

ANALYSIS OF REGRESSION DISCONTINUITY DESIGNS USING CENSORED DATA

YOUNGJOO CHO

Department of Applied Statistics, Konkuk University, Seoul, Republic of Korea

Email: yvc5154@konkuk.ac.kr

CHEN HU

*Division of Biostatistics and Bioinformatics, Sidney Kimmel Comprehensive Cancer Center
Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA*

Email: chu22@jhmi.edu

DEBASHIS GHOSH*

*Department of Biostatistics and Informatics
Colorado School of Public Health, Aurora, CO 80045, USA*

Email: debashis.ghosh@cuanschutz.edu

SUMMARY

In many medical and scientific settings, the choice of treatment or intervention may be determined by a covariate threshold. For example, elderly men may receive more thorough diagnosis if their prostate-specific antigen (PSA) level is high. In these cases, the causal treatment effect is often of great interest, especially when there is a lack of evidence from randomized clinical trials. From the social science literature, a class of methods known as regression discontinuity (RD) designs can be used to estimate the treatment effect in this situation. Under certain assumptions, such an estimand enjoys a causal interpretation. We show how to estimate causal effects under the regression discontinuity design for censored data. The proposed estimation procedure employs a class of censoring unbiased transformations that includes inverse probability censored weighting and doubly robust transformation schemes. Simulation studies are used to evaluate the finite-sample properties of the proposed estimator. We also illustrate the proposed method by evaluating the causal effect of PSA-dependent screening strategies.

Keywords and phrases: Causal effect; Double robustness; Instrumental variable; Observational studies; Survival analysis.

* Corresponding author

© Institute of Statistical Research and Training (ISRT), University of Dhaka, Dhaka 1000, Bangladesh.

1 Introduction

In observational studies, scientific interest can typically be formulated in terms of causal estimands. The presence of confounding variables makes their estimation difficult. To perform causal inference, the analyst typically relies on several assumptions. One important assumption is the “no unmeasured confounders” assumption, which implies that treatment assignment is independent of potential outcome given confounders (Rosenbaum and Rubin, 1983). This has also been referred to as the unconfoundedness assumption. However, this assumption is typically not empirically testable.

Regression discontinuity (RD) designs have been widely used in the social sciences. One appealing feature of the RD design is that the treatment assignment is either deterministically or probabilistically determined by a continuous variable of interest, termed the forcing variable. For such a design, the no unmeasured confounders assumption is not required for inferring causality. In the neighborhood of threshold, we have a so-called “randomization” environment so that it is possible to deduce causality. The idea is that with lack of manipulation of the threshold, the observed and unobserved confounders have same distribution in the neighborhood of threshold (Lee, 2008; Lee and Lemieux, 2010). The study of RD designs was initiated by Thistlethwaite and Campbell (1960) and has been developed further in many subsequent studies. For example, Hahn et al. (1999, 2001) proved theoretical results on the identification of RD estimates and asymptotic properties. Ludwig and Miller (2007) proposed bandwidth selection procedures for local linear regression and applied it to evaluate the effects of funding on educational programs. To obtain optimal bandwidth, Imbens and Kalyanaraman (2012) proposed a selection procedure to provide theoretical basis to select bandwidth. To address the bias caused by the local linear regression estimates in Hahn et al. (1999, 2001); Calonico et al. (2014) developed bias-corrected nonparametric estimation approaches whose confidence intervals demonstrate improved coverage relative to those from other RD estimators.

Much of the aforementioned studies dealt with the case of uncensored data. The outcome of interest often represents the disease risks and is measured as a failure time, which is naturally subject to right censoring. For example, it remains controversial whether prostate-specific antigen (PSA)-based screening strategies can meaningfully reduce prostate-cancer specific incidence, which are both measured as time-to-event data and subject to censoring. In practice, a PSA level ≥ 4.0 mg/nl is often used as a “magic number” to identify those with a “high risk” of prostate cancer, and often is often accompanied by a series of additional diagnosis tests to detect if prostate cancer is present. As introduced earlier, in the absence of randomized clinical trials, RD design provides an ideal opportunity to answer this question, for example, whether the additional tests prompted by a PSA level ≥ 4.0 mg/nl can meaningfully improve clinical outcomes. Shoag et al. (2015) attempted to answer this interesting question using the standard RD design method for binary outcomes, without fully utilizing the time-to-event information and accounting for right censoring.

There has been limited research in the area of RD designs for censored data. Lesik (2007) used a proportional hazard model for discrete-time data to investigate the causal effect of developmental mathematics programs on student retention by examining student retention as time-to-event data. Recently, Bor et al. (2014); Moscoe et al. (2015) discussed the use of RD designs in medical and epidemiological studies, and applied their proposed method to estimate the effect of early versus late treatment initiation for HIV patients on survival. They proposed using hazard-based models to study

risk of HIV patients on the early versus late treatment initiation from CD4 counts. In this case, the CD4 count creates discontinuity, and they find that there discontinuity exists on CD4 counts greater or equal to 200 cells/ μL . They estimate the mortality rate from the hazard-based model. However, the aforementioned methods have not been rigorously studied, either in theory or through numerical simulations.

A key question in prostate cancer data is “Is there meaningful difference of time to any first cancer incidence or prostate cancer between PSA level ≥ 4.0 mg/nl and < 4.0 mg/nl?” If there is a difference, then 4.0 mg/nl can be used for decision-making purposes. However, previous approaches in the RD design are not appropriate to answer this question because they are either only applicable to uncensored data or are difficult to examine in terms of differences of time-to-event, which is directly connected to survival.

Directly applying standard nonparametric RD estimation procedures without accounting for censoring is not appropriate. One way to solve this issue is to use existing RD estimation procedures with a transformation of the response that behaves in a manner analogous to the uncensored data case. Fan and Gijbels (1994) propose using local linear regression based on a transformed response for censored data. Their proposed transformation includes the inverse probability weighted censoring (IPCW) method, which is a commonly used method in the missing data literature to handle censoring. However, this method is inefficient in that it does not include information of censored observations in the estimation. Rubin and Van der Laan (2007) overcame this difficulty by proposing a doubly robust transformation of the response, which requires modeling of failure time distribution as well as censoring distribution. This approach shows promise compared to IPCW methods in prognostic modeling (Steingrímsson et al., 2016, 2019), but no studies have shown its efficiency gains for estimation of parameters in regression modeling with the purpose of inference.

We propose a class of estimation procedures in the RD design for censored data. We prove mathematical properties of our estimators and examine its performance using numerical studies. Moreover, we directly model the survival time, so our interpretation regarding survival is direct. We first find a relevant quantity based on observed data for the outcome and then apply a local linear regression approach. In Section 2, we review the relevant data structures and discuss approaches of RD design for uncensored data. Section 3 and 4 describe the extension of RD designs to censored data as well as laying out the methodology with attendant asymptotic results. In Section 5, simulation studies are presented to evaluate the finite-sample properties of our proposal. In Section 6, we apply our method to the Prostate, Lung, Colorectal, and Ovarian (PLCO) dataset to test the effect of treatment assignment by PSA. Some discussion concludes Section 7.

2 Review of RD Design for Uncensored Data

Before discussing the proposed methodology, we first introduce the RD design for uncensored data using a potential outcome framework. Let $(Y_*^{(1)}, Y_*^{(0)})$ be the potential outcomes under treatment and control; we use Z to define treatment. We define \mathbf{W} to be a vector of forcing variables; for simplicity, we consider only one forcing variable W . As the name suggests, the forcing variable determines the treatment (Imbens and Lemieux, 2008). Since only one of the potential outcomes is

observable, the observed response is $Y_* = ZY_*^{(1)} + (1 - Z)Y_*^{(0)}$. The main characteristic of the RD design is that the treatment assignment Z depends on a function of W , which can be deterministic or probabilistic. This corresponds to the sharp and fuzzy RD designs, respectively.

One key assumption in RD designs is that there is no gaming of the forcing variable W (McCrary, 2008). This is checked in practice by empirically plotting the distribution of W and checking to see that there is no clumping around the cutoff value of interest. When the assumption of no gaming on the forcing variable holds, we have a “locally randomized” study (Lee and Lemieux, 2010). The logic of this result is as follows. The distribution of the forcing variable given observed and unobserved confounders is continuous. Then by Bayes’ rule, the joint distribution of observed and unobserved confounders given the forcing variable is continuous at the cutoff point, which implies that the entire confounder distribution is identical at the neighborhood of the cut off (Lee and Lemieux, 2010). This “local randomization” is a compelling feature compared to standard observational studies and allows for establishing causality as in randomized experiments (Bor et al., 2014).

In the sharp RD design, treatment assignment is decided by a deterministic function of forcing variable. Let H^* be a known discontinuous function; for the sharp RD, $Z = H^*(W)$. The main causal effect of interest is the average treatment effect at the discontinuity point w_0 . By design, if the value of the forcing variable is greater than or equal to the cutoff, $E\{Y_*^{(1)}|W\} = E(Y_*|W)$. Similarly, $E\{Y_*^{(0)}|W\} = E(Y_*|W)$ if the value of the forcing variable is less than the cut point. Since our interest focuses on the causal effect at w_0 , with a continuity assumption for $E\{Y_*^{(1)}|W\}$ and $E\{Y_*^{(0)}|W\}$, we can identify limits around the threshold (Bor et al., 2014).

$$E[Y_*^{(1)} - Y_*^{(0)}|W = w_0] = \lim_{w \downarrow w_0} E(Y_*|W = w) - \lim_{w \uparrow w_0} E(Y_*|W = w). \quad (2.1)$$

In the fuzzy RD design, treatment assignment is a probabilistic function of the forcing variable. A jump in the probability of treatment assignment exists at the threshold (Imbens and Lemieux, 2008) but it is less than one and depends on the forcing variable. Let q_1 and q_2 be monotone functions of the forcing variable. For fuzzy RD designs, the probability of receiving treatment is

$$P(Z = 1|W) = \begin{cases} q_1(W) & \text{if } W < w_0, \\ q_2(W) & \text{if } W \geq w_0. \end{cases}$$

By contrast, for sharp RD designs, the probability of receiving treatment given that the forcing variable is greater than a certain cutoff is one. Hence, the treatment status and treatment assignment are the same. In fuzzy RD designs,

$$\lim_{w \downarrow w_0} P(Z = 1|W = w) \neq \lim_{w \uparrow w_0} P(Z = 1|W = w)$$

with the difference of $\lim_{w \downarrow w_0} P(Z = 1|W = w)$ and $\lim_{w \uparrow w_0} P(Z = 1|W = w)$ not being equal to one. Hence, for fuzzy RD, the treatment assignment is not equivalent to the treatment status. One can consider the sharp RD as a special case of the fuzzy RD when $\lim_{w \downarrow w_0} P(Z = 1|W = w) - \lim_{w \uparrow w_0} P(Z = 1|W = w) = 1$. Now expression (2.1) is not the average treatment

effect in a fuzzy RD design. However, since the forcing variable determines treatment assignment, it effectively functions as an instrumental variable. By using the arguments in Angrist et al. (1996), we can obtain the so-called complier average treatment effect. This effect is the main identifiable causal estimand in the fuzzy RD design and is important in practice because participants in the study may not comply with the initial treatment assignment. Formally, we have

$$\begin{aligned} & \frac{\lim_{w \downarrow w_0} E(Y_* | W = w) - \lim_{w \uparrow w_0} E(Y_* | W = w)}{\lim_{w \downarrow w_0} P(Z | W = w) - \lim_{w \uparrow w_0} P(Z | W = w)} \\ &= E[Y_*^{(1)} - Y_*^{(0)} | W = w_0, \text{subject is a complier}]. \end{aligned}$$

In this case, the unconfoundedness assumption is not realistic because people with similar values of the forcing variable away from the cut-point may receive different treatment. These two subjects will thus not be comparable (Imbens and Lemieux, 2008).

3 Extension of RD Designs to Censored Data: Data Structure and Assumptions

As in the uncensored data case, we define $(T^{(1)}, T^{(0)})$ as potential outcomes under treatment and control assignments, respectively. In survival data, $(T^{(1)}, T^{(0)})$ are potential times to event for the treatment and control. Without censoring, time to failure T is only observable for either of treatment and control group, that is, $T = ZT^{(1)} + (1 - Z)T^{(0)}$. Let $X = (T, W, Z)$ be full data. Let C be time to censoring. We can only observe $\tilde{T} = T \wedge C$, $\Delta = I(T \leq C)$. Define $O = (\tilde{T}, \Delta, Z, W)$. The observable data are $O_i = (\tilde{T}_i, W_i, Z_i, \Delta_i)$, $i = 1, \dots, n$. We assume that the observed data are independent and identically distributed. We apply a logarithm transform to the response. Define

$$Y^{(1)} = \log T^{(1)}, \quad Y^{(0)} = \log T^{(0)}, \quad Y = \log T, \quad \tilde{Y} = Y \wedge \log C,$$

and Y_i and \tilde{Y}_i are individual realizations of Y and \tilde{Y} , respectively. Usually, the censoring time is not affected by treatment assignment so that it is reasonable to assume that $C^{(1)} = C^{(0)} = C$ (Bai et al., 2013), where $(C^{(1)}, C^{(0)})$ are the potential outcomes for censoring under treatment and control. Let w_0 be a cut-point for the forcing variable. We now discuss the necessary assumptions for identification of causal effects in RD designs with censored data.

1. Participants in the study do not have any ability to manipulate the cutoff.
2. $E(Y^{(1)} | W = w)$ and $E(Y^{(0)} | W = w)$ are continuous in w .
3. $(Y^{(0)}, Y^{(1)})$ and C are independent.
4. Censoring is random, i.e., $C \perp (W, Z)$.
5. Defining $S^{(1)}(t|w) = P(T^{(1)} > t | W = w)$ and $S^{(0)}(t|w) = P(T^{(0)} > t | W = w)$, $S^{(1)}(t|w)$ and $S^{(0)}(t|w)$ are continuous at w for all t . Denote $S(t|w) = P(T > t | W = w)$.
6. Let $G(t) = P(C > t)$. $G(t)$ is continuous for all t .

7. For the fuzzy RD, $P(Z = 1|W = w)$ is continuous at w except for $w = w_0$ and

$$\lim_{w \downarrow w_0} P(Z = 1|W = w) \neq \lim_{w \uparrow w_0} P(Z = 1|W = w).$$

8. For the fuzzy RD, $Z(w^*)$ is nondecreasing in w^* at $w^* = w_0$, where $Z(w)$ is potential treatment status given point w .

Condition 1 ensures an effectively “randomized environment” in the cutoff. Condition 2 is a smoothness assumption for the mean potential outcome functions around the cutoff point w_0 . Condition 3 is the typical noninformative censoring assumption in survival analysis. Condition 4 states that C is independent of W and Z . Conditions 5 and 6 guarantee smoothness for the failure time and censoring distributions. Condition 7 is a standard assumption for fuzzy RD. Condition 8 states that potential treatment status is monotonic in the cutoff point (Imbens and Lemieux, 2008). For survival data, in the sharp RD design, the average causal effect of treatment given the forcing variable is $E(Y^{(1)} - Y^{(0)}|W = w)$. Under these assumptions,

$$\begin{aligned} E\{Y^{(0)}|W = w_0\} &= \lim_{w \uparrow w_0} E\{Y^{(0)}|Z = 0, W = w_0\} = \lim_{w \uparrow w_0} E(Y|W = w), \\ E\{Y^{(1)}|W = w_0\} &= \lim_{w \downarrow w_0} E\{Y^{(1)}|Z = 1, W = w_0\} = \lim_{w \downarrow w_0} E(Y|W = w). \end{aligned}$$

Under a sharp RD, the average treatment effect is then

$$\tau_{SRD}(w_0) = \lim_{w \downarrow w_0} E[Y|W = w] - \lim_{w \uparrow w_0} E[Y|W = w].$$

For a fuzzy RD, the average treatment effect is

$$\tau_{FRD}(w_0) = \frac{\lim_{w \downarrow w_0} E(Y|W = w) - \lim_{w \uparrow w_0} E(Y|W = w)}{\lim_{w \downarrow w_0} P(Z = 1|W = w) - \lim_{w \uparrow w_0} P(Z = 1|W = w)}.$$

We denote $\tau_{SRD}(w_0) \equiv \tau_{SRD}$ and $\tau_{FRD}(w_0) \equiv \tau_{FRD}$.

4 Proposed Methodology

4.1 Censoring unbiased transformations

In sharp RD with uncensored data, we can use nonparametric estimation directly with the response variable. However, for censored data, it is very difficult to use response directly due to censoring. In our case, the issue is to find a transformation q^* based on observed data such that $E\{q^*(O)\} = E(Y|W)$. The transformation is referred to as a censoring unbiased transformation (Fan and Gijbels, 1994; Rubin and van der Laan, 2007; Steingrimsson et al., 2019). One approach is to use an inverse probability censoring weighted (IPCW) method to obtain $E(Y|W)$, which is

$$Y_{IPCW} = \frac{\Delta Y}{G(T)}.$$

It is easy to show that Y_{IPCW} is censoring unbiased transformation. However, this approach requires the censoring distribution to be correctly specified and yields an inefficient estimator. Rubin and van der Laan (2007) propose a doubly robust (DR) censoring unbiased transformation for the failure time. Let

$$M_G(u) = I(\tilde{T} \leq u, \Delta = 0) - \int_0^u I(\tilde{T} \geq s)\lambda_G(s)ds,$$

where $\lambda_G(s)$ is true hazard function of G . The form is

$$Y_{DR} = \frac{\Delta Y}{G(T)} + \int_0^{\tilde{T}} \frac{Q_Y(u|W)}{G(u)} dM_G(u),$$

where $Q_Y(u, W)$ is $E\{Y|T > u, W\}$. This DR transformation requires estimation of both censoring and failure time distributions. It is a combination of an IPCW term and a mean zero martingale transform term. An IPCW term is a special case of DR transformation. This martingale transformation term utilizes information from censored observations, which yields greater efficiency than using the IPCW approach. Since it is also a censoring unbiased transformation, it guarantees that $E(Y_{DR}(T)|W = w) = E(Y|W = w)$ for any w (Steingrímsson et al., 2019). Moreover, if either the censoring distribution or failure time distribution is correctly specified, then the estimator from DR transformation based on the estimated Q and G is consistent for $E(Y|W)$. Moreover, if models for Q and G are both correctly specified, then the resulting estimator from the DR transformation is the most efficient given class of the estimator of $E(Y|W)$ (Steingrímsson et al., 2019).

4.2 Asymptotic theory

Using the transformation from Section 4.1 enables us to use uncensored data techniques in the censored situation by applying the data-dependent transformation to the observed outcome. As can be seen in the previous section, these transformations depend on either the censoring distribution or both the censoring and failure time distributions. In this section, we discuss RD-based estimation procedures of the treatment effect and their asymptotic properties with censored data. The IPCW and DR transformations using observed data are given by

$$Y_{IPCW,i}(O_i; G) = \frac{\Delta_i Y_i}{G(T_i)}, \quad Y_{DR,i}(O_i; G, S) = \frac{\Delta_i Y_i}{G(T_i)} + \int_0^{\tilde{T}_i} \frac{Q_Y(u|W_i; S)}{G(u)} dM_{G,i}(u),$$

where S represents a model for failure time and $Q_Y(\cdot; S) = E(Y|T \geq \cdot; S)$ and

$$M_{G,i}(u) = I(\tilde{T}_i \leq u, \Delta_i = 0) - \int_0^u I(\tilde{T}_i \geq s)\lambda_G(s)ds.$$

We only focus on the DR transformation. To obtain limits in the sharp and fuzzy RD designs, parametric methods are not very attractive because modeling and the discontinuity point depend on the particular parametric distribution. Hahn et al. (2001) demonstrate that a causal effect in RD designs is nonparametrically estimable. To handle the discontinuity at the cutoff, local linear

regression method is widely used in RD literature (Hahn et al., 1999; Imbens and Lemieux, 2008). We now apply Fan and Gijbels (1996) local linear regression method to the transformed response and Z . Applying kernel smoothing for binary variable is also advocated by many authors (Hahn et al., 1999; Imbens and Lemieux, 2008; Li and Racine, 2003; Okumura, 2011). Let $K(\cdot)$ be a kernel function and h be bandwidth. For fuzzy RD, we consider the following loss functions

$$\begin{aligned}
 & U_{R,DR}^{FRD,Y}(\alpha_R^Y, \beta_R^Y; G, S) \\
 &= \sum_{i=1}^n I(W_i \geq w_0) \{Y_{DR,i}(O_i; G, S) - \alpha_R^{(Y)} - \beta_R^{(Y)}(W_i - w_0)\}^2 K\left(\frac{W_i - w_0}{h}\right), \\
 & U_{L,DR}^{FRD,Y}(\alpha_L^Y, \beta_L^Y; G, S) \\
 &= \sum_{i=1}^n I(W_i < w_0) \{Y_{DR,i}(O_i; G, S) - \alpha_L^{(Y)} - \beta_L^{(Y)}(W_i - w_0)\}^2 K\left(\frac{W_i - w_0}{h}\right), \\
 & U_R^{FRD,Z}(\alpha_R^Z, \beta_R^Z) \\
 &= \sum_{i=1}^n I(W_i \geq w_0) \{Z_i - \alpha_R^Z - \beta_R^Z(W_i - w_0)\}^2 K\left(\frac{W_i - w_0}{h}\right), \quad \text{and} \\
 & U_L^{FRD,Z}(\alpha_L^Z, \beta_L^Z) \\
 &= \sum_{i=1}^n I(W_i < w_0) \{Z_i - \alpha_L^Z - \beta_L^Z(W_i - w_0)\}^2 K\left(\frac{W_i - w_0}{h}\right). \tag{4.1}
 \end{aligned}$$

We can similarly define loss functions for the IPCW transformation; denote them by

$$U_{R,IPCW}^{FRD,Y}(\alpha_R^Y, \beta_R^Y; G), \quad \text{and} \quad U_{L,IPCW}^{FRD,Y}(\alpha_L^Y, \beta_L^Y; G).$$

In sharp RD designs, since the treatment assignment is deterministic, the loss functions with transformation responses are only necessary and they are identical to those in fuzzy RD design. To estimate parameters, we first estimate G and S . Next, we estimate loss functions in (4.1) and we apply standard least squares with respect to estimated loss functions. After we estimate G and S , our method is simple.

Let $\{\hat{\alpha}_{R,DR}^{FRD,Y}(G, S), \hat{\beta}_{R,DR}^{FRD,Y}(G, S), \hat{\alpha}_{L,DR}^{FRD,Y}(G, S), \hat{\beta}_{L,DR}^{FRD,Y}(G, S)\}$ be the estimators using the DR transformation in fuzzy RD design, respectively. Furthermore, we define $\{\hat{\alpha}_R^{FRD,Z}, \hat{\beta}_R^{FRD,Z}, \hat{\alpha}_L^{FRD,Z}, \hat{\beta}_L^{FRD,Z}\}$ for the estimators of modeling treatment assignment. We can similarly define estimators for the sharp RD designs; denote them by $\{\hat{\alpha}_{R,DR}^{SRD,Y}(G, S), \hat{\beta}_{R,DR}^{SRD,Y}(G, S), \hat{\alpha}_{L,DR}^{SRD,Y}(G, S), \hat{\beta}_{L,DR}^{SRD,Y}(G, S)\}$. Then we can derive estimators for the fuzzy and sharp RD designs:

$$\begin{aligned}
 \hat{\tau}_{FRD}^{DR}(G, S) &= \left(\hat{\alpha}_{R,DR}^{FRD,Y}(G) - \hat{\alpha}_{L,DR}^{FRD,Y}(G)\right) / \left(\hat{\alpha}_R^{FRD,Z} - \hat{\alpha}_L^{FRD,Z}\right), \quad \text{and} \\
 \hat{\tau}_{SRD}^{DR}(G, S) &= \left(\hat{\alpha}_{R,DR}^{SRD,Y}(G, S) - \hat{\alpha}_{L,DR}^{SRD,Y}(G, S)\right).
 \end{aligned}$$

Note that we have suppressed dependence on the bandwidth in the definition of these estimators. We discuss how to estimate bandwidth formally in Section 4.3. In the next set of results, we show

that under certain conditions, we can prove asymptotic convergence results for the sharp and fuzzy RD estimators. As can be seen, our estimators depend on G and S . Let G_0 and S_0 be the true distributions of failure and censoring times, respectively. Let \hat{G} and \hat{S} be estimated distributions of failure and censoring times, respectively. In the estimation, we correctly estimate censoring distribution while we may incorrectly estimate survival distribution. As discussed in the Supplementary Materials, we assume uniform consistency of \hat{G} to G_0 and \hat{S} to S^* , where S^* is possibly an incorrect model of S . We discuss the estimation of these two distributions in the next subsection. For these two theoretical results, the IPCW and DR estimators are asymptotically normal with some bias. Regularity conditions needed for Theorem 1 are the following:

$$(C1) \quad I_1 = \int_0^\infty \log(u) \frac{G_0(u)}{G(u-)} dF_0(u|w) < \infty.$$

(C2) For $a > 0$,

$$D_1(a) = \int_0^a \frac{S_0(u|w)}{S(u|w)} \frac{d\bar{G}_0(u)}{G(u-)} < \infty, \quad D_2(a) = \int_0^a \frac{G_0(u)S_0(u|w)}{G(u)S(u|w)} \frac{d\bar{G}(u)}{G(u-)} < \infty.$$

$$(C3) \quad I_2 = \int_0^\infty \log(u)[D_1(a-) - D_2(a-)]dF(u|w) < \infty.$$

$$(C4) \quad \int_0^\infty \frac{[\log(u)]^2}{G_0(u)} dF_0(u|w) < \infty.$$

$$(C5) \quad D_3(a) = \int_0^a \frac{Q_Y(u, w, S)}{\{G_0(u)\}^2} d\bar{G}_0(u) < \infty \text{ for each } a > 0.$$

(C6) \hat{G} is uniformly consistent to G_0 .

(C7) \hat{S} is uniformly consistent to S^* where S^* is possibly incorrect model of S .

As Steingrímsson, Diao, and Strawderman (2019) have shown, from conditions (C1)-(C5), we can prove that

$$E(Y_{DR}(O; G_0, S)|W) = E(Y_{DR}(O; G, S_0)|W) = E(Y_{DR}(O; G_0, S_0)|W) = E(Y|W) = \mu(W).$$

This is necessary for proving asymptotic normality of $\hat{\tau}_{FRD}^{IPCW}(\hat{G})$, $\hat{\tau}_{FRD}^{DR}(\hat{G}, \hat{S})$, $\hat{\tau}_{SRD}^{IPCW}(\hat{G})$, and $\hat{\tau}_{SRD}^{DR}(\hat{G}, \hat{S})$. Let $p(w) = E(Z|W = w)$ and define

$$\begin{aligned} \mu^+(w) &= \lim_{w \downarrow w_0} E(Y|W = w), & \mu^-(w) &= \lim_{w \uparrow w_0} E(Y|W = w), \\ p^+(w) &= \lim_{w \downarrow w_0} P(Z = 1|W = w), & p^-(w) &= \lim_{w \uparrow w_0} P(Z = 1|W = w). \end{aligned}$$

Now we need conditions similar to those in Hahn et al. (1999). Define

$$\begin{aligned} Y_{IPCW}(O; G) &= \frac{\Delta Y}{G(T)}, \quad Y^{DR}(O; G, S) = \frac{\Delta Y}{G(T)} + \int_0^{\bar{T}} \frac{Q_Y(u, W, S)}{G(u)} dM_G(u), \\ Y_{IPCW^*}(O; G) &= Y_{IPCW}(O; G) - \mu^+(w_0) - \mu'^+(w_0)(W - w_0), \\ Y_{DR^*}(O; G, S) &= Y^{DR}(O; G, S) - \mu^+(w_0) - \mu'^+(w_0)(W - w_0), \\ Z^* &= Z - p^+(w_0) - p'^+(w_0)(W - w_0), \\ L_{ih}^+ &= I(W_i \geq w_0)K\left(\frac{W_i - w_0}{h}\right), \quad L_{ih}^- = I(W_i < w_0)K\left(\frac{W_i - w_0}{h}\right). \end{aligned}$$

We further define

$$\begin{aligned} \sigma_{DR}^2(w; G, S) &= Var(Y_{DR}(O; G, S)|W = w), \\ \sigma_{DR}^{2+}(w_0; G, S) &= \lim_{\epsilon \downarrow w_0} Var(Y_{DR}(O; G, S)|W = w), \\ \sigma_{DR}^{2-}(w_0; G, S) &= \lim_{\epsilon \uparrow w_0} Var(Y_{DR}(O; G, S)|W = w), \\ \eta_{DR}(w; G, S) &= Cov(Y_{DR}(O; G, S), Z|W = w), \\ \eta_{DR}^+(w_0; G, S) &= \lim_{w \downarrow w_0} Cov(Y_{DR}(O; G, S), Z|W = w), \\ \eta_{DR}^-(w_0; G, S) &= \lim_{w \uparrow w_0} Cov(Y_{DR}(O; G, S), Z|W = w). \end{aligned}$$

We can make similar definitions for $\sigma_{IPCW}^2(G, S)$, $\sigma_{IPCW}^{2+}(w_0; G, S)$, $\sigma_{IPCW}^{2-}(w_0; G, S)$, $\eta_{IPCW}(w_0; G, S)$, $\eta_{IPCW}^+(w_0; G, S)$ and $\eta_{IPCW}^-(w_0; G, S)$. Define the matrices

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

and

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

For our first theorem, we need the following standard assumptions from the RD estimation literature.

(R1) For $W \neq w_0$, let $\mu(w)$ and $p(w)$ be twice continuously differentiable functions. Let $\mu'(w)$ and $\mu''(w)$ be the first and second derivatives of $\mu(w)$ and similarly for $p'(w)$ and $p''(w)$. Let

$\mu^{+}(w)$ and $\mu''^{+}(w)$ be the first and second derivatives of $\mu^{+}(w)$, and $p^{+}(w)$ and $p''^{+}(w)$ are the first and second derivatives of $p(w)$. Define $\mu'^{-}(w)$ and $\mu''^{-}(w)$ to be the first and second derivative of $\mu^{-}(w)$ and $p'^{-}(w)$ and $p''^{-}(w)$ are first and second derivative of $p^{-}(w)$. Assume there exists $B > 0$ such that $|\mu^{+}(w)|, |\mu'^{+}(w)|, |\mu''^{+}(w)|$ and $|p^{+}(w)|, |p'^{+}(w)|, |p''^{+}(w)|$ are uniformly bounded on $(w_0, w_0 + B]$. Similarly, $|\mu^{-}(w)|, |\mu'^{-}(w)|, |\mu''^{-}(w)|$ and $|p^{-}(w)|, |p'^{-}(w)|, |p''^{-}(w)|$ are uniformly bounded on $[w_0 - B, w_0)$.

- (R2) Assume that $\mu^{+}(w_0), \mu'^{+}(w_0), \mu''^{+}(w_0), \mu^{-}(w_0), \mu'^{-}(w_0), \mu''^{-}(w_0), p^{+}(w_0), p'^{+}(w_0), p''^{+}(w_0), p^{-}(w_0), p'^{-}(w_0)$ and $p''^{-}(w_0)$ are finite.
- (R3) 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 .
- (R4) $\sigma_{IPCW}^2(w; G_0), \sigma_{DR}^2(w; G_0, S^*)$ and $\eta_{IPCW}(w; G_0), \eta_{DR}(w; G_0, S^*)$ are uniformly bounded in a neighborhood of w_0 .
- (R5) Assume that $\sigma_{IPCW}^{2+}(w_0; G_0), \sigma_{DR}^{2+}(w_0; G_0, S^*), \sigma_{IPCW}^{2-}(w_0; G_0), \sigma_{DR}^{2-}(w_0; G_0, S^*)$ and $\eta_{IPCW}^{+}(w_0; G_0), \eta_{DR}^{+}(w_0; G_0, S^*), \eta_{IPCW}^{-}(w_0; G_0), \eta_{DR}^{-}(w_0; G_0, S^*)$ are finite.
- (R6) $\lim_{W_i \uparrow w_0} E \left[|Y_{IPCW,i}(O_i; G_0) - \mu(W_i)|^r \middle| W_i \right]$ and $\lim_{W_i \uparrow w_0} E \left[|Y_{DR,i}(O_i; G_0, S^*) - \mu(W_i)|^r \middle| W_i \right], r = 1, 2, 3$ are finite. We assume similarly when $W_i \downarrow w_0$.
- (R7) K is continuous and symmetric. Moreover, support of K is compact and for any $u, K(u) \geq 0$.
- (R8) The bandwidth satisfies $h \sim n^{-1/5}$ where \sim indicates ‘‘asymptotically equivalent’’.
- (R9) Let $V_n = o_p(1)$. Then

$$E \left[\left(\frac{W_i - w_0}{h} \right)^{j_1} (L_{ih}^+)^{j_2} V_n \right] = O(1); \quad j_1 = 0, \dots, 6, \quad j_2 = 1, 2, 3.$$

Now we show our main result.

Theorem 1. Assume that conditions (C1)-(C5), (R1)-(R9) hold. By Lemma 1-6 in the Supplementary Materials,

$$\begin{aligned} n^{2/5} (\hat{\tau}_{FRD}^{IPCW}(\hat{G}) - \tau_{FRD} - \varphi_{FRD}) &\xrightarrow{d} N(0, \Sigma_{FRD}^{IPCW}(G_0)), \\ n^{2/5} (\hat{\tau}_{FRD}^{DR}(\hat{G}, \hat{S}) - \tau_{FRD} - \varphi_{FRD}) &\xrightarrow{d} N(0, \Sigma_{FRD}^{DR}(G_0, S^*)), \end{aligned}$$

where $\varphi_{FRD}, \Sigma_{FRD}^{IPCW}(G)$ and $\Sigma_{FRD}^{DR}(G, S)$ are defined in the Supplementary Materials.

For the sharp RD case, the result follows easily from Theorem 1 because there is no need to model $Z|W$ for fuzzy RD.

Corollary 4.1. Suppose that conditions (C1)-(C5) and (R1)-(R9) hold. By Lemma 1-6 and Theorem 1 therein,

$$\begin{aligned} n^{2/5}(\hat{\tau}_{SRD}^{IPCW}(\hat{G}) - \tau_{SRD} - \varphi_{SRD}) &\xrightarrow{d} N(0, \Sigma_{SRD}^{IPCW}(G_0)), \\ n^{2/5}(\hat{\tau}_{SRD}^{DR}(\hat{G}, \hat{S}) - \tau_{SRD} - \varphi_{SRD}) &\xrightarrow{d} N(0, \Sigma_{SRD}^{DR}(G_0, S^*)), \end{aligned}$$

where φ_{SRD} , $\Sigma_{SRD}^{IPCW}(G)$ and $\Sigma_{SRD}^{DR}(G, S)$ are defined in the Supplementary Materials.

Now we demonstrate an efficiency result similar to that given in Steingrimsso et al. (2019) for a separate problem. For the sharp RD estimator, due to the “local randomization” result, the DR estimator with true censoring and failure time distributions is a more efficient estimator than the estimators with IPCW and DR transformations, which only involve true censoring distribution. We formally state these in the theorem below:

Theorem 2. *Suppose that conditions (C1)-(C5) and (R1)-(R9) in the Supplementary Materials hold. Let AV ar denote asymptotic variance. Then for the sharp RD estimator,*

$$AV ar\left(\hat{\tau}_{SRD}^{DR}(G_0, S_0)\right) \leq \min \left\{ AV ar\left(\hat{\tau}_{SRD}^{IPCW}(G_0)\right), AV ar\left(\hat{\tau}_{SRD}^{DR}(G_0, S)\right) \right\}.$$

4.3 Bandwidth selection

Once we estimate the censoring unbiased transformation with respect to failure time, we can then apply the existing methodology for RD designs with censored data. The estimated transformation is

$$\begin{aligned} \hat{Y}_{IPCW,i}(O_i; \hat{G}) &= \frac{\Delta_i \tilde{Y}_i}{\hat{G}(\tilde{T}_i)}, \\ \hat{Y}_{DR,i}(O_i; \hat{G}, \hat{S}) &= \frac{\Delta_i \tilde{Y}_i}{\hat{G}(\tilde{T}_i)} + \int_0^{\tilde{T}_i} \frac{\hat{Q}_Y(u|W_i; \hat{S})}{\hat{G}(u)} d\hat{M}_{G,i}(u), \end{aligned}$$

where $\hat{Q}_Y(\cdot|\cdot; \hat{S})$ is an estimator of $Q_Y(\cdot|\cdot; S)$ and

$$\hat{M}_{G,i}(u|W_i) = I(\tilde{T}_i \leq u, \Delta_i = 0) - \int_0^u I(\tilde{T}_i \geq s) d\hat{\Lambda}_G(s),$$

and $\hat{\Lambda}_G(\cdot)$ is a Nelson-Aalen estimator for G . Details of computation of \hat{G} and $\hat{Q}_Y(\cdot|\cdot; \hat{S})$ are provided in the Supplementary Materials.

In standard nonparametric regression, one important issue is bandwidth selection. Ludwig and Miller (2007) propose a mean squared error (MSE) type cross-validation criterion. Let $\hat{a}_L(W, \xi, L)$ be the ξ quantile of the empirical distribution of W using observations $W_i < w_0$ and let $\hat{a}_R(W, 1-\xi)$ be the $1 - \xi$ quantile of the empirical distribution of W using observations $W_i \geq w_0$. Moreover, let $\hat{\alpha}_L^{(Y)}$ and $\hat{\alpha}_R^{(Y)}$ be estimated parameters for α_L and α_R . Criterion from Ludwig and Miller (2007) for uncensored data is

$$\frac{1}{n} \sum_{\hat{a}_L(W, \xi) \leq W_i \leq \hat{a}_R(W, 1-\xi)} \left(Y_i - \hat{\gamma}(W_i) \right)^2,$$

where

$$\hat{\gamma}(w) = \begin{cases} \hat{\alpha}_L(w) & \text{if } w < w_0, \\ \hat{\alpha}_R(w) & \text{if } w \geq w_0. \end{cases}$$

This criterion still works for the unbiased censoring transformation because it uses an MSE-type criterion and does not depend on the variance of Y . We now modify the proposal of Ludwig and Miller (2007) (LM) for censored data. For the sharp RD estimator, we consider

$$\begin{aligned} & (\hat{\alpha}_R^Y(w; \hat{G}, \hat{S}), \hat{\beta}_R^Y(w; \hat{G}, \hat{S})) \\ &= \arg \min \sum_{i=1}^n I(W_i \geq w) \left\{ Y_{DR,i}(O_i; \hat{G}, \hat{S}) - \alpha_R^Y - \beta_R^Y(W_i - w) \right\}^2 K\left(\frac{W_i - w_0}{h}\right), \\ & (\hat{\alpha}_L^Y(w; \hat{G}, \hat{S}), \hat{\beta}_L^Y(w; \hat{G}, \hat{S})) \\ &= \arg \min \sum_{i=1}^n I(W_i < w) \left\{ Y_{DR,i}(O_i; \hat{G}, \hat{S}) - \alpha_L^Y - \beta_L^Y(W_i - w) \right\}^2 K\left(\frac{W_i - w_0}{h}\right). \end{aligned}$$

The LM criterion for sharp RD estimator in censored data is then given by

$$CV_{Y_{DR}}(h; \hat{G}, \hat{S}) = \frac{1}{n} \sum_{\hat{a}_L(W, \xi) \leq W_i \leq \hat{a}_R(W, 1-\xi)} \left(\hat{Y}_{DR,i}(O_i; \hat{G}, \hat{S}) - \hat{\gamma}_{DR}^Y(W_i) \right)^2,$$

where

$$\hat{\gamma}_{DR}^Y(w) = \begin{cases} \hat{\alpha}_{L,DR}^Y(w; \hat{G}, \hat{S}) & \text{if } w < w_0, \\ \hat{\alpha}_{R,DR}^Y(w; \hat{G}, \hat{S}) & \text{if } w \geq w_0. \end{cases}$$

We then choose $\hat{h}_{DR}(\hat{G}, \hat{S}) = \arg \min_h CV_{Y_{DR}}(h; \hat{G}, \hat{S})$. We can derive a similar quantity for $\hat{Y}_{IPCW,i}(O_i; \hat{G}), i = 1, \dots, n$. For the fuzzy RD estimator, we define

$$\begin{aligned} (\hat{\alpha}_R^Z(w), \hat{\beta}_R^Z(w)) &= \arg \min_{\alpha_R^Z, \beta_R^Z} \sum_{i=1}^n I(W_i \geq w) \left\{ Z_i - \alpha_R^Z - \beta_R^Z(W_i - w) \right\}^2 K\left(\frac{W_i - w}{h}\right), \\ (\hat{\alpha}_L^Z(w), \hat{\beta}_L^Z(w)) &= \arg \min_{\alpha_L^Z, \beta_L^Z} \sum_{i=1}^n I(W_i < w) \left\{ Z_i - \alpha_L^Z - \beta_L^Z(W_i - w) \right\}^2 K\left(\frac{W_i - w}{h}\right). \end{aligned}$$

We then obtain

$$CV_Z(h) = \frac{1}{n} \sum_{\hat{a}_L(W, \xi) \leq W_i \leq \hat{a}_R(W, 1-\xi)} \left(Z_i - \hat{\gamma}_Z(W_i) \right)^2,$$

where

$$\hat{\gamma}_Z(w) = \begin{cases} \hat{\alpha}_L^Z(w) & \text{if } w < w_0, \\ \hat{\alpha}_R^Z(w) & \text{if } w \geq w_0. \end{cases}$$

We then obtain $\hat{h}_Z = \arg \min_h CV_Z(h)$. A smaller bandwidth is preferable to reduce the bias of the estimator. Hence for the fuzzy RD estimator, we consider $\min\{\hat{h}_{IPCW}(\hat{G}), \hat{h}_Z\}$ (IPCW) and $\min\{\hat{h}_{DR}(\hat{G}, \hat{S}), \hat{h}_Z\}$ (DR).

4.4 Variance estimation

With the estimation of the censoring and failure time distributions and bandwidth selection, we obtain

$$\begin{aligned} \left(\hat{\alpha}_{R,DR,\hat{h}}^{FRD,Y}(\hat{G}, \hat{S}), \hat{\beta}_{R,DR,\hat{h}}^Y(\hat{G}, \hat{S})\right) &= \arg \min_{\alpha_R^Y, \beta_R^Y} U_{R,DR,\hat{h}}^{FRD,Y}(\alpha_R^Y, \beta_R^Y; \hat{G}, \hat{S}), \\ \left(\hat{\alpha}_{L,DR,\hat{h}}^{FRD,Y}(\hat{G}, \hat{S}), \hat{\beta}_{L,DR,\hat{h}}^Y(\hat{G}, \hat{S})\right) &= \arg \min_{\alpha_L^Y, \beta_L^Y} U_{L,DR,\hat{h}}^{FRD,Y}(\alpha_L^Y, \beta_L^Y; \hat{G}, \hat{S}), \\ \left(\hat{\alpha}_{R,\hat{h}}^{FRD,Z}, \hat{\beta}_{R,\hat{h}}^{FRD,Z}\right) &= \arg \min_{\alpha_R^Z, \beta_R^Z} U_{R,\hat{h}}^{FRD,Z}(\alpha_R^Z, \beta_R^Z), \quad \left(\hat{\alpha}_L^Z, \hat{\beta}_L^Z\right) = \arg \min_{\alpha_L^Z, \beta_L^Z} U_{L,\hat{h}}^{FRD,Z}(\alpha_L^Z, \beta_L^Z), \end{aligned}$$

where $U_{R,DR,\hat{h}}^{FRD,Y}$ and $U_{R,\hat{h}}^{FRD,Z}$ correspond to $U_{R,DR}^{FRD,Y}$ and $U_R^{FRD,Z}$ with an estimated bandwidth, respectively. Estimation functions with IPCW transformation can be similarly defined. In addition to fuzzy RD estimators, estimators for sharp RD designs can also be similarly defined. Hence the proposed fuzzy RD and sharp RD estimators based on \hat{G} and \hat{S} are

$$\begin{aligned} \hat{\tau}_{FRD,\hat{h}}^{DR}(\hat{G}, \hat{S}) &= \left(\hat{\alpha}_{R,DR,\hat{h}}^{FRD,Y}(\hat{G}, \hat{S}) - \hat{\alpha}_{L,DR,\hat{h}}^{FRD,Y}(\hat{G}, \hat{S})\right) / \left(\hat{\alpha}_{R,\hat{h}}^Z - \hat{\alpha}_{L,\hat{h}}^Z\right), \quad \text{and} \\ \hat{\tau}_{SRD,\hat{h}}^{DR}(\hat{G}, \hat{S}) &= \left(\hat{\alpha}_{R,DR,\hat{h}}^{SRD,Y}(\hat{G}, \hat{S}) - \hat{\alpha}_{L,DR,\hat{h}}^{SRD,Y}(\hat{G}, \hat{S})\right), \end{aligned}$$

where $\{\hat{\alpha}_{R,DR,\hat{h}}^{SRD,Y}(\hat{G}, \hat{S}), \hat{\alpha}_{L,DR,\hat{h}}^{SRD,Y}(\hat{G}, \hat{S})\}$ are sharp RD estimators with estimated bandwidth from DR transformation. For variance estimation, one may use the methods based on the asymptotic results in Section 4.2. Let $e_1 = (1, 0)^T$ and $a(u) = (1, u)^T$. Define $g(\cdot)$ to common density of W_i . Using the expressions in the Supplementary Materials, the asymptotic variance of $\hat{\tau}_{SRD}^{DR}(\hat{G}, \hat{S})$ can be expressed as

$$\frac{1}{n} e_1^T \left(\mathbf{\Gamma}_{h+}^{-1} \phi_{YY+,DR} \mathbf{\Gamma}_{h+}^{-1} + \mathbf{\Gamma}_{h-}^{-1} \phi_{YY-,DR} \mathbf{\Gamma}_{h-}^{-1} \right) e_1,$$

where $\mathbf{\Gamma}_{h+}$, $\mathbf{\Gamma}_{h-}$, $\phi_{YY+,DR}$, $\phi_{YY-,DR}$ are defined in the Supplementary Materials. The first approach is to use plug-in residuals for the sandwich variance estimator. The idea of the plugged-in residuals is to compute residuals from transformed DR response and the causal estimate, and next estimate $\phi_{YY+,DR}$ and $\phi_{YY-,DR}$ with residuals, and finally estimate $\mathbf{\Gamma}_{h+}^{-1}$ and $\mathbf{\Gamma}_{h-}^{-1}$.

The second approach is to use a nonparametric nearest neighbor (NN) variance estimator as in Calonico et al. (2014). In this method, we choose observations and compute “local residuals” close to forcing variable in each i and estimate $\phi_{YY+,DR}$ and $\phi_{YY-,DR}$ with these “local residuals”, and finally estimate $\mathbf{\Gamma}_{h+}^{-1}$ and $\mathbf{\Gamma}_{h-}^{-1}$. This procedure gives more weight on the observation close to forcing variable in each i . This approach is advantageous in that it does not require nonparametric

smoothing and is robust (Abadie and Imbens, 2006). The nonparametric bootstrap is another method we consider for standard error estimation.

For implementation of our method, we can use existing software for uncensored data. The R package `rdboost` (Calonico et al., 2015b) is a powerful tool to perform statistical inference for RD designs. To implement the proposed methods, one can simply transform the response by methods in Section 4.3. Then with the transformed quantities as a new response, we can estimate regression coefficients along with standard errors.

5 Simulation Results

We performed simulation studies to evaluate the finite-sample properties of our proposed estimators. The forcing variable W is generated as a $Unif(0, 1)$ random variable. The error variable is generated as $\epsilon \sim N(0, 0.5)$. Regression coefficients are set to be $\beta_{10} = 2$, $\beta_{20} = 1$, and $\beta_{30} = 1$. The response is generated from the following model:

$$T = \exp(\beta_{10} + \beta_{20}W + \beta_{30}I(W \geq 0.5) + \epsilon).$$

Censoring is generated as a $Unif(0, 50)$ random variable that is independent of T and W . Three models for the conditional expectation were considered in the simulation study: Cox, Log-normal, and Log-Logistic models. We use the Kaplan-Meier estimator to estimate G . To ensure positivity of \hat{G} , we truncate \hat{T} by ω where ω is the 95th percentile of observed time for estimation of G (Steingrimsson et al., 2016). The censoring distribution is estimated using the Kaplan-Meier estimator. Sample sizes are $n = 200$ and $n = 400$. The number of bootstraps within each simulation is 50. To select the bandwidth by cross-validation, it is important to select ξ based on the range of dataset. The value of ξ is 0.5. The amount of observed censoring is approximately 51% across the simulations. The kernel function is a triangular function, which is

$$K(u) = 1 - |u|.$$

Table 1 shows finite-sample properties of the estimator $\hat{\tau}_{SRD}$. In the columns denoting standard error calculation and coverage, NN, Plug-in, and Boot denote the nearest neighborhood, plug-in residual, and bootstrap approaches, respectively. For coverage, all the calculations are based on the normal approximation. The IPCW approach is more biased than the DR approach. Except for the bootstrap, in general, the coverage of the estimators satisfies the 95% nominal level. The efficiency gain of the DR approach compared to the IPCW approach is noticeable. The performances of the DR approaches across the conditional expectations are very stable. The results from the DR approach confirm the augmentation theory results from Tsiatis (2007).

In the next set of simulation studies, we consider the fuzzy RD based on a modification of the simulation setting in Yang (2013). We generate $W \sim Unif(-1, 1)$, and let $V = I(W \geq 0)$. Next, we generate $\kappa \sim N(0, 0.25)$ and independent of all these aforementioned variables. The treatment variable is then defined as $Z = I(-0.5 + V + W + \kappa > 0)$. We then generate $\epsilon \sim N(0, 0.25)$, which is independent of aforementioned variables. Failure time is defined as $T = \exp(\beta_{10} + \beta_{20}W +$

$\beta_{30}Z + \epsilon$), where regression coefficients are set to be $\beta_{10} = 2$, $\beta_{20} = 1$, and $\beta_{30} = 1$. The censoring variable is generated as $Unif(0, 50)$. The average censoring rate is approximately 39%. In this case, the denominator for the true value is calculated as

$$\lim_{w \downarrow 0} P(Z = 1|W = w) = P(\kappa > -0.5) = 1 - P(\kappa \leq -0.5) = 1 - \Phi(-2),$$

$$\lim_{w \uparrow 0} P(Z = 1|W = w) = P(\kappa > 0.5) = 1 - P(\kappa \leq 0.5) = 1 - \Phi(2),$$

where Φ is an inverse function of the standard normal cumulative distribution function. Hence, the denominator should be $\Phi(2) - \Phi(-2)$. For the numerator,

$$\lim_{w \downarrow 0} E\{\log(T)|W = w\} = \beta_{10} + \beta_{20} \times 0 + \lim_{w \downarrow 0} P(Z = 1|W = w),$$

$$\lim_{w \uparrow 0} E\{\log(T)|W = w\} = \beta_{10} + \beta_{20} \times 0 + \lim_{w \uparrow 0} P(Z = 1|W = w).$$

Hence, the numerator and denominator are equal so that the average treatment effect for those who comply with the treatment assignment is 1. We use the same conditional expectation methods as in the sharp RD case. Table 2 presents the numerical results for sample sizes $n = 250$ and $n = 500$. For all approaches, the bias is greater than those reported in Table 1. This makes sense because the estimator in fuzzy RD has a denominator that requires estimation via a nonparametric method, which introduces bias. As with the sharp RD situation, the DR method shows good performance regardless of the choice of conditional expectation. The IPCW method has a larger bias than the DR methods. The coverage probability tends to perform better in larger sample sizes.

Table 1: Numerical results when sample size $n = 200$ and $n = 400$ in sharp RD. EPD : empirical standard deviation, SE: mean of standard error, Cover : 95% coverage rate, NN : nearest neighborhood approach, Plug-in : plug-in residuals approach, Boot : bootstrap

n		Bias	EPD	SE			Cover		
				NN	Plug-in	Boot	NN	Plug-in	Boot
200	IPCW	0.074	1.671	1.422	1.390	1.542	0.900	0.894	0.930
	Cox	-0.004	0.122	0.123	0.119	0.121	0.942	0.944	0.946
	Log-norm	-0.008	0.136	0.136	0.132	0.136	0.946	0.944	0.948
	Log-log	-0.008	0.137	0.136	0.132	0.139	0.942	0.942	0.950
400	IPCW	0.091	1.066	1.008	0.995	1.084	0.930	0.930	0.956
	Cox	-0.003	0.093	0.086	0.085	0.085	0.940	0.932	0.934
	Log-norm	-0.004	0.101	0.095	0.093	0.095	0.922	0.920	0.930
	Log-log	-0.004	0.104	0.096	0.094	0.097	0.920	0.924	0.932

Table 2: Numerical results for mean response when sample size $n = 250$ and $n = 500$ in fuzzy RD. EPD: empirical standard deviation, SE: mean of standard error, Cover: 95% coverage rate, NN: nearest neighborhood approach, Plug-in: plug-in residuals approach, Boot: bootstrap

n		Bias	EPD	SE			Cover		
				NN	Plug-in	Boot	NN	Plug-in	Boot
250	IPCW	0.054	1.570	1.270	1.146	1.269	0.905	0.875	0.895
	Cox	0.011	0.237	0.204	0.179	0.210	0.909	0.879	0.895
	Log-norm	0.014	0.249	0.214	0.188	0.219	0.918	0.879	0.901
	Log-log	0.013	0.249	0.214	0.188	0.219	0.920	0.879	0.903
500	IPCW	0.139	1.045	0.919	0.834	1.155	0.928	0.913	0.950
	Cox	0.021	0.186	0.144	0.130	0.182	0.915	0.909	0.962
	Log-norm	0.025	0.191	0.151	0.137	0.189	0.934	0.924	0.956
	Log-log	0.025	0.191	0.151	0.136	0.189	0.932	0.920	0.954

Based on simulation studies, the DR method is recommended as compared to the IPCW method. In practice, since calculations using the AFT model are easier than Cox model and due to theoretical advantages of using a parametric model (as shown in the Supplementary Materials), the AFT model is recommended for the conditional expectation calculation. The choice of error distribution in the AFT model does not appear to make a noticeable difference to the estimation.

6 Real Data Analysis

We now apply the proposed methodology to evaluate whether PSA-based screening strategies have a meaningful impact on prostate-cancer incidence, as well as first cancer incidence of any type as time-to-event data. The PLCO cancer screening trial (Andriole et al., 2009; Shoag et al., 2015) randomized 76,678 men to receive either annual PSA screening for 6 years or no annual screening. Among those who received annual PSA screening from 1993 to 2001, those with a PSA of 4.0 ng/ml at any time were recommended for further workup and biopsy, for example, PSA-based screening strategy, for prostate cancer diagnosis. In the context of the RD design, this practice naturally creates a sharp RD design when treating PSA as the forcing variable. We therefore evaluated the role of additional workup and biopsy, as prompted by a PSA cutoff of 4.0 ng/ml, in cancer incidence occurred at the first (the first cancer incidence) and for prostate cancer (prostate cancer incidence). To simplify our discussions, we focus on the role of PSA-based screening at the time of study entry among those who were not previously tested for PSA. While the role of PSA-based screening among those who had been exposed to PSA may be also of interest, its analysis involves methodology that is still being developed and is not discussed here.

Table 3: Descriptive statistics for baseline covariates. SD: standard deviation DRE: Digital Rectal Examination BPH: benign prostatic hypertrophy

	Mean (SD)	Number of people
PSA	1.935 (10.385)	33014
PSA \geq 4	9.304 (35.879)	2628 (8%)
PSA < 4	1.297 (0.882)	30386 (92%)
DRE	-	33014
Negative or Abnormal, Non-Suspicious	-	29348 (88.9%)
Abnormal, Suspicious	-	2385 (7.2%)
Others	-	1281 (3.9%)
Ethnicity	-	33014
Black	-	1427 (4.3%)
Non-black	-	31587 (95.7%)
Hispanic	-	33014
Yes	-	704 (2.1%)
No	-	32310 (97.9%)
Marital	-	33014
Married and living as married	-	27550 (83.4%)
Others	-	5464 (16.6%)
Existence of enlarged prostate or BPH	-	33014
Yes	-	7347 (22.3%)
No	-	25667 (77.7%)

Table 4: Exact matching with Digital Rectal Examination (DRE) result, ethnicity, Hispanic, marital status, existence of enlarged prostate or benign prostatic hypertrophy (BPH).

PSA	First cancer incidence			Prostate cancer incidence		
	Est	Std.err	p-value	Est	Std.err	p-value
0-15 ($n = 32814$)	-0.077	0.207	0.708	-0.082	0.286	0.775
0-25 ($n = 32950$)	-0.056	0.201	0.779	-0.07	0.292	0.810
0-50 ($n = 32987$)	-0.017	0.174	0.922	-0.08	0.322	0.803
0-100 ($n = 33000$)	-0.019	0.185	0.919	-0.093	0.358	0.796
All ($n = 33014$)	-0.025	0.228	0.912	-0.028	0.434	0.949

Although the local randomization property holds for RD design, this property assumes that treatment assignment is independent of other covariates. To alleviate the potential concerns due to the associations between PSA-based screening strategy and covariates at study entry, we take a view that is conceptually similar to a double propensity score matching approach (Austin, 2017). In addition to RD design, we also conduct a propensity score matching with Digital Rectal Examination (DRE) result, ethnicity, Hispanic, marital status, existence of enlarged prostate, or benign prostatic hypertrophy (BPH) at the baseline.

After matching, there were 33,014 men, including 2628 with PSA level greater than 4.0 ng/ml and 30,386 with PSA level less than or equal to 4.0 ng/ml at study entry. The range of PSA by 2.5% and 97.5% quantiles is 0.24 and 6.88, respectively. Table 3 presents descriptive statistics for the final dataset. In this sample, censoring rates for the first cancer incidence and prostate cancer incidence are 75.9% and 87.6%, respectively.

We select the DR approach with conditional expectation method on the parametric AFT log-normal model. Since the distribution of PSA is highly right-skewed, we investigate treatment effect in various ranges of PSA. Table 4 reports results from sharp RD. We use the nearest neighbor (NN) approach for calculation of standard error. Based on Table 4, there is no significant screening effect from baseline PSA threshold level 4.0 mg/nL for either the first cancer incidence or prostate cancer-specific incidences. The screening effects with a cutoff of 4.0 mg/nL for the first and prostate cancer incidence are slightly negative, which implies that screening slightly decreases time to cancer diagnosis (either prostate cancer only or any cancer) although the effect is not statistically significant.

We also create data-driven RD plots (Calonico et al., 2015a); the results are shown in Figure 1 and 2. These data-driven plots approximate regression functions by local sample means with evenly- or quantile-spaced bins. These plots also reflect the variability in the data (Calonico et al., 2015b). We use 40 bins on each side of the cutoff. These plots are useful to capture the variability of the transformed response with cutoffs and to check causal effect graphically. These four plots also indicate no treatment effect using a baseline PSA threshold level of 4.0 mg/nL and they show decreasing trend of transformed time of logarithm of failure time across PSA level. Our qualitative findings of no effect of the PSA threshold on survival are concordant with those in Shoag et al.

(2015), although there are major differences between the two analyses. Their approach compares survival functions, while our methodology directly models the time to death. The techniques in this paper allow for identification of an individual-specific causal effect, which is not available in Shoag et al. (2015).

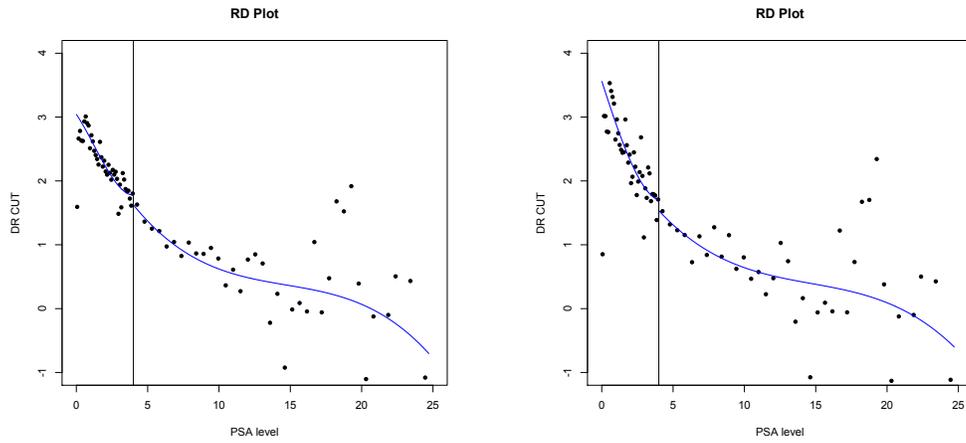


Figure 1: Data-driven RD plot by Calonico, Cattaneo, and Titiunik (2015a) for the first cancer incidence (left) and prostate cancer incidence (right) with PSA range 0-25

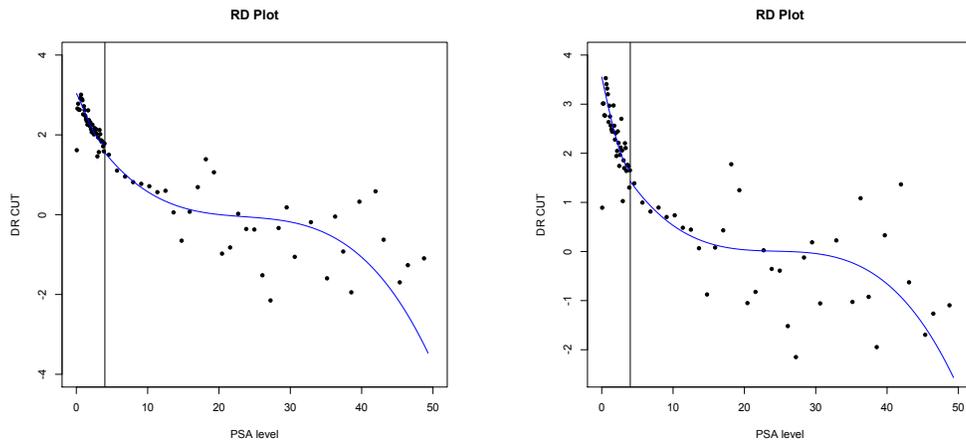


Figure 2: Data-driven RD plot by Calonico, Cattaneo, and Titiunik (2015a) for the first cancer incidence (left) and prostate cancer incidence (right) with PSA range 0-50

7 Conclusion

We have proposed new estimators of causal effects with censored data in RD designs. Simulation studies reveal that the DR approach yields a more efficient estimator than IPCW that appears to have better finite-sample properties in the simulation studies we considered. Moreover, the bias of the DR method is smaller than that of the IPCW method.

The main contribution of our paper is that we provide, for the first time to our knowledge, a formal framework to estimate the causal effect in regression discontinuity designs using right censored time-to-event data. However, there are limitations of our approach. Given the close connections of our approach to those of RD methods for uncensored data, we are unable to estimate the adjusted survival distributions of each treatment group. In addition, our approach requires specification of the censoring and outcome regression models. If one can correctly specify the outcome regression model correctly, then its efficiency will be higher than our proposed approach. In addition, if both the censoring and outcome regression models are misspecified, then this will lead to very biased answers. To enable implementation by users, we have made our code available as R scripts at <https://github.com/Ghoshlab/SurvRD>.

We consider only one forcing variable for analysis. However, in practice, data may contain several forcing variables, and they may provide additional information for treatment effect. There are two possible scenarios: (i) the forcing variable is a function of multiple covariates and (ii) there is one forcing variable correlated with other covariates, as shown in our data analysis. Imbens and Zajonc (2009); Zajonc (2012) examine the situation of multiple forcing variables. Recently, Calonico et al. (2019) proposed a covariate adjustment approach in RD. It is of significant interest to include covariates or consider the composite forcing variable in RD analysis. Our future work is to propose an estimation procedure of the treatment effect in RD adjusting for effects of other covariates. In this case, we may need to model censoring distribution given other covariates. Although we only have one forcing variable without covariates, the forcing variable is expected to have multiple cut-offs. Cattaneo et al. (2016) discuss the multi-cutoff problem. This is also an interesting future work. In the data analysis, we use matching process and apply our RD procedure. However, we do not consider randomness of matching in the variance estimation. It is also an compelling future work to consider matching in the RD.

We have adapted the LM approach for bandwidth selection. Imbens and Kalyanaraman (2012) propose optimal bandwidth selection based on mean square error approximation and Calonico et al. (2014) propose bandwidth selection that helps bias correction. These have elegant asymptotic theory regarding their bandwidth selection proposals. This is currently under investigation.

Acknowledgements

The second author would like to acknowledge the support of NCI U10 CA180822. The third author would like to acknowledge the support of NIH R01 CA129102 and NSF DMS 1914937.

References

- Abadie, A. and Imbens, G. W. (2006), "Large sample properties of matching estimators for average treatment effects," *Econometrica*, 74, 235–267.
- Andriole, G. L., Crawford, E. D., Grubb III, R. L., Buys, S. S., Chia, D., Church, T. R., Fouad, M. N., Gelmann, E. P., Kvale, P. A., Reding, D. J., et al. (2009), "Mortality results from a randomized prostate-cancer screening trial," *New England Journal of Medicine*, 360, 1310–1319.
- Austin, P. C. (2017), "Double propensity-score adjustment: a solution to design bias or bias due to incomplete matching," *Statistical Methods in Medical Research*, 26, 201–222.
- Bai, X., Tsiatis, A. A., and O'Brien, S. M. (2013), "Doubly-Robust Estimators of Treatment-Specific Survival Distributions in Observational Studies with Stratified Sampling," *Biometrics*, 69, 830–839.
- Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L., and Bärnighausen, T. (2014), "Regression discontinuity designs in epidemiology: causal inference without randomized trials," *Epidemiology (Cambridge, Mass.)*, 25, 729.
- Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019), "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 101, 442–451.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014), "Robust nonparametric confidence intervals for regression-discontinuity designs," *Econometrica*, 82, 2295–2326.
- (2015a), "Optimal data-driven regression discontinuity plots," *Journal of the American Statistical Association*, 110, 1753–1769.
- (2015b), "rdrubust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R J.*, 7, 38.
- Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., and Keele, L. (2016), "Interpreting regression discontinuity designs with multiple cutoffs," *The Journal of Politics*, 78, 1229–1248.
- Fan, J. and Gijbels, I. (1994), "Censored regression: local linear approximations and their applications," *Journal of the American Statistical Association*, 89, 560–570.
- (1996), *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, vol. 66, CRC Press.

- Hahn, J., Todd, P., and Van der Klaauw, W. (1999), "Evaluating the effect of an antidiscrimination law using a regression-discontinuity design," Tech. rep., National bureau of economic research.
- (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.
- Imbens, G. and Zajonc, T. (2009), "Regression discontinuity design with vector-argument assignment rules," *Unpublished paper*.
- Imbens, G. W. and Lemieux, T. (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142, 615–635.
- Lee, D. S. (2008), "Randomized experiments from non-random selection in US House elections," *Journal of Econometrics*, 142, 675–697.
- Lee, D. S. and Lemieux, T. (2010), "Regression discontinuity designs in economics," *Journal of Economic Literature*, 48, 281–355.
- Lesik, S. A. (2007), "Do developmental mathematics programs have a causal impact on student retention? An application of discrete-time survival and regression-discontinuity analysis," *Research in Higher Education*, 48, 583–608.
- Li, Q. and Racine, J. (2003), "Nonparametric estimation of distributions with categorical and continuous data," *Journal of Multivariate Analysis*, 86, 266–292.
- Ludwig, J. and Miller, D. L. (2007), "Does Head Start improve children's life chances? Evidence from a regression discontinuity design," *The Quarterly Journal of Economics*, 122, 159–208.
- McCrary, J. (2008), "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.
- 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.
- Okumura, H. (2011), "Kernel regression for binary response data," *Memoirs of The Graduate School of Science and Engineering, Shimane University Series B: Mathematics*, 44, 33–53.
- Rosenbaum, P. R. and Rubin, D. B. (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41–55.
- Rubin, D. and van der Laan, M. J. (2007), "A doubly robust censoring unbiased transformation," *International Journal of Biostatistics*, 3, 1–19.

- Shoag, J., Halpern, J., Eisner, B., Lee, R., Mittal, S., Barbieri, C. E., 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.
- Steingrimsson, J. A., Diao, L., Molinaro, A. M., and Strawderman, R. L. (2016), “Doubly robust survival trees,” *Statistics in Medicine*, 35, 3595–3612.
- Steingrimsson, J. A., Diao, L., and Strawderman, R. L. (2019), “Censoring unbiased regression trees and ensembles,” *Journal of the American Statistical Association*, 114, 370–383.
- Thistlethwaite, D. L. and Campbell, D. T. (1960), “Regression-discontinuity analysis: An alternative to the ex post facto experiment,” *Journal of Educational psychology*, 51, 309.
- Tsiatis, A. (2007), *Semiparametric theory and missing data*, Springer Science & Business Media.
- Zajonc, T. (2012), “Essays on causal inference for public policy,” Ph.D. thesis, Harvard University.

Received: February 27, 2021

Accepted: April 7, 2021