

Regression Discontinuity Designs: A Decision Theoretic Approach

Panayiota Constantinou¹ and Aidan G. O’Keeffe*²

¹*School of Mathematics, University of Bristol, BS8 1TH, UK.*

²*Department of Statistical Science, University College London, WC1E 6BT, UK.*

January 5, 2016

Abstract

The regression discontinuity design (RDD) is a quasi-experimental design that can be used to identify and estimate the causal effect of a treatment using observational data. In an RDD, a pre-specified rule is used for treatment assignment, whereby a subject is assigned to the treatment (control) group whenever their observed value of a specific continuous variable is greater than or equal to (is less than) a fixed threshold. Sharp RDDs occur when guidelines are strictly adhered to and fuzzy RDDs occur when the guidelines are not strictly adhered to. In this paper, we take a rigorous decision theoretic approach to formally study causal effect identification and estimation in both sharp and fuzzy RDDs. We use the language and calculus of conditional independence to express and explore in a clear and precise manner the conditions implied by each RDD and investigate additional assumptions under which the identification of the average causal effect at the threshold can be achieved. We apply the methodology in an example concerning the relationship between statins (a class of cholesterol-lowering drugs) and low density lipoprotein (LDL) cholesterol using a real set of primary care data.

1 Introduction

Regression discontinuity designs (RDDs) were first developed during the 1960s with the aim of estimating the causal effect of an intervention using observational data (Thistlethwaite and Campbell (1960)). Such designs occur where a decision to apply a treatment is linked to some continuous ‘assignment variable’ through a decision rule. Here, a ‘treatment’ refers to any action that is taken with the aim of affecting some response in an outcome variable of interest. For example, in an educational context, the ‘treatment’ might be the assignment of students to a particular school based on an admissions test result (the assignment variable), where students are given a place at the school based on their attainment of a test ‘pass mark’ (the decision rule). In a medical context the ‘treatment’ might be the prescription of a drug based on the value of a subject’s diagnostic test result (the assignment variable). Typically, a decision rule takes the form of a ‘treatment threshold’ whereby a subject receives the treatment if the subject’s assignment variable value lies at or above the treatment threshold and the subject does not receive the treatment if the subject’s assignment variable lies below the treatment threshold, at a pre-specified point in time.

Assuming that subjects with similar assignment variable values are exchangeable and that the decision rule is identical for all subjects, a comparison of outcome variable values between subjects whose assignment variables lie ‘just above’ the threshold and ‘just below’ the threshold might be considered appropriate for the calculation of a causal effect of the treatment on the outcome of interest. For example, in a medical context, statin therapy - that is, prescription of lipid-modifying drugs (the treatment) may be prescribed to patients only if their 10-year risk of developing cardiovascular disease (the assignment variable) lies above a pre-specified threshold of 20% (the treatment threshold). Patients whose risk scores lie close to 20% are perhaps likely to be similar with respect to their age, lifestyle and other variables implying that we might have a population of exchangeable subjects to whom an external decision rule is applied to determine allocation to statins.

*E-mail address: a.o’keeffe@ucl.ac.uk; Corresponding author

RDDs have been applied in many areas of economics and social science (see, for example, Berk and Leeuw (1999), van der Klaauw (2002), van der Klaauw (2008), Lalive (2008)). RDDs have not been used extensively in medicine and epidemiology, although their use and development is increasing in this field (Linden et al. (2006), Almond and Doyle (2011), Linden and Adams (2012), Bor et al. (2014), Vandembroucke and le Cessie (2014), O’Keeffe et al. (2014)). The increasing availability of large observational healthcare databases suggests that the RDD could be considered more widely and developed for use in biostatistics and epidemiology.

Decision theory has been developed and used as an approach to causal inference (Rubin (1978), Heckerman and Shachter (1995), Dawid (2000)). By definition, an RDD includes a decision rule and, as such, the link between the decision theoretic approach to causal inference and the RDD appears natural. A decision theoretic approach typically involves a set of assumptions made about the probabilistic behaviour of variables of interest across a variety of regimes. Different regimes might refer to, for example, different time points, locations or contexts. In our context, we consider a regime to be either ‘interventional’, when it results from an external intervention (for example, by the statistician) which forces the treatment variable to take on a particular value, or ‘observational’, when it results from an observed setting where the treatment was determined by a number of known or unknown factors (e.g. doctor or patient preference for treatment, governmental guidelines etc.). RDDs represent examples of ‘observational regimes’ and are commonly classified as either ‘sharp’ or ‘fuzzy’. In a sharp RDD, subjects are assigned the treatment strictly according to the treatment threshold. In other words, at the point where the decision to assign a treatment is taken, all subjects who are given the treatment will exhibit an assignment variable value at or above the treatment threshold and, conversely, all subjects who do not receive the treatment will exhibit an assignment variable value below the treatment threshold. In a fuzzy RDD, this might not be the case; there may be some subjects whose receipt (or not) of the treatment is contrary to what is indicated by the decision rule. Figure 1 shows plots of example fuzzy and sharp RDDs, for a continuous assignment variable $x \in \mathcal{X} \subseteq \mathbb{R}$ such that a subject receives the treatment if $x \geq x_0$, where $x_0 \in \mathcal{X}$ defines the treatment threshold level. In these example plots, we assume a continuous outcome variable of interest, denoted $y \in \mathcal{Y} \subseteq \mathbb{R}$.

In this work, we define and formulate a decision-theoretic approach to the RDD which, to our knowledge, has not been attempted. We use ideas of conditional independence (Dawid (1979), Dawid (1980)) to describe assumptions concerning the variables defined in a RDD and how these assumptions relate to considered regimes. We derive appropriate causal effect estimators for sharp and fuzzy RDDs and present examples of treatment effect evaluation using a real example concerning the initiation of statin therapy for the prevention of cardiovascular disease in UK primary care.

2 Notation and Terminology within a Decision Theoretic framework

In this section, we define the notation and terminology that will be used within the Decision Theoretic (DT) framework of statistical causality. The DT framework was introduced by Dawid (2000) and has been used to address a variety of problems in causality (Geneletti, 2007; Dawid and Didelez, 2010; Dawid and Constantinou, 2014). In this framework, we acknowledge that the distributions of stochastic variables of interest are different, in principle, between different regimes (observational or interventional). We explore relationships that may be assumed between the differing distributions using the language and calculus of conditional independence (Dawid, 1979, 1980; Constantinou, 2014). The example that will be used throughout the paper is the one described previously, where statins may be prescribed to patients whose risk of developing cardiovascular disease (CVD) within 10 years exceeds 20%. Our objective is to make causal inference about the effect of the treatment for those individuals who have measurements around the threshold and thus inform policy makers about the effectiveness of statins on lowering LDL cholesterol when prescribed according to the guidelines. We note that everything we describe using this specific example, may be generalised to a variety of contexts.

First, we consider the treatment variable T (prescription of statins). For simplicity, we consider T to be binary, taking the value 0 when the treatment is not administered and value 1 when the treatment

is administered. We consider X to be a continuous assignment variable (10-year CVD risk score) using which the treatment guidelines are applied. In addition, we consider a threshold x_0 , where, when $X \geq x_0$ the guidelines recommend treatment and when $X < x_0$ the guidelines recommend no treatment. In our example, $x_0 = 20\%$, as recommended by United Kingdom’s National Institute for Health and Care Excellence (NICE) guidelines (NICE (2008)). Associated with these measurements is the binary threshold indicator Z which denotes the choice for treatment as imposed by the guidelines. So $Z = 0$, when $X < x_0$ and $Z = 1$, when $X \geq x_0$. We consider the response variable of interest (LDL cholesterol level) and denote this by Y . We aim to identify the causal effect of T on Y and estimate this for patients around the threshold. Further to the above variables, we consider confounding variables; in other words, variables that can influence the treatment choice and at the same time affect the response variable Y . We denote these variables by C and note that a subset of C may be taken into account for determining X . In our example, some variables such as age, gender, smoking status, cholesterol levels, etc., are those using which the 10-year risk score is calculated but some variables, such as a doctor’s preference for treatment, patient’s preference for treatment, allergies, reactions to previous treatments, etc., might be additional to those encompassed in X (i.e., X is a function of C). In the sharp design none of these variables (except through X) can affect the treatment that will be administered. However, in the fuzzy design some of these variables might outweigh the guidelines, especially if the doctor believes that the guidelines did not put appropriate weight on specific measurements considered important and that the patient would benefit from a different treatment than that which the guidelines recommend.

Further to the stochastic variables, we introduce a non-stochastic variable, Σ , which will indicate the operating regime. Interpreting as *causal* the *effect of an intervention*, we want to compare the distribution of the response variable Y , around the threshold x_0 , between the two interventional regimes; the regime where we intervene to give treatment with the regime where we intervene not to give treatment. These two interventional regimes are considered otherwise identical. However, the regimes from which we obtain data (sharp or fuzzy RDD) do not necessarily represent interventional settings but only provide a (usually incomplete) record of observations as they have been generated by Nature and collected by the GP/hospital. Thus, in order to make inference for the interventional regimes using information from the observational regime, we need to introduce and justify assumptions that connect the probabilistic behaviour of the random variables of interest across the differing regimes. Here, we emphasise that the regime indicator Σ has no uncertainty associated with it, and serves only as a parameter to index the regimes. More precisely,

$$\Sigma = \begin{cases} 0, & \text{for the interventional regime under control treatment,} \\ 1, & \text{for the interventional regime under active treatment,} \\ s, & \text{for the observational regime under a sharp RDD,} \\ f, & \text{for the observational regime under a fuzzy RDD.} \end{cases}$$

We emphasise that any probabilistic statement concerning the stochastic variables should be made, explicitly or implicitly, conditional on the value of the regime. Thus, we will write $\mathbb{P}_0(Y)$ to denote the distribution of Y under the interventional regime $\Sigma = 0$, $\mathbb{E}_s(Y | X = x_0)$ to denote the conditional distribution of Y given $X = x_0$, under the observational sharp RDD $\Sigma = s$, and use similar notation throughout the paper. Note that when an observational regime is in operation, the distribution of all the random variables involved, and in particular the treatment variable T , will be determined by Nature. On the contrary, when an interventional regime is in operation, the treatment T will take the same value as the one of the regime with probability 1, i.e., $\mathbb{P}_0(T = 0) = 1$ and $\mathbb{P}_1(T = 1) = 1$. Hence, in the interventional regimes, T is a degenerate random variable with all probability on one value, the same as the value of Σ .

Our principal aim is to measure the impact of the treatment intervention on Y , at the threshold x_0 . As such, the quantity we seek to identify and estimate is the *Average Causal Effect at the threshold x_0* , denoted by ACE_{x_0} , where

$$ACE_{x_0} := \mathbb{E}_1(Y | X = x_0) - \mathbb{E}_0(Y | X = x_0). \quad (1)$$

The ACE_{x_0} is the difference in expectation of Y , between the two interventional regimes $\Sigma = 0$ and $\Sigma = 1$, at the threshold x_0 . Thus, a non-zero value admits a causal interpretation. In the following

sections, we will explore assumptions, expressed in the language of conditional independence, that allow the identification of the ACE_{x_0} from the corresponding observational regime (sharp or fuzzy RDD).

3 Conditional independence

Conditional independence in a form that encompasses simultaneously stochastic and non-stochastic variables (extended conditional independence) was first introduced by Dawid (1979, 1980) and the required calculus, known as the axioms of conditional independence, has been established under conditions by Constantinou (2014). In this section, we give definitions for extended conditional independence as they will be used in the paper and state the properties that will be needed in the following section. Where theorems and lemmas are given, corresponding proofs are provided in the appendix for this paper.

We denote by \mathcal{S} the set of contemplated regimes, where $\mathcal{S} \subseteq \{0, 1, s, f\}$. Also, X_1, X_2, \dots denote stochastic variables and Σ denotes the non-stochastic regime indicator that takes values σ in the set \mathcal{S} . To specify explicitly the underlying probability measure considered in regime σ , we write \mathbb{P}_σ .

Definition 3.1. *We say that X_1 is (conditionally) independent of X_2 given (X_3, Σ) and write $X_1 \perp\!\!\!\perp X_2 \mid (X_3, \Sigma)$ if, for any $\sigma \in \mathcal{S}$ and any (bounded, real and measurable) function $h(X_1)$,*

$$\mathbb{E}_\sigma[h(X_1) \mid X_2, X_3] = \mathbb{E}_\sigma[h(X_1) \mid X_3] \quad \text{a.s. } [\mathbb{P}_\sigma]. \quad (2)$$

In other words, $X_1 \perp\!\!\!\perp X_2 \mid (X_3, \Sigma)$ if, separately in each regime $\sigma \in \mathcal{S}$, the distribution of X_1 given (X_2, X_3) depends, in fact, only on X_3 . So under each regime, upon information on X_3 , further information on X_2 becomes redundant for making probabilistic inference about X_1 . If such a statement is not true, we will write instead $X_1 \not\perp\!\!\!\perp X_2 \mid (X_3, \Sigma)$.

Definition 3.2. *We say that X_1 is (conditionally) independent of Σ given X_2 , and write $X_1 \perp\!\!\!\perp \Sigma \mid X_2$, if for any (bounded, real and measurable) function $h(X_1)$, there exists a function $w(Z)$ such that, for all $\sigma \in \mathcal{S}$,*

$$\mathbb{E}_\sigma[h(X_1) \mid X_2] = w(X_2) \quad \text{a.s. } [\mathbb{P}_\sigma]. \quad (3)$$

Interpreting the above definition, $X_1 \perp\!\!\!\perp \Sigma \mid X_2$ requires that, there exists a function $w(X_2)$, which does not depend on the regime, and can serve as a version of the conditional expectation $\mathbb{E}[h(X_1) \mid X_2]$ simultaneously in all regimes. So (3) provides a link between the (conditional) distribution of $X_1 \mid X_2$ in all $s \in \mathcal{S}$. Intuitively, $X_1 \perp\!\!\!\perp \Sigma \mid X_2$ reflects the understanding that, upon information on X_2 , further information on the operating regime becomes irrelevant for making probabilistic inference about X_1 . Similarly to the above, if such a statement is not true, we will write instead $X_1 \not\perp\!\!\!\perp \Sigma \mid X_2$.

Starting with a set of assumptions (which we believe represent the problem under study), we want to explore if they lead to identification of the ACE_{x_0} from observational data. Identification can follow directly (by definition of such assumptions) or indirectly (by further properties which can be induced by such assumptions). In order to explore such paths, we use the axioms of conditional independence. In Theorem 3.3, we write $W \preceq Y$ to mean that W is a function of Y .

Theorem 3.3 (Axioms of conditional independence). *Let X, Y, Z, W be random variables. Then the following properties hold.*

- P1. [Symmetry] $X \perp\!\!\!\perp Y \mid Z \Rightarrow Y \perp\!\!\!\perp X \mid Z$.
- P2. $X \perp\!\!\!\perp Y \mid X$.
- P3. [Decomposition] $X \perp\!\!\!\perp Y \mid Z$ and $W \preceq Y \Rightarrow X \perp\!\!\!\perp W \mid Z$.
- P4. [Weak Union] $X \perp\!\!\!\perp Y \mid Z$ and $W \preceq Y \Rightarrow X \perp\!\!\!\perp Y \mid (W, Z)$.
- P5. [Contraction] $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z) \Rightarrow X \perp\!\!\!\perp (Y, W) \mid Z$.

Theorem 3.3, stated above for stochastic variables, also holds under conditions when some of the variables involved are non-stochastic and conditional independence is defined as in Definition 3.1 and Definition 3.2 (Constantinou, 2014). More precisely, suppose that we are given a collection of extended conditional independence properties. Any deduction made using the axioms of stochastic conditional independence will be valid, so long as, in both premises and conclusions, no non-stochastic variables appear in the left-most term in a conditional independence statement and one of the following conditions holds: the regime space is discrete or the stochastic variables involved are discrete or there exists a dominating regime. In our context, we only consider a finite regime space ($\mathcal{S} = \{0, 1, s\}$ or $\mathcal{S} = \{0, 1, f\}$) which validates the use of Theorem 3.3 for stochastic and non-stochastic variables together.

4 Identification of the Average Causal Effect at the Threshold

To compute the ACE_{x_0} , we explore conditions that allow us to identify the conditional expectations $\mathbb{E}_1(Y | X = x_0)$ and $\mathbb{E}_0(Y | X = x_0)$ from observational data. Before we examine the sharp and fuzzy RDD separately, we assume that the following two properties ((4a) and (4b)) hold regardless of which is the operating observational regime.

Condition 4.1 (Sufficiency).

$$(C, X) \perp\!\!\!\perp \Sigma \tag{4a}$$

$$Y \perp\!\!\!\perp \Sigma \mid (C, X, T) \tag{4b}$$

Variables (C, X) which satisfy (4a) are called *covariates* and variables (C, X) which satisfy (4a) and (4b) are called *sufficient covariates* (Guo and Dawid, 2010). While the formal definitions for these properties follow from Definition 3.2 (for either $\mathcal{S} = \{s, 0, 1\}$ or $\mathcal{S} = \{f, 0, 1\}$), here we emphasize on the intuitive understanding. Property (4a) requires that the distribution of (C, X) is the same in all regimes. We consider this assumption appropriate as (C, X) represent attributes of the individuals which are determined independently of *how* the treatment is allocated. Thus the joint distribution of (C, X) is considered independent of the regime Σ . Property (4b) requires that the conditional distribution of Y given (C, X, T) also does not depend on the regime. Informally, this means that conditioning on (C, X, T) we do not need further information on the regime Σ to make probabilistic inference about Y . Property (4b) is problem-specific as there may be several (distinct) choices for C , or none at all that validate this property. It lies on the researcher to identify appropriate (possibly multivariate) covariates C so that (4b) holds, but without it further progress seems impossible. Property (4b) has also been described as ‘strongly ignorable treatment assignment, given U’ (Rosenbaum and Rubin, 1983).

Conditional independence properties (4a) and (4b) can be represented graphically (see Figure 2). In general, we shall present corresponding influence diagrams (when such diagrams exist) but, for a detailed account of the semantics of influence diagrams and how they can be used to graphically derive further conditional independence properties, the reader is referred to Dawid (2002) and Cowell et al. (2007). We note that graphical representation, when available, may be helpful as it provides a visually transparent way of deriving further conditional independence properties. However, it is never essential and all that can be achieved using the graphical approach, and more, can be achieved using an algebraic approach.

4.1 Sharp Design

The property that characterises the sharp RDD and separates it from the fuzzy RDD is that the guidelines are strictly adhered to. This property can be expressed mathematically as follows.

Property 4.2.

$$T \perp\!\!\!\perp C \mid (X, \Sigma = s) \tag{5}$$

Property 4.2 states that in the observational regime $\Sigma = s$, T does not depend on C , given X . That is, the only information that is used to decide on the treatment T , is that provided by the assignment variable

X , and no further information that might be contained in C is relevant. This property is trivially true for the interventional regimes as well. Once we condition on the value of the interventional regime, the treatment takes the same value as the value of the regime with probability one and further information on any other variable becomes irrelevant. Thus, when the observational regime represents a sharp RDD and $\mathcal{S} = \{0, 1, s\}$, Condition 4.3 below holds.

Condition 4.3.

$$T \perp\!\!\!\perp C \mid (X, \Sigma) \quad (6)$$

We can represent (4a), (4b) and (6) graphically using an influence diagram (see Figure 3). Combining Condition 4.1 and Condition 4.3, we can show that $Y \perp\!\!\!\perp \Sigma \mid (X, T)$.

Theorem 4.4. *Let (C, X) be sufficient covariates and assume that Condition 4.3 holds. Then,*

$$Y \perp\!\!\!\perp \Sigma \mid (X, T). \quad (7)$$

Theorem 4.4 implies that the outcome of interest Y , no longer depends on the regime Σ , once we have information on the assignment variable X and the treatment variable T . What is important to notice about this result, is that it allows us to transfer probabilistic information about Y , between the regimes, completely disregarding the extra information that is provided by C . This result (see Theorem 4.6), allows us to identify the ACE_{x_0} from the observational sharp RDD.

The last condition that is required, in order to justify estimation of the ACE_{x_0} from observations *around* the threshold (as opposed to observations *at* the threshold), is the continuity assumption, expressed in Condition 4.5.

Condition 4.5 (Continuity in the interventional regimes).

$$\mathbb{E}_0(Y \mid X = x_0) = \lim_{\epsilon \downarrow 0} \mathbb{E}_0(Y \mid X = x_0 - \epsilon) \quad (8a)$$

$$\mathbb{E}_1(Y \mid X = x_0) = \lim_{\epsilon \downarrow 0} \mathbb{E}_1(Y \mid X = x_0 + \epsilon) \quad (8b)$$

Condition 4.5 requires that the expectation of the variable of interest Y is left (right) continuous at the threshold x_0 , in the interventional regime $\Sigma = 0$ ($\Sigma = 1$). Violation of this assumption would suggest that a change in the expectation of Y is due to a change in X (around the threshold x_0) and not necessarily a change in T . Analogues of this assumption can also be found in the counterfactual framework (Hahn et al., 2001; Lee, 2008).

Theorem 4.6. *Let (C, X) be sufficient covariates and assume that Condition 4.3 and Condition 4.5 hold. Then,*

$$ACE_{x_0} = \lim_{\epsilon \downarrow 0} [\mathbb{E}_s(Y \mid X = x_0 + \epsilon, T = 1) - \mathbb{E}_s(Y \mid X = x_0 - \epsilon, T = 0)].$$

This theorem shall be used to prove Theorem 4.7 (see supporting material).

The RDD has been linked with the instrumental variables (IV) framework as the threshold indicator Z satisfies the properties of a binary IV (Imbens and Lemieux, 2008; Didelez et al., 2010; Geneletti et al., 2015). For the sharp RDD, it appears that we do not need to use the properties of an IV, in order to prove an alternative identification formula based on the threshold indicator Z .

Theorem 4.7. *Let (C, X) be sufficient covariates and assume that Condition 4.3 and Condition 4.5 hold. Then,*

$$ACE_{x_0} = \lim_{\epsilon \downarrow 0} [\mathbb{E}_s(Y \mid X = x_0 + \epsilon, Z = 1) - \mathbb{E}_s(Y \mid X = x_0 - \epsilon, Z = 0)].$$

Hence, the ACE_{x_0} is identified under a sharp RDD. We now turn our attention to the, more common, fuzzy RDD.

4.2 Fuzzy Design

In the fuzzy RDD, in addition to information contained in X , further information contained in C is taken into account in order to determine the treatment T . Thus $T \not\perp\!\!\!\perp C \mid (X, \Sigma = f)$ and, as a consequence, we cannot derive (7) which allows us to identify the ACE_{x_0} from the observational regime. In this section, we explore different conditions which allow identification of the ACE_{x_0} from a fuzzy RDD.

Condition 4.8. For $t = 0, 1$,

$$\mathbb{P}_f(T = t \mid C, X) > 0 \quad \text{a.s. } [\mathbb{P}_f] \quad (9)$$

Variables (C, X) which satisfy (4a), (4b) and (9) are called *strongly sufficient covariates* (Guo and Dawid, 2010). Condition 4.8 states that, given any observable values of (C, X) , there is positive probability to observe both active and control treatment in the observational regime. For values $(C, X) = (c, x)$ around the threshold $X = x_0$, we expect this condition to hold as both treatments will be observed with high probability. Here we also require that both treatments are observed with probability greater than zero for values $(C, X) = (c, x)$ away from the threshold $X = x_0$.

Condition 4.9 (Continuity in the observational regime). For $t = 0, 1$, the conditional expectation $\mathbb{E}_f(Y \mid C, X, T = t)$ is continuous in x at x_0 .

Condition 4.9 requires that for given values of the treatment T (active and control), the conditional expectation of Y , given C , X and T is continuous at the threshold x_0 . This condition suggests that it is the treatment and not any of the other variables that is responsible for the discontinuity in the outcome around the threshold. Similar assumptions can be found in (Hahn et al., 2001; Lee, 2008; Geneletti et al., 2015).

Theorem 4.10. Let (C, X) be strongly sufficient covariates. Then for $t = 0, 1$ and any versions of the conditional expectations,

$$\mathbb{E}_t(Y \mid C, X) = \mathbb{E}_f(Y \mid C, X, T = t) \quad (10)$$

almost surely in any regime. In particular,

$$\mathbb{E}_t(Y \mid C, X) = \mathbb{E}_f(Y \mid C, X, T = t) \quad \text{a.s. } [\mathbb{P}_f]. \quad (11)$$

Theorem 4.11. Let (C, X) be strongly sufficient covariates and assume that Condition 4.9 holds. Then,

$$ACE_{x_0} = \mathbb{E}_f[\lim_{\epsilon \downarrow 0} h(C) \mid X = x_0]$$

where

$$h(C) = \mathbb{E}_f(Y \mid C, X = x_0 + \epsilon, T = 1) - \mathbb{E}_f(Y \mid C, X = x_0 - \epsilon, T = 0).$$

Theorem 4.11 provides a formula for identifying the ACE_{x_0} purely from observational data as we have shown that the ACE_{x_0} is expressible purely in terms of properties of the observational joint distribution of (Y, C, X, T) , where (C, X) are strongly sufficient covariates.

To identify the ACE_{x_0} in terms of the threshold indicator Z which is considered a special case of a binary IV (Imbens and Lemieux, 2008; Didelez et al., 2010; Geneletti et al., 2015), we invoke the following IV property for $\mathcal{S} = \{0, 1, f\}$.

Property 4.12 (IV property).

$$Y \perp\!\!\!\perp Z \mid (C, X, T, \Sigma) \quad (12)$$

Property 4.12 requires that the conditional distribution of Y given (C, X, T) in any regime $\sigma \in \mathcal{S} = \{0, 1, f\}$ does not depend on Z . This property readily follows as Z is a function of X and, once X is given, further information about Z becomes redundant for Y .

Lemma 4.13. *Assume that Condition 4.9 and Property 4.12 hold. Then a.s. $[\mathbb{P}_f]$,*

$$\begin{aligned} & \lim_{\epsilon \downarrow 0} [\mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0)] \\ &= \lim_{\epsilon \downarrow 0} \frac{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)}{\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)}. \end{aligned}$$

Lemma 4.13 leads to the following Theorem, in which the average causal effect is identified.

Theorem 4.14. *Let (C, X) be strongly sufficient covariates and assume that Condition 4.9 and Property 4.12 hold. Then,*

$$ACE_{x_0} = \mathbb{E}_f[\lim_{\epsilon \downarrow 0} h(C) | X = x_0]$$

where

$$h(C) = \frac{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)}{\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)}.$$

In the next section, we present a worked example in which the methods described are applied to real example concerning the prescription of statins in UK primary care.

5 Example: Prescription of Statins in UK Primary Care

We describe an example in which the methods outlined in Sections 2–4 are applied to a set of real data. We use The Health Improvement Network (THIN) database, a large source of UK primary care data, consisting of routine, anonymised, patient records collected during patient consultations at over 500 general practices (GPs) (www.epic-uk.org). We wish to examine the causal relationship between the initiation of statin therapy and the low density lipoprotein (LDL) cholesterol level. Many randomised trials and other studies have established the effect of statins on LDL cholesterol level (Scandinavian Simvastatin Survival Study Group (1994), Heart Protection Study Collaborative Group (2002), Baigent et al. (2005), Brugts et al. (2009)) which makes this choice of example useful for demonstrative purposes.

In January 2006, the UK National Institute of Health and Care Excellence (NICE) issued guidelines stating that individuals whose 10-year risk of cardiovascular disease (CVD) development was greater than 20% should be prescribed statins, a class of cholesterol-lowering drugs (NICE (2008)). Here, 10-year CVD risk is measured using an appropriate risk predictor, such as the 10-year Framingham risk score (Wilson et al. (1998)) or Q-RISK score (Hippisley-Cox et al. (2007)). We consider an individual’s 10-year CVD risk score as the ‘assignment variable’ for statin therapy.

We chose to use data from 1386 male patients in THIN, for whom a 10-year CVD risk was calculated between 01 January 2007 and 31 December 2008, aged 50–70 who were non-diabetic, non-smokers, had not previously experienced a cardiovascular event (e.g. stroke, myocardial infarction) and who had not previously been prescribed statins. We define three key variables of interest in this example, where $i \in \{1, \dots, 1386\}$ denotes the patient index:

- $X_i \in [0, 1]$: Risk score of the i^{th} patient (the **assignment variable**).
- $Y_i \in [0, \infty)$: LDL cholesterol level of the i^{th} patient (the **continuous outcome variable**).
- $T_i \in \{0, 1\}$: Treatment indicator for the i^{th} patient (= 0 if patient is not prescribed statins, = 1 if patient is prescribed statins).

We note that Y_i is measured between one and six months after the risk score calculation, so that the effect of statins on LDL cholesterol level can be observed. Furthermore, we define the threshold attainment indicator Z_i ($Z_i = 1$ if risk score is at or above the threshold and $Z_i = 0$ otherwise).

Figure 4 shows a simple scatter plot of the assignment variable (10-year CVD risk score) against the outcome variable (LDL cholesterol level), with different icons to differentiate between patients who did and did not receive statins. We see that, as would be expected, the RDD is fuzzy in this example. Nonetheless, there appears to be an obvious separation in the LDL cholesterol levels of the treated and

the untreated at the treatment threshold suggesting, at first sight, that the use of a fuzzy RDD may be appropriate for these data.

We consider the possible confounders. Since 10-year risk score is calculated based on: age, gender, smoking status, diabetic status, systolic and diastolic blood pressures and total cholesterol level, it is perhaps reasonable to consider these variables as the main confounders in this example. The data have been stratified by choosing only males aged 50–70 who are non-diabetic and non-smokers. We assume that other confounders are similarly distributed for patients whose risk scores lie close enough to the treatment threshold. The notion of patients ‘lying close to the treatment threshold’ is reflected in the choice of RDD bandwidth, which we denote ζ (with $\zeta > 0$). Patients’ data are considered in the RDD if their assignment variable value lies within the interval $(x_0 - \zeta, x_0 + \zeta)$, representing the range of assignment variable values for which patients are considered to be exchangeable in the RDD.

For most patients, we would expect this assumption of confounders being similarly distributed above and below the threshold to be reasonable, especially after stratification and where the chosen bandwidth is reasonably small, although the assumption is untestable. However, as is often seen in randomised trials, it can be useful to compare basic summary statistics of possible confounders between treatment groups. Summary statistics for confounders compared between groups above and below the threshold are shown in Table 1, for $\zeta \in \{0.05, 0.10, 0.15\}$. These show that, in general, confounders appear to be similarly distributed in groups above and below the threshold.

The THIN data have been generated under an observational regime. In accordance with the ideas presented in Section 4, we must examine whether or not we can identify the average causal effect of statins on the LDL cholesterol level using these data. Condition (4.1a) implies that the distributions of the 10-year risk score and the confounding variables should be independent of the regime under which the data were obtained. Whilst this condition is untestable, pragmatically, it appears reasonable that the way in which treatment arises is unlikely to have affected the distributions of confounders and the 10-year CVD risk score, especially in patients whose 10-year CVD risk scores lie close to the threshold. For Condition (4.1b) to be satisfied, we consider the further assumption that the LDL cholesterol level is independent of the regime, conditional on the confounders, 10-year CVD risk score and treatment allocation. Again, although this is untestable, it seems reasonable to assume there may be correlation between the confounders, the risk score and LDL cholesterol level and, if statins affect LDL cholesterol level, correlation between LDL cholesterol level and treatment allocation via some biological mechanism through which statins work on human subjects. Once these relationships have been accounted for, we assume that the mechanism through which the data concerning treatment (or non-treatment) arose (i.e. interventional or observational) is of limited importance. Consequently, we argue that Condition (4.1b) is likely to be satisfied in this example.

Condition 4.8 states that the both the probability of treatment and the probability of non-treatment should be non-zero, almost surely, under the fuzzy RDD, conditional on the assignment variable and the set of confounders. In this example, Figure 4 and the results in Table 1 show that the probability of treatment is non-zero under many different levels of the possible confounding variables and the 10-year risk score. Hence, we argue that the confounders and the 10-year risk score are *strongly sufficient* covariates.

Condition 4.9 states that the expected value of the outcome variable should be continuous in the assignment variable, x , at the threshold conditional on the confounders, 10-year risk score and where treatment does not change. In this example, it is unlikely that the expected LDL cholesterol value is likely to ‘jump’ where treatment assignment does not change.

Finally, we need to argue that Property 4.12 holds with these data. In essence, Property 4.12 implies that the outcome variable value is independent of the threshold variable conditional on the assignment variable, treatment, regime indicator and confounders. In this example, the treatment rule (prescription of statins if the 10-year CVD risk score is 20% or greater) is determined by the UK government and is not patient-specific. Furthermore, since Z is determined solely by the 10-year CVD risk score, conditioning on this variable implies directly that further knowledge of Z is unnecessary to make inference regarding the LDL cholesterol level. Hence, we apply the result of Theorem 4.14 to estimate the causal effect of statin therapy on LDL cholesterol level, using a fuzzy RDD and the THIN subset of data.

5.1 Models and Estimation

For a chosen bandwidth ζ ($\zeta \in (0, 1)$), we define \mathcal{A}_ζ and \mathcal{B}_ζ to be the sets of patients whose 10-year CVD risk scores lie above and below the threshold, respectively. Setting the treatment threshold to be $x_0 = 0.2$ and defining $X_i^C = X_i - 0.2$ to be the centred 10-year CVD risk score for the i^{th} patient, we consider the following linear models:

$$\begin{aligned} Y_i &= \beta_{0a} + \beta_{1a}X_i^C + \omega_i \quad (\text{for } i \in \mathcal{A}_\zeta); \\ Y_i &= \beta_{0b} + \beta_{1b}X_i^C + \omega_i \quad (\text{for } i \in \mathcal{B}_\zeta). \end{aligned}$$

with $\omega_i \sim \mathcal{N}(0, \sigma^2)$ for $i \in \mathcal{A}_\zeta \cup \mathcal{B}_\zeta$. The function that we wish to evaluate is

$$h(C) = \frac{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)}{\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)}. \quad (13)$$

The numerator of (13) is estimated by $\hat{\beta}_{0a} - \hat{\beta}_{0b}$. To estimate the denominator of (13) we consider the following estimates:

$$\begin{aligned} \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) &\approx \hat{\pi}_a = \frac{1}{|\mathcal{A}_\zeta|} \sum_{i \in \mathcal{A}_\zeta} T_i \\ \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0) &\approx \hat{\pi}_b = \frac{1}{|\mathcal{B}_\zeta|} \sum_{i \in \mathcal{B}_\zeta} T_i \end{aligned}$$

Hence, the causal effect estimate for the effect of statins on LDL cholesterol level is given by

$$\hat{\eta} = \frac{\hat{\beta}_{0a} - \hat{\beta}_{0b}}{\hat{\pi}_a - \hat{\pi}_b}$$

where $(\hat{\beta}_{0a}, \hat{\beta}_{0b})$ denote the maximum likelihood estimates of (β_{0a}, β_{0b}) . We consider the calculation of η using the THIN subset data and a choice of three RDD bandwidths ($\zeta = 0.05, 0.1$ and 0.15). Estimated values, $\hat{\eta}$, together with associated 95% confidence intervals, are shown in Table 2.

Examining Table 2, we see that for each chosen RDD bandwidth, the estimated causal effect of statins on LDL cholesterol level appears to differ from zero, suggesting that statins appear to reduce the LDL cholesterol level in general. This would be expected, especially in light of the many large-scale randomised trials that have provided substantial evidence in favour of the beneficial effect of statins with respect to LDL cholesterol level. We have used the arguments and results presented in earlier sections of the paper to justify and construct a regression discontinuity design on real observational data, through a decision theoretic approach. We note that the use of a restricted sample of patients for this demonstrative example implies that the results reported are not necessarily applicable to the general population, or of strict clinical/epidemiological significance.

6 Discussion

We have developed a formal, decision theoretic approach to the regression discontinuity design using the language of conditional independence. Conditions under which the causal effect of a treatment on a continuous outcome of interest can be identified were defined and the corresponding causal effect estimator derived, for both sharp and fuzzy RDDs. Full theoretical justification for causal effect identification was made.

In line with other methods for causal inference in observational studies, assumptions involving confounders are necessary, though we argue that, where such assumptions are reasonable, carefully-argued and example-specific, estimation of a causal effect can be made in either a sharp or fuzzy RDD. Specification of a full and appropriate set of confounders remains important, as with other causal inference methods. With an RDD, the comparison of groups above and below the treatment threshold, but whose

assignment variable values lie close to the treatment threshold, aims to ensure that the distributions of confounders, both observed and unobserved, are balanced between comparative groups. This assumption highlights the link between the RDD in an observational setting and the randomised controlled trial in an experimental setting. Although this assumption is untestable, it may seem reasonable in many scenarios and simple methods, such as basic plots and summary statistics, can be helpful in highlighting fundamental distributional discrepancies in confounders between groups above and below the treatment threshold.

We applied the methodology to an example based on real data concerning the prescription of statins to patients at moderate risk of cardiovascular disease development, according to a 10-year CVD risk score. In this example, the main observed confounders were specified and values compared between groups above and below the treatment threshold. A basic scatter plot of the assignment variable against the outcome of interest (Figure 4) suggested that the use of a fuzzy RDD was likely to be appropriate for these data. It is important to note the value of plots such as Figure 4, which should always be produced whenever the use of an RDD is considered for any observational dataset. Justification of each assumption required in our decision theoretic approach was made and the causal effect of statins on LDL cholesterol level was calculated using a selection of pre-specified bandwidths. This example showcases the use of a rigorous approach to RDD analyses in medical studies and epidemiology.

Further methodological extensions to our work include the development of RDD methods for non-continuous outcomes and dynamic processes. We encourage the wider use of the RDD as a valid method for treatment effect estimation in biostatistics and epidemiology. We hope that the use of RDD methods will become more widespread in this field, particularly with the increasing availability of large-scale electronic healthcare databases.

Acknowledgements

We thank Vanessa Didelez for helpful and insightful discussions concerning this work. Panayiota Constantinou received funding from the UK Engineering and Physical Sciences Research Council (EPSRC grant: D063485/1). Aidan O’Keeffe received funding from the UK Medical Research Council (MRC grant: MR/K014838/1). Approval for the use of the THIN data was granted by a Scientific Review Committee in August 2014 (ref. 14-021).

References

- Almond, D. and Doyle, J. J. (2011). After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy*, 3:1–34.
- Baigent, C., Keech, A., Kearney, P. M., Blackwell, L., Pollicino, G. B. C., Kirby, A., Sourjina, T., Peto, R., Collins, R., Simes, R., and Collaborators, C. T. T. (2005). Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet*, 366(9493):1267–1278.
- Berk, R. and Leeuw, J. (1999). An evaluation of California’s inmate classification system using a generalised regression discontinuity design. *Journal of the American Statistical Association*, 94(448):1045–1052.
- Bor, J., Moscoe, E., Mutevezdi, P., Newell, M. L., and Barnighausen, T. (2014). Regression discontinuity designs in epidemiology: Causal inference without randomised trials. *Epidemiology*, 25(5):729–737.
- Brugts, J. J., Yetgin, T., Hoeks, S. E., Gotto, A. M., Shepherd, J., Westendorp, R. G. J., de Craen, A. J. M., Knopp, R. H., Nakamura, H., Ridker, P., van Domburg, R., and Deckers, J. W. (2009). The benefits of statins in people without established cardiovascular disease but with cardiovascular risk factors: meta-analysis of randomised controlled trials. *BMJ*, 338.
- Constantinou, P. (2014). *Conditional Independence and Applications in Statistical Causality*. PhD thesis, University of Cambridge.

- Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. (2007). *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. Springer Publishing Company, Incorporated, 1st edition.
- Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 41(1):1–31.
- Dawid, A. P. (1980). Conditional independence for statistical operations. *Annals of Statistics*, 8(3):598–617.
- Dawid, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424.
- Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review / Revue Internationale de Statistique*, 70(2):pp. 161–189.
- Dawid, A. P. and Constantinou, P. (2014). A formal treatment of sequential ignorability. *Statistics in Biosciences*, 6(2):166–188.
- Dawid, A. P. and Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Statist. Surv.*, 4:184–231.
- Didelez, V., Meng, S., and Sheehan, N. A. (2010). Assumptions of IV methods for observational epidemiology. *Statist. Sci.*, 25(1):22–40.
- Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):199–215.
- Geneletti, S., O’Keeffe, A. G., Sharples, L. D., Richardson, S., and Baio, G. (2015). Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, 34:2334–2352.
- Guo, H. and Dawid, A. P. (2010). Sufficient covariates and linear propensity analysis. In Teh, Y. W. and Titterton, D. M., editors, *In Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics (AISTATS) 2010*, Journal of Machine Learning Research Workshop and Conference Proceedings 9, pages 598–617, Chia Laguna, Sardinia, Italy.
- Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.
- Heart Protection Study Collaborative Group (2002). MRC/BHF Heart Protection Study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo controlled trial. *Lancet*, 360(9326):7–22.
- Heckerman, D. and Shachter, R. (1995). Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research*, 3:405–430.
- Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., and Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*, 335(7611):136.
- Imbens, G. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2):615–635.
- Lalive, R. (2008). How do extended benefits affect unemployment duration: A regression discontinuity approach. *Journal of Econometrics*, 142(2):785–806.
- Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks*, 20(5):491–505.
- Lee, D. S. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics*, 142(2):675–697.

- Linden, A. and Adams, J. L. (2012). Combining the regression discontinuity design and propensity score-based weighting to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, 18(2):317–325.
- Linden, A., Adams, J. L., and Roberts, N. (2006). Evaluating disease management programme effectiveness: an introduction to the regression discontinuity design. *Journal of Evaluation of Clinical Practice*, 12(2):124–131.
- NICE (2008). *Quick reference guide: Statins for the prevention of cardiovascular events*.
- O’Keeffe, A. G., Geneletti, S., Baio, G., Sharples, L. D., Nazareth, I., and Petersen, I. (2014). Regression discontinuity designs: An approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ*, 339:g5293.
- Pearl, J. (1986). A constraintpropagation approach to probabilistic reasoning. In Kanal, L. N. and Lemmer, J. F., editors, *In Uncertainty in Artificial Intelligence*, page 357370, North-Holland, Amsterdam.
- Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 6:34–68.
- Scandinavian Simvastatin Survival Study Group (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet*, 344(8934):1383–1389.
- Thistlethwaite, D. L. and Campbell, D. T. (1960). Regression-discontinuity analysis - an alternative to the ex-post-facto experiment. *Journal of Educational Psychology*, 51(6):309–317.
- van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: Regression discontinuity approach. *International Economic Review*, 43(4):1249–1287.
- van der Klaauw, W. (2008). Regression-discontinuity analysis: a survey of recent developments in economics. *Labour*, 22(2):219–245.
- Vandenbroucke, J. P. and le Cessie, S. (2014). Regression discontinuity design: Let’s give it a try to evaluate medical and public health interventions. *Epidemiology*, 25(5):738–741.
- Verma, T. and Pearl, J. (1990). Causal networks: Semantics and expressiveness. In Shachter, R. D., Levitt, T. S., Kanal, L. N., and Lemmer, J. F., editors, *In Uncertainty in Artificial Intelligence 4*, page 6976, North-Holland, Amsterdam.
- Wilson, P. W. F., D’Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., and Kannel, W. B. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation*, 97:1837–1847.

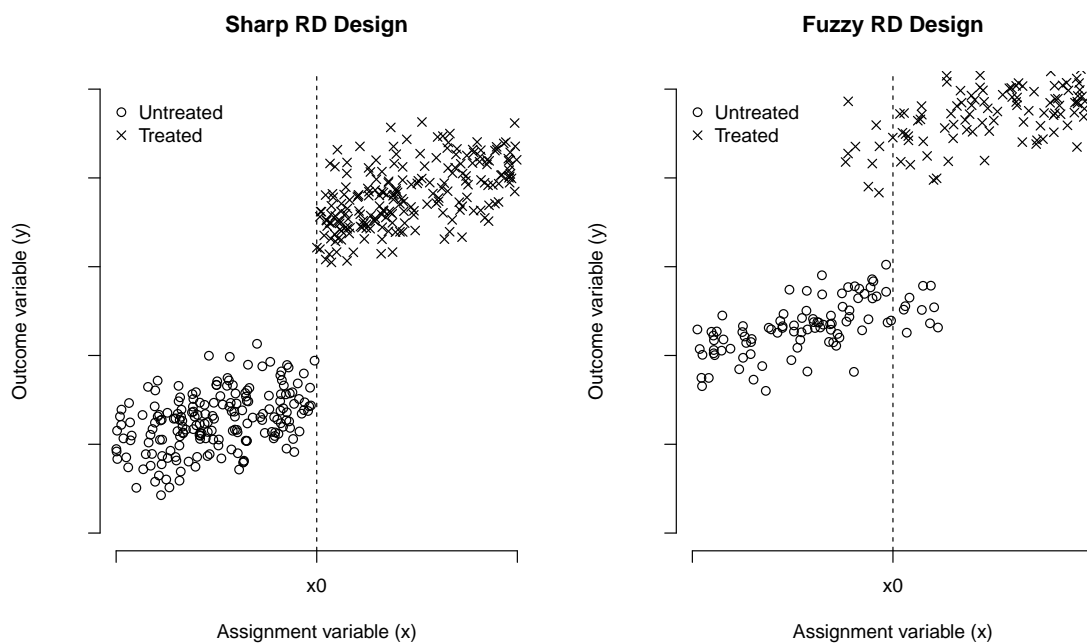


Figure 1: Graphical representations of example sharp (left-hand plot) and fuzzy (right-hand plot) regression discontinuity designs. ‘Untreated’ icons denote subjects who do not receive the treatment whereas ‘Treated’ icons denote subjects who received the treatment. The treatment threshold is depicted by a dashed line at $x = x_0$.

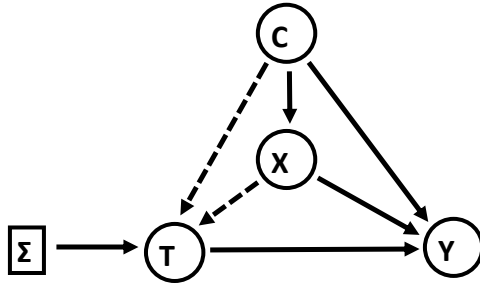


Figure 2: Influence Diagram representing (4a) and (4b). Property (4a) is represented by the absence of arrows from Σ to (C, X) and property (4b) is represented by the absence of a direct arrow from Σ to Y . The dotted arrows from X to T and from C to T indicate a link that disappears under an interventional regime.

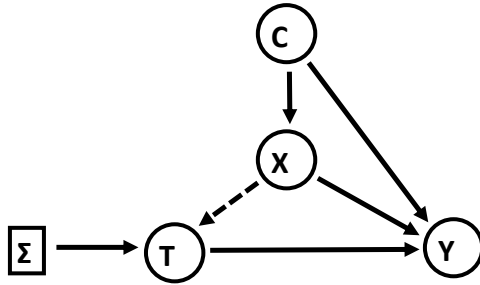


Figure 3: Influence Diagram representing (4a), (4b) and (6). Property (6) is represented by the absence of an arrow from C to T .

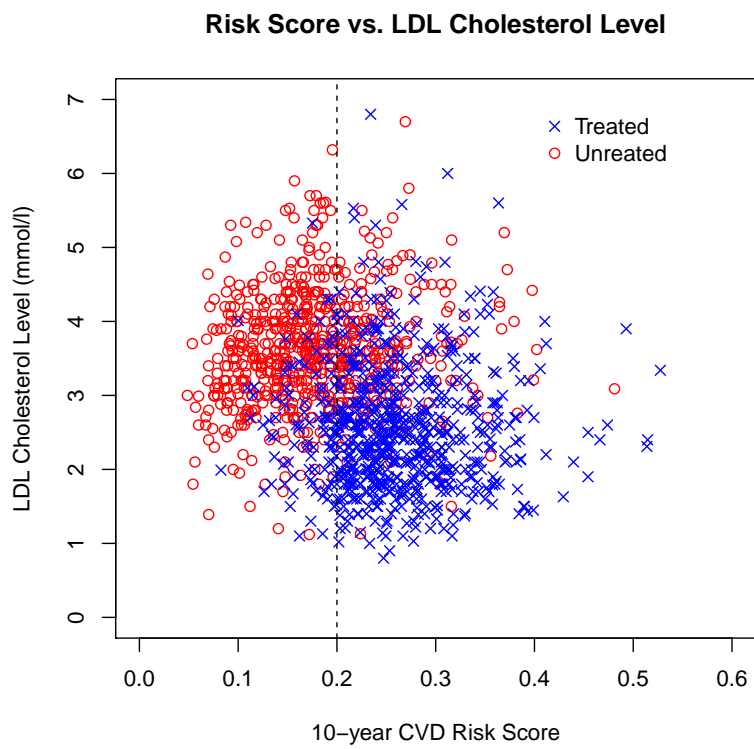


Figure 4: Scatter plot showing 10-year CVD risk score against LDL cholesterol level for the THIN subset data. Patients who receive statins are indicated by blue icons and patients who do not receive statins are indicated by red icons. The dashed vertical line denotes the treatment threshold at the 20% risk score.

Table 1: Table showing summary statistics for a variety of possible confounding variables for patients above and below the threshold, using a variety of RDD bandwidths ($\zeta = 0.05, 0.1$ or 0.15), for the THIN data subset.

<i>Bandwidth = 0.05</i>						
Variable	Group	Mean	Median	Std. Dev.	Minimum	Maximum
Systolic BP (mmHg)	$Z_i = 0$	133.7	134.0	9.9	106.0	162.0
	$Z_i = 1$	139.5	139.0	11.7	110.0	184.0
Diastolic BP (mmHg)	$Z_i = 0$	81.6	80.0	8.3	59.0	110.0
	$Z_i = 1$	82.0	80.0	8.8	58.0	110.0
LDL Cholesterol (mmol/L)	$Z_i = 0$	3.96	3.93	0.82	1.50	6.20
	$Z_i = 1$	3.93	3.90	0.77	1.24	6.20
HDL Cholesterol (mmol/L)	$Z_i = 0$	1.34	1.30	0.31	0.73	2.85
	$Z_i = 1$	1.25	1.20	0.27	0.76	2.28
Triglycerides (mmol/L)	$Z_i = 0$	1.68	1.60	0.78	0.38	5.10
	$Z_i = 1$	1.78	1.60	0.91	0.55	6.78
<i>Bandwidth = 0.10</i>						
Variable	Group	Mean	Median	Std. Dev.	Minimum	Maximum
Systolic BP (mmHg)	$Z_i = 0$	131.7	132.0	10.6	102.0	176.0
	$Z_i = 1$	141.2	140.0	12.4	110.0	199.0
Diastolic BP (mmHg)	$Z_i = 0$	81.1	80.0	8.1	59.0	110.0
	$Z_i = 1$	82.5	81.0	9.1	58.0	110.0
LDL Cholesterol (mmol/L)	$Z_i = 0$	3.82	3.76	0.82	0.90	6.20
	$Z_i = 1$	3.99	3.90	0.80	1.24	6.90
HDL Cholesterol (mmol/L)	$Z_i = 0$	1.37	1.30	0.32	0.73	2.85
	$Z_i = 1$	1.24	1.20	0.26	0.60	2.28
Triglycerides (mmol/L)	$Z_i = 0$	1.57	1.40	0.74	0.38	5.10
	$Z_i = 1$	1.80	1.60	0.91	0.55	9.00
<i>Bandwidth = 0.15</i>						
Variable	Group	Mean	Median	Std. Dev.	Minimum	Maximum
Systolic BP (mmHg)	$Z_i = 0$	130.3	130.0	11.2	97.0	176.0
	$Z_i = 1$	142.5	140.0	12.9	110.0	202.0
Diastolic BP (mmHg)	$Z_i = 0$	80.5	80.0	8.2	59.0	110.0
	$Z_i = 1$	83.0	82.0	9.2	58.0	119.0
LDL Cholesterol (mmol/L)	$Z_i = 0$	3.75	3.70	0.83	0.90	6.20
	$Z_i = 1$	4.00	3.90	0.78	1.24	6.90
HDL Cholesterol (mmol/L)	$Z_i = 0$	1.39	1.33	0.33	0.73	2.85
	$Z_i = 1$	1.21	1.20	0.26	0.60	2.28
Triglycerides (mmol/L)	$Z_i = 0$	1.52	1.36	0.73	0.38	5.10
	$Z_i = 1$	1.85	1.62	0.92	0.55	9.00

Table 2: Table showing the estimated causal effect of statin therapy on LDL cholesterol level, together with associated standard errors and 95% confidence intervals using the THIN data subset for chosen bandwidths: 0.05, 0.10 and 0.15.

Bandwidth (ζ)	Estimate $\hat{\eta}$ (Standard Error)	95% Confidence Interval
0.05	-0.766 (0.373)	(-1.497, -0.035)
0.10	-0.889 (0.209)	(-1.299, -0.479)
0.15	-0.934 (0.165)	(-1.258, -0.610)

A Proofs of Theorems in Section 4

A.1 Proof of Theorem 4.4

Theorem 4.4. *Let (C, X) be sufficient covariates and assume that Condition 4.3 holds. Then,*

$$Y \perp\!\!\!\perp \Sigma \mid (X, T). \quad (14)$$

Proof. Applying the axioms of Theorem 3.3, we obtain:

$$\begin{aligned} (4a) \quad & \xrightarrow{P1} \Sigma \perp\!\!\!\perp (C, X) \\ & \xrightarrow{P4} \Sigma \perp\!\!\!\perp (C, X) \mid X \\ & \xrightarrow{P3} \Sigma \perp\!\!\!\perp C \mid X \\ & \xrightarrow{P1} C \perp\!\!\!\perp \Sigma \mid X. \end{aligned} \quad (15)$$

Also,

$$(6) \quad \xrightarrow{P1} C \perp\!\!\!\perp T \mid (X, \Sigma), \quad (16)$$

and applying $P5$ to (15) and (16), we see that $C \perp\!\!\!\perp (T, \Sigma) \mid X$. Then applying in turn $P4$, $P3$ and $P1$ we obtain

$$\Sigma \perp\!\!\!\perp C \mid (X, T). \quad (17)$$

Lastly,

$$(4b) \quad \xrightarrow{P1} \Sigma \perp\!\!\!\perp Y \mid (C, X, T) \quad (18)$$

and applying $P5$ to (17) and (18) we see that $\Sigma \perp\!\!\!\perp (C, Y) \mid (X, T)$. Applying in turn $P3$ and $P1$ concludes the proof. \square

We note that the result of Theorem 4.4, can also be verified using graphical techniques (in particular, the d-separation (Pearl, 1986; Verma and Pearl, 1990), or the equivalent moralization criterion (Lauritzen et al., 1990)).

A.2 Proof of Theorem 4.6

Theorem 4.6 *Let (C, X) be sufficient covariates and assume that Condition 4.3 and Condition 4.5 hold. Then,*

$$ACE_{x_0} = \lim_{\epsilon \downarrow 0} [\mathbb{E}_s(Y \mid X = x_0 + \epsilon, T = 1) - \mathbb{E}_s(Y \mid X = x_0 - \epsilon, T = 0)].$$

Proof.

$$\begin{aligned} ACE_{x_0} &:= \mathbb{E}_1(Y \mid X = x_0) - \mathbb{E}_0(Y \mid X = x_0) \\ &= \lim_{\epsilon \downarrow 0} [\mathbb{E}_1(Y \mid X = x_0 + \epsilon) - \mathbb{E}_0(Y \mid X = x_0 - \epsilon)] \quad (\text{Condition 4.5}) \\ &= \lim_{\epsilon \downarrow 0} [\mathbb{E}_1(Y \mid X = x_0 + \epsilon, T = 1) - \mathbb{E}_0(Y \mid X = x_0 - \epsilon, T = 0)] \\ &= \lim_{\epsilon \downarrow 0} [\mathbb{E}_s(Y \mid X = x_0 + \epsilon, T = 1) - \mathbb{E}_s(Y \mid X = x_0 - \epsilon, T = 0)] \quad (\text{Theorem 4.4}) \end{aligned}$$

\square

A.3 Proof of Theorem 4.7

Theorem 4.7 *Let (C, X) be sufficient covariates and assume that Condition 4.3 and Condition 4.5 hold. Then,*

$$ACE_{x_0} = \lim_{\epsilon \downarrow 0} [\mathbb{E}_s(Y | X = x_0 + \epsilon, Z = 1) - \mathbb{E}_s(Y | X = x_0 - \epsilon, Z = 0)].$$

Proof. Noting that for the sharp observational regime, $\mathbb{P}_s(T = Z) = 1$, the proof follows directly from Theorem 4.6. \square

A.4 Proof of Theorem 4.10

Theorem 4.10 *Let (C, X) be strongly sufficient covariates. Then for $t = 0, 1$ and any versions of the conditional expectations,*

$$\mathbb{E}_t(Y | C, X) = \mathbb{E}_f(Y | C, X, T = t) \quad (19)$$

almost surely in any regime. In particular,

$$\mathbb{E}_t(Y | C, X) = \mathbb{E}_f(Y | C, X, T = t) \quad \text{a.s. } [\mathbb{P}_f]. \quad (20)$$

Proof. See Theorem 1 in Guo and Dawid (2010). \square

A.5 Proof of Theorem 4.11

Theorem 4.11 *Let (C, X) be strongly sufficient covariates and assume that Condition 4.9 holds. Then,*

$$ACE_{x_0} = \mathbb{E}_f[\lim_{\epsilon \downarrow 0} h(C) | X = x_0]$$

where

$$h(C) = \mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0).$$

Proof.

$$\begin{aligned} ACE_{x_0} &:= \mathbb{E}_1(Y | X = x_0) - \mathbb{E}_0(Y | X = x_0) \\ &= \mathbb{E}_1[\mathbb{E}_1(Y | C, X = x_0) | X = x_0] - \mathbb{E}_0[\mathbb{E}_0(Y | C, X = x_0) | X = x_0] \\ &= \mathbb{E}_f[\mathbb{E}_1(Y | C, X = x_0) | X = x_0] - \mathbb{E}_f[\mathbb{E}_0(Y | C, X = x_0) | X = x_0] \quad (4a) \\ &= \mathbb{E}_f[\mathbb{E}_f(Y | C, X = x_0, T = 1) - \mathbb{E}_f(Y | C, X = x_0, T = 0) | X = x_0] \\ &\quad (\text{Theorem 4.10}) \\ &= \mathbb{E}_f[\lim_{\epsilon \downarrow 0} h(C) | X = x_0] \quad (\text{Condition 4.9}) \end{aligned}$$

\square

A.6 Proof of Lemma 4.13

Lemma 4.13 *Assume that Condition 4.9 and Property 4.12 hold. Then a.s. $[\mathbb{P}_f]$,*

$$\begin{aligned} &\lim_{\epsilon \downarrow 0} [\mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0)] \\ &= \lim_{\epsilon \downarrow 0} \frac{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)}{\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)}. \end{aligned}$$

Proof.

$$\begin{aligned}
& \lim_{\epsilon \downarrow 0} [\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)] \\
&= \lim_{\epsilon \downarrow 0} [\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1, T = 0) \times \mathbb{P}_f(T = 0 | C, X = x_0 + \epsilon, Z = 1) \\
&\quad + \mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0, T = 0) \times \mathbb{P}_f(T = 0 | C, X = x_0 - \epsilon, Z = 0) \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)] \\
&= \lim_{\epsilon \downarrow 0} \{E_f(Y | C, X = x_0 + \epsilon, Z = 1, T = 0) \times [1 - \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1)] \\
&\quad + \mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0, T = 0) \times [1 - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)] \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)\} \\
&= \lim_{\epsilon \downarrow 0} \{E_f(Y | C, X = x_0 + \epsilon, T = 0) \times [1 - \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1)] \\
&\quad + \mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0) \times [1 - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)] \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 1) \times \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)\} \quad (\text{from (16)}) \\
&= \lim_{\epsilon \downarrow 0} \{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) \\
&\quad \times [P_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)] \\
&\quad - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0) \\
&\quad \times [P_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)]\} \quad (\text{Condition 4.9}) \\
&= \lim_{\epsilon \downarrow 0} \{[\mathbb{E}_f(Y | C, X = x_0 + \epsilon, T = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, T = 0)] \\
&\quad \times [\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)]\}
\end{aligned}$$

where all the equalities hold a.s. $[P_f]$. Rearranging we conclude the proof. \square

A.7 Proof of Theorem 4.14

Theorem 4.14 *Let (C, X) be strongly sufficient covariates and assume that Condition 4.9 and Property 4.12 hold. Then,*

$$ACE_{x_0} = \mathbb{E}_f[\lim_{\epsilon \downarrow 0} h(C) | X = x_0]$$

where

$$h(C) = \frac{\mathbb{E}_f(Y | C, X = x_0 + \epsilon, Z = 1) - \mathbb{E}_f(Y | C, X = x_0 - \epsilon, Z = 0)}{\mathbb{P}_f(T = 1 | C, X = x_0 + \epsilon, Z = 1) - \mathbb{P}_f(T = 1 | C, X = x_0 - \epsilon, Z = 0)}.$$

Proof. Applying the result of Lemma 4.13 in Theorem 4.11 we conclude the proof. \square