistics*, **101**, 442–451.

Cho Y, Hu C, and Ghosh D (2021). Analysis of regression discontinuity designs using censored data, *Journal of Statistical Research*, **55**, 225–248.

Cho Y, Molinaro AM, Hu C, and Strawderman RL (2020). Regression trees for cumulative incidence functions, Available from: *arXiv preprint* arXiv:2011.06706

Cho Y, Molinaro AM, Hu C, and Strawderman RL (2022). Regression trees and ensembles for cumulative incidence functions, *The International Journal of Biostatistics*, **18**, 397–419.

Cox DR (1972). Regression models and life-tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, **34**, 187–202.

Fan J and Gijbels I (1994). Censored regression: Local linear approximations and their applications, *Journal of the American Statistical Association*, **89**, 560–570.

Fan J and Gijbels I (1996). *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability 66*, CRC Press, London.

Gerds TA and Schumacher M (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times, *Biometrical Journal*, **48**, 1029–1040.

Graf E, Schmoor C, Sauerbrei W, and Schumacher M (1999). Assessment and comparison of prognostic classification schemes for survival data, *Statistics in Medicine*, **18**, 2529–2545.

Hahn J, Todd P, and Van der Klaauw W (1999). Evaluating the effect of an antidiscrimination law using a regression-discontinuity design (Technical report), National Bureau of Economic Research, Retrieved Jan. 17, 2024, Available from: https://www.nber.org/system/files/working_papers/w7131/w7131.pdf

Hahn J, Todd P, and Van der Klaauw W (2001). Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica*, **69**, 201–209.

Imbens G and Kalyanaraman K (2012). Optimal bandwidth choice for the regression discontinuity estimator, *The Review of Economic Studies*, **79**, 933–959.

Zajonc T (2012). Essays on Causal Inference for Public Policy (Doctoral dissertation), Harvard University, Cambridge, MA.

Imbens G and Lemieux T (2008). Regression discontinuity designs: A guide to practice, *Journal of*

Econometrics, **142**, 615–635.

Lee DS (2008). Randomized experiments from non-random selection in us house elections, *Journal of Econometrics*, **142**, 675–697.

Lin DY and Ying Z (1994). Semiparametric analysis of the additive risk model, *Biometrika*, **81**, 61–71.

Lostritto K, Strawderman RL, and Molinaro AM (2012). A partitioning deletion/substitution/addition algorithm for creating survival risk groups, *Biometrics*, **68**, 1146–1156.

Ludwig J and Miller DL (2007). Does head start improve children's life chances? evidence from a regression discontinuity design, *Quarterly Jounral of Economics*, **122**, 159–208.

Moscoe E, Bor J, and Bärnighausen T (2015). Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice, *Journal of Clinical Epidemiology*, **68**, 132–143.

Prorok PC, Andriole GL, Bresalier RS *et al.* (2000). Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial, *Controlled Clinical Trials*, **21**, 273S–309S.

Robins JM, Rotnitzky A, and Zhao LP (1994). Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, **89**, 846–866.

Rubin DB (1974). Estimating causal effects of treatments in randomized and nonrandomized studies, *Journal of Educational Psychology*, **66**, 688–701.

Rubin D and van der Laan MJ (2007). A doubly robust censoring unbiased transformation, *The International Journal of Biostatistics*, **3**, 1–19.

Shoag J, Halpern J, Eisner B, Lee R, Mittal S, Barbieri CE, and Shoag D (2015). Efficacy of prostate-specific antigen screening: Use of regression discontinuity in the plco cancer screening trial, *JAMA Oncology*, **1**, 984–986.

Silverman BW (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. CRC press, London.

Steingrimsson JA, Diao L, and Strawderman RL (2019). Censoring unbiased regression trees and ensembles, *Journal of the American Statistical Association*, **114**, 370–383.

Strawderman RL (2000). Estimating the mean of an increasing stochastic process at a censored stopping time, *Journal of the American Statistical Association*, **95**, 1192–1208.

Thistlethwaite DL and Campbell DT (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment, *Journal of Educational Psychology*, **51**, 309–317.

Tsiatis A (2007). *Semiparametric Theory and Missing Data*, Springer Science & Business Media, New York.

Received July 4, 2023; Revised November 8, 2023; Accepted December 8, 2023