Approaches to risk ratio estimation in a regression discontinuity design: Application to the prescription of statins for cholesterol reduction in UK Primary Care

Mariam O. Adeleke[1], Aidan G. O'Keeffe*[2,3], and Gianluca Baio[1]

[1]*Department of Statistical Science, University College London, WC1E 6BT, UK*
[2]*School of Mathematical Sciences, University of Nottingham, NG7 2RD, UK*
[3]*Institute of Epidemiology and Health Care, University College London, WC1E 6BT, UK*

Abstract

In recent years regression discontinuity (RD) designs have been used increasingly for the estimation of treatment effects in observational medical data where a rule-based decision to apply a treatment is taken using a continuous assignment variable. Most RD design applications have focused on effect estimation where the outcome of interest is continuous, with scenarios with binary outcomes receiving less attention, despite their ubiquity in medical studies. In this work we develop an approach to estimation of the risk ratio in a fuzzy RD design (where treatment is not always strictly applied according to the decision rule), derived using common RD design assumptions. This method compares favourably to other risk ratio estimation approaches: the established Wald estimator and a risk ratio estimate from a multiplicative structural mean model, with promising results from extensive simulation studies. A demonstration and further comparison is made using a real exam-

*E-mail address: aidan.o'keeffe@nottingham.ac.uk; Corresponding author

ple to evaluate the effect of statins (where a statin prescription is made based on a patient's 10-year cardiovascular disease risk score) on LDL cholesterol reduction in UK Primary Care.

1 Introduction

Regression discontinuity (RD) designs are an established approach to treatment effect estimation using observational data where treatment allocation is made according to a decision rule defined by an external variable [1, 2, 3]. Typically, given a sample of subjects, an 'assignment variable' is observed for each subject and a decision rule used to assign a treatment where a subject should receive the treatment if their assignment variable lies above (or below) an externally defined threshold. Consequently, an RD design may be defined using a local randomisation approach where the treatment threshold separates the 'treated' and 'untreated' and, under certain assumptions, this threshold may be viewed as a quasi-randomisation device in that subjects whose assignment variable value lies 'just below' the threshold may be viewed as similar to those whose assignment variable value lies 'just above' the threshold [4]. In particular, we might view these groups as balanced with regard to confounding variables in an analogous manner to the 'treatment' and 'control' groups in a two-group individually randomised controlled trial.

In recent years, the local randomisation approach to RD designs has been explored, developed and applied in the literature [5, 6, 7] and the differences between this approach to RD designs and the continuity-based RD design framework have been noted (see, for example, [8]). In medical applications, the natural analogy between the quasi-randomisation seen in the local randomisation approach to RD designs and randomisation in an individually randomised controlled trial suggests that a local randomisation RD design can be a practical and intuitive approach to treatment effect estimation in situations where an external decision rule is used to determine

treatment.

This approach to RD designs has been applied to medical studies in the literature, although the usual focus has been on situations where the treatment effect of interest is a difference in means for a continuously distributed outcome of interest [9, 10, 11, 12]. Conversely, binary outcomes have received less attention in RD designs applied to medical studies, with only a few recent examples in the literature of local randomisation RD designs in medical contexts. Lesik [13] used a two-stage least squares approach for odds ratio estimation in an RD design which is only appropriate when the event of interest is rare [14]. Xu [15] and van Leeuwen et al. [16] each considered inference for binary outcomes in sharp RD designs only (where the threshold-treatment rule is applied strictly). Bor et al. [17] applied a complier average causal effect estimate using local linear regression models. Using similarities between local randomisation RD designs and treatment effect estimation using instrumental variables, Geneletti et al. [18] argued that the risk ratio can be estimated using an RD design via a Bayesian approach to a multiplicative structural mean model. Prior constraints were made to ensure that effect estimates and associated uncertainty intervals remained positive but methods were not applicable in a non-Bayesian setting.

In this paper an estimator for the risk ratio in a local randomisation RD design is proposed and derived using design assumptions and compared to alternative approaches. The estimator exhibits favourable properties when evaluated through extensive simulation studies, being suitable for estimating a range of risk ratio sizes in the presence of confounding using only standard RD design assumptions and without reliance on the event of interest being rare. A real example is presented to demonstrate its use, applied to the prescription of statins for the reduction of LDL cholesterol based on 10-year cardiovascular risk score in UK primary care.

This paper is organised as follows: Section 2 outlines the RD design and key assump-

tions required for inference. Section 3 describes methods for risk ratio estimation in an RD design, including the new approach presented in this work. Simulation studies to compare this approach to alternatives are performed and evaluated in Section 4. An example on the prescription of statins in UK primary care based on cardiovascular disease risk score is presented in Section 5 and an overall discussion is provided in Section 6.

2 RD Design - Background and Definitions

Suppose that data are observed for n subjects where, for the i^{th} subject, $Y_i \in \{0, 1\}$ is a binary outcome of interest and X_i is a continuously distributed 'assignment variable'. An externally-defined decision rule is set such that the i^{th} subject should receive a given treatment or intervention if X_i exceeds a pre-defined threshold x_0. The 'threshold indicator' Z_i is defined to indicate this decision where $Z_i = 1$ if $X_i \geq x_0$ and $Z_i = 0$ if $X_i < x_0$.

In addition, a treatment indicator T_i is defined where $T_i = 1$ if the i^{th} subject receives the treatment and $T_i = 0$ otherwise. In an ideal situation, known as a *sharp* RD design, the decision rule would be applied strictly so that $T_i \equiv Z_i$ for each subject. However, in reality – and especially in medical contexts – the decision rule may not be adhered to universally. This is known as a *fuzzy* RD design and, despite this lack of adherence, treatment effects can be estimated in fuzzy RD designs subject to some assumptions which will be outlined. In this work, we focus on treatment effect estimation for binary outcomes in fuzzy RD designs because fuzziness is common in medical applications, although methods discussed can be applied to sharp RD designs. Potential subject-specific confounding variables are denoted by $\mathcal{C}_i$ and we note that these may be observed or unobserved.

As a running example that will be explored in this paper, we consider the prescription of statins – a class of cholesterol-lowering drugs – based on 10-year cardiovascular

disease risk score (10-year CVD risk). In the UK, a 2008 guideline from the National Institute for Health and Care Excellence (NICE) stipulated that statins should be prescribed routinely to adults if their 10-year CVD risk score, calculated using the Framingham method [19] or QRISK approach [20], exceeds 20%. Risk scores are calculated by inputting patient variables such as sex, age, smoking status and blood pressure measurements into a validated risk prediction model. It is widely known that increased low density lipoprotein (LDL) cholesterol can contribute to cardiovascular disease and statins are prescribed with the aim of reducing LDL cholesterol to prevent future CVD (e.g. a stroke or myocardial infarction). An LDL cholesterol reduction of 1mmol/L or more has been described as clinically beneficial for those who may be at risk of CVD [21]. Hence, we will present methods with reference to an RD design to assess the effect of statins on obtaining a significant (1mmol/L or greater) reduction in LDL cholesterol – a binary outcome of clinical interest.

In this example, the assignment variable is a patient's 10-year CVD risk score, the threshold is a score of 0.2 (20%), the binary outcome is a reduction of 1mmol/L or more in LDL cholesterol (1 = 'yes', 0 = 'no') and the treatment is the prescription of statins. Data are sourced for 1384 males who were non smokers and not diabetic from The Health Improvement Network (THIN) - a large source of anonymised UK primary care data from over 500 general practices (family doctors) and broadly representative of the UK population at the time of extraction (2008–2014) [22, 23].

To place this example in the context of a local randomisation RD design, we might assume that a group of patients whose 10-year CVD risk scores lie 'just above' the 20% threshold (and therefore who should receive statins) might be considered similar to a group of patients whose 10-year CVD risk scores lie 'just below' the 20% threshold (and therefore who should not receive statins). In using the term 'similar' we mean that these groups may be balanced with respect to other variables – potentially confounding variables (e.g. age, body mass index, blood pressure

level etc.) – in a similar manner to groups in a two-group individually randomised controlled trial. The concept of 'just above' and 'just below' the threshold is quantified by defining a window $(x_0 - h, x_0 + h)$ for some positive constant h where a subject is included in the RD analysis only if their observed assignment variable value x_i lies in the range $(x_0 - h, x_0 + h)$. In a continuity-based RD design, h – often known as the 'bandwidth' – may be selected according to desired optimality criteria to provide an unbiased treatment effect estimate at the threshold, typically using a local polynomial fit [24, 25]. In a local randomisation framework, which we use in this work, the bandwidth h is typically chosen a priori to define a region in which balance with respect to unobserved confounders is likely to hold. This may be obtained by clinical input on discussion with experts who may advise on likely patient characteristics. Alternatively, a number of bandwidths could be considered and, for each bandwidth, distributions of potential confounding variables compared for patients with scores above and below the threshold to obtain an appropriate bandwidth prior to using the RD design [5]. More recently, a data-driven Bayesian approach to optimal bandwidth selection has been proposed for use in the local randomisation framework [26].

As mentioned previously, with medical scenarios it is usual that the treatment decision rule will not be followed strictly for all subjects. In the statins example it is possible that some general practitioners (GPs) may prescribe statins to those with 10-year risk scores below 20% because - knowing a patient personally - they feel that treatment would be beneficial. In contrast, other patients may choose to decline treatment and instead reduce their risk of cardiovascular disease development by other means (for example, by changing their lifestyle or diet). That the threshold decision rule is not followed for some subjects does not imply that the RD design cannot be used for reliable inference regarding a treatment effect, although an acknowledgment of the 'fuzziness' in the design should be made and methods developed and used for inference should account for this.

2.1 RD Design Assumptions

A number of assumptions are required to estimate a treatment effect using an RD design [2, 27]. We outline these using the language of conditional independence [28] although we note that the assumptions can also be stated using a potential outcomes framework [2, 29]. Assumptions are stated mathematically but Appendix A provides further details and a description of each assumption in the context of the statin prescription example to aid understanding of these assumptions in practice. We use the following notation: if A, B and C are random variables then $A \perp\!\!\!\perp B$ implies that 'A and B are independent' whereas $A \not\perp\!\!\!\perp B$ implies that 'A and B are not independent'. Similarly $A \perp\!\!\!\perp B|C$ implies that 'A and B are independent conditional on C' and $A \not\perp\!\!\!\perp B|C$ implies that 'A and B are not independent conditional on C'.

A1 The probability of receiving treatment, conditional on the assignment variable, is discontinuous at the threshold:

$$\lim_{x \to x_0^-} \mathbb{P}(T_i = 1 \mid X_i = x) \neq \lim_{x \to x_0^+} \mathbb{P}(T_i = 1 \mid X_i = x).$$

A2 The threshold indicator and treatment indicator are not independent:

$$T_i \not\perp\!\!\!\perp Z_i.$$

A3 The threshold indicator is independent from confounding variables, conditional on the assignment variable:

$$Z_i \perp\!\!\!\perp \mathcal{C}_i | X_i.$$

A4 The expectation of the binary outcome Y_i, conditional on the assignment variable, is continuous at the threshold given treatment allocation:

$$\mathbb{E}(Y_i | X_i = x, T_i = t) \text{ is continuous at } x = x_0 \text{ for each of } t \in \{0, 1\}.$$

A5 The binary outcome is independent of the threshold indicator, given the other variables. That is:

$$Y_i \perp\!\!\!\perp Z_i | (T_i, X_i, \mathcal{C}_i).$$

A6 The population of interest (and sample of data used) does not contain 'defiers' – that is, subjects who would *never* take the treatment if it were offered or systems where the opposite of the decision rule would be applied to some subjects. This assumption implies that the treatment effect estimate is valid only for populations where the treatment would be taken if offered, thereby precluding the estimation of an overall, average, treatment effect using an RD design.

With any modelling approach it is important to consider whether or not assumptions are valid prior to working with a given dataset. Some of the above assumptions can be explored/checked prior to fitting an RD design. For example, A1 is often explored by producing a plot of the probability of treatment within small assignment variable bins either side of the threshold and producing a scatterplot to see if there is visual evidence of a discontinuity. A2 may be checked by calculating the correlation between the treatment and threshold indicators or using an appropriate hypothesis test of association. A3 cannot be tested formally as the treatment decision rule is devised externally (rather than on individual subject characteristics). However, it may be appropriate in many scenarios to assume that - in a region around the threshold - a subject's confounding variables cannot be manipulated to influence where their assignment variable lies in relation to the threshold/external decision rule. Assumption A4 can be explored by producing plots of the assignment variable against the probability of treatment within small assignment variable bins for the separate treated and untreated subject groups. Assumption A5 is also not testable

but may hold subject to the user having chosen a bandwidth such that the groups of subjects above and below the threshold are balanced with respect to confounding variables. Often we may produce summary statistics for potential confounders for these two groups prior to applying a design to check that there is balance between groups in a similar manner to the checking of randomised groups at baseline in a randomised controlled trial.

3 Estimating the Risk Ratio

Our focus in on the use of an RD design where the outcome of interest is binary and our interest lies in estimating the risk ratio with regard to the binary outcome of interest, comparing treated and untreated groups in a region close to the threshold. We assume that a window around the threshold exists such that where $X_i \in (x_0 - h, x_0 + h)$ subjects are exchangeable and groups of subjects with assignment variable values above and below the threshold may be considered similar with regard to potential confounding variables, $\mathcal{C}_i$.

Our estimand of interest is the risk ratio at the threshold, defined

$$\text{RR} = \frac{\lim_{x \to x_0} \mathbb{E}(Y|T = 1, X = x)}{\lim_{x \to x_0} \mathbb{E}(Y|T = 0, X = x)}.$$

and, in words, this is simply the ratio of the probability of the event of interest occurring for the treated and untreated groups.

We note that odds ratio estimation to describe a treatment effect is commonplace in many settings. However, the odds ratio is a non-collapsible measure, which can sometimes be problematic when estimating [30] and, in general, causal odds ratio estimators may be more biased than causal risk ratio estimators [14].

We outline two common approaches to risk ratio estimation using observational data: the Wald estimator and a multiplicative structural mean model. We then

derive a new risk ratio estimator based on RD design assumptions within a local randomisation framework that will be compared to the two aforementioned methods.

3.1 Wald Estimator

The Wald estimator is an adapted version of that for continuous outcomes [31] and expresses the log of the risk ratio as a local average treatment effect estimator, using a difference in natural logarithms of the expectation of the outcome of interest, defined

$$\log(\text{WALD-RR}) = \frac{\log[\mathbb{E}(Y_i|Z_i = 1)] - \log[\mathbb{E}(Y_i|Z_i = 0)]}{\mathbb{E}(T_i|Z_i = 1) - \mathbb{E}(T_i|Z_i = 0)}.$$

For this estimator to be valid and estimate a causal risk ratio at the threshold, we require the two following assumptions to hold [14, 32]:

W1: $\log[\mathbb{E}(Y_i|T_i = t)]$ is linear in t;

W2: $\mathbb{E}(T_i|Z_i = z)$ is linear in z.

Further explanation of these assumptions and their purpose for WALD-RR estimation is provided in Appendix B.

This estimator has been used in the instrumental variables (IV) framework and has been shown to be consistent for the risk ratio where the treatment effect is small and the event of interest is rare [32]. Considering the RD design threshold as analogous to an instrument in an IV framework and, subject to the RD design assumptions holding, this estimator can be used for RD designs with rare outcomes and a small treatment effect. However, in other situations we may wish to consider alternative approaches.

3.2 Multiplicative Structural Mean Model

Structural models based on potential outcomes have been established as a method for causal effect estimation [33, 34, 35]. For the binary outcome Y_i we define $Y_i^{(1)}$

and $Y_i^{(0)}$ to be the potential outcomes for the i^{th} subject if treated or not treated, respectively. Although each outcome is defined, only one may be observed in a given dataset because the same subject cannot be both treated and not treated.

The multiplicative structural mean model (MSMM) aims to compare the log expectations of the potential outcomes for the treated group [35], written

$$\log \mathbb{E}(Y_i^{(1)}|Z_i, T_i = 1) - \log \mathbb{E}(Y_i^{(0)}|Z_i, T_i = 1) = \psi_0 + \psi_1 Z_i.$$

Here $\exp(\psi_0)$ and $\exp(\psi_0+\psi_1)$ are risk ratios for the treated (RRT) for subjects with assignment variable values below and above the threshold, respectively. This model can be used to produce a consistent estimate of the RRT subject to the following additional assumptions [32, 34]:

M1: There is no interaction between the threshold indicator, Z_i, and the treatment variable, T_i (i.e. no effect modification).

M2: $\log[\mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z)] - \log[\mathbb{E}(Y_i^{(0)}|T_i = t, Z_i = z)]$ is linear in t.

Under assumption **M1**, $\psi_1 = 0$ and it follows that the treatment effect of interest is $\exp(\psi_0)$. Furthermore, a relationship between the overall risk ratio (known as the 'causal risk ratio' - CRR) and the RRT is shown below.

$$\frac{\mathbb{E}(Y_i^{(1)}|T_i = 1)}{\mathbb{E}(Y_i^{(0)}|T_i = 1)} = \frac{\mathbb{E}(Y_i^{(1)}|T_i = 1)}{\mathbb{E}(Y_i^{(0)}|T_i = 0)} \frac{\mathbb{E}(Y_i^{(0)}|T_i = 0)}{\mathbb{E}(Y_i^{(0)}|T_i = 1)};$$

$$\text{i.e. RRT} = \text{CRR} \times \text{SB}.$$

Here 'SB' denotes the 'selection bias' which measures the ratio of the expectation of the observed outcome for untreated subjects and the expectation of the counterfactual outcome for treated subjects. If treatment allocation is random and treated and untreated groups are balanced with respect to confounding variables then this ratio should be 1 and there would be no selection bias. In reality, and subject to

the RD design assumptions holding, we should expect that for subjects 'just above' and 'just below' the threshold (i.e. where $X_i \in (x_0 - h, x_0 + h)$) groups are balanced with respect to confounders and that selection bias will be minimal in the above expression.

An analytic expression for the RRT, derived using a MSMM, was provided by Hernán and Robins [34] and is shown below

$$\text{RRT} = 1 - \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\mathbb{E}(Y_i(1 - T_i)|Z_i = 1) - \mathbb{E}(Y_i(1 - T_i)|Z_i = 0)}. \tag{1}$$

In practice, because of assumption **M2**, generalised least squares models (with a log link function) are typically used to estimate individual components of (1). However, when Y_i is a binary outcome a logit link function would be more natural as this would ensure that probability estimates lie between 0 and 1. Geneletti et al. [18] have shown that using logistic regression models in (1) yields a similar estimate to that obtained when log-linear models are used but with the added advantage of probability estimates that always remain in the correct range. Moreover it can be shown that when Y_i is a binary variable and logistic regression models are used to estimate components of (1) then Assumption **M2** is satisfied (see Appendix C).

3.3 RD Design Method

As noted previously, the Wald estimator (defined in Section 3.1) relies on additional assumptions **W1** and **W2** and is valid only where the binary event of interest is rare and the treatment effect is small. In addition, the MSMM estimator (defined in Section 3.2) relies on additional assumptions **M1** and **M2**. As as result, we propose and derive an estimator for the risk ratio using only the RD design assumptions stated in Section 2.1 that is applicable to non-rare binary events and a variety of treatment effect magnitudes.

To define this estimator, we note that the treatment effect of interest is the risk

ratio at the threshold, written

$$\lim_{x \to x_0} \frac{\mathbb{E}(Y_i|T_i = 1, X_i = x)}{\mathbb{E}(Y_i|T_i = 0, X_i = x)}.$$

Owing to the effect of possible unobserved confounding, the numerator and denominator of the above expression cannot be estimated directly (for example, using a logistic regression model for Y_i with T_i and X_i as explanatory variables). However, in the RD design, the threshold is defined externally and, subject to the assumptions outlined in Section 2.1, we can exploit the threshold as a quasi-randomising device to derive the following estimator for the risk ratio at the threshold, which we term RDD-RR.

$$\text{RDD-RR} = 1 - \lim_{x \to x_0} \frac{\mathbb{E}(Y_i|Z_i = 1, X_i = x) - \mathbb{E}(Y_i|Z_i = 0, X_i = x)}{\mathbb{E}(Y_i|Z_i = 1, X_i = x)\mathbb{E}(T_i|Z_i = 0) - \mathbb{E}(Y_i|Z_i = 0, X_i = x)\mathbb{E}(T_i|Z_i = 1)}.$$

A proof of this result is provided in Appendix D.

In practice, because Y_i and T_i are binary, logistic regression models may be fitted to estimate each component of the above expression. Standard error estimation is less straightforward and, in this work, we use a bootstrap approach to obtain standard error estimates for each of the three estimation methods. Specifically, bootstrapping is used to estimate the standard error of the log risk-ratio using 2000 bootstrap replications. A corresponding 95% confidence interval can then be calculated for the log risk ratio using a normal approximation.

4 Simulation Studies

Having outlined three methods for risk ratio estimation in an RD design we perform simulation studies to evaluate and compare these estimation approaches. This is done for a variety of designs with differing levels of fuzziness and confounding.

4.1 Data Extraction

In an effort to make the simulation studies realistic, we extracted a dataset from The Health Improvement Network (THIN) - a large source of anonymised electronic patient records collected from over 500 UK general practice (GP) surgeries in which patients are generally representative of the UK population [23] - on the prescription of statins based on 10-year CVD risk score. Data from 1384 male patients aged between 50 and 70 years who were non-diabetic, non-smokers, had never experienced a cardiovascular event (stroke or myocardial infarction) and for whom a 10-year CVD risk score was calculated between January 2007 and December 2008 were extracted. In this example, the 10-year CVD risk score is the assignment variable with the treatment threshold set as a 10-year CVD risk score of 0.2 or greater, using the NICE guidance that was in place at the time these data were collected. The treatment is the prescription of statins and the outcome of interest is the binary event Y_i, which records whether or not subject i's LDL cholesterol level decreases by 1mmol/L or more, defined

$$Y_i = \begin{cases} 1 & \text{if subject } i\text{'s LDL cholesterol level decreases by 1mmol/L or more;} \\ 0 & \text{otherwise.} \end{cases}$$

Considering the time scale for measuring the change in LDL cholesterol, in the dataset each subject's LDL cholesterol level is recorded at the same time as the risk score is calculated. We then took the next LDL cholesterol measurement at least one month after the risk score calculation and used this value to determine the change in LDL cholesterol level and define Y_i.

4.2 Simulation Study set-up

We use the 1384 10-year CVD risk scores from the extracted data as assignment variables, X_i ($i \in \{1, \ldots, 1384\}$), with the threshold set as $x_0 = 0.2$. We simulated

2000 datasets, each with 1384 patients, under a variety of design fuzziness levels and confounding scenarios, which we will outline. First, we describe the general simulation algorithm.

Step 1. For the i^{th} subject, the assignment variable X_i is taken and a centred assignment variable X_i^c and threshold indicator Z_i are defined as

$$X_i^c = X_i - 0.2$$
$$Z_i = I\{X_i^c \geq 0\}.$$

Step 2. A confounding variable, U_i, is simulated from a continuous uniform distribution where

$$U_i \sim \text{Uniform}(0, 1).$$

Step 3. The probability of receiving treatment (statins), p_i, is defined

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Z_i + \beta_2 U_i + \beta_3 X_i^c.$$

The parameters β_1 and β_3 govern the strength of the treatment guideline (the effect of the threshold and assignment variable on the probability of treatment). β_2 allows the probability of receiving treatment to be affected by the confounding variable.

Step 4. Using the probability calculated in Step 3, a treatment indicator, T_i, is simulated where

$$T_i \sim \text{Bernoulli}(p_i).$$

Step 5. The expectation of the binary outcome of interest, p_i^y, is calculated as

$$\log\left(\frac{p_i^y}{1-p_i^y}\right) = \beta_4 + \beta_5 T_i + \beta_6 U_i.$$

Here β_5 governs the strength of the relationship between treatment and the binary outcome and β_6 allows an association between the binary outcome and confounding variable.

Step 6. Using p_i^y from Step 5, the binary outcome Y_i is simulated as follows

$$Y_i \sim \text{Bernoulli}(p_i^y).$$

Step 7. Steps 1–6 are repeated for each $i \in \{1, \ldots, 1384\}$ to create a dataset containing data from 1384 subjects.

Step 8. Steps 1–7 are repeated M times to create M simulated datasets.

The true causal risk ratio is calculated as

$$\text{RR} = \frac{\int_{\mathcal{U}} \text{expit}(\beta_4 + \beta_5 + \beta_6 u)\mathrm{d}u}{\int_{\mathcal{U}} \text{expit}(\beta_4 + \beta_6 u)\mathrm{d}u}$$

where $\text{expit}(x) = \frac{\exp(x)}{1+\exp(x)}$.

The pre-specified values of the parameters $\beta_0, \ldots, \beta_6$ are chosen to reflect the real THIN dataset and also adjusted to produce various levels of fuzziness and confounding. Table 1 shows choices of β_1, β_2 and β_6 used to create these scenarios, in which we use the terms 'weak' and 'strong' fuzziness and 'no', 'low' and 'high' confounding. Having outlined these parameters, the parameter β_0 is selected such that the probability of receiving treatment (across all subjects) is approximately 0.5. This is in line with the equivalent probability in the THIN statins dataset described in Section 4.1 and the value of β_0 was obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n} \text{expit}\left(\beta_0 + \beta_1 Z_i + \beta_2 U_i + \beta_3 X_i^c\right) = 0.5.$$

Similarly, for each scenario, β_4 was set such that the occurrence of the binary out-

come of interest has probability 0.44 across all subjects, similar to that seen in the THIN data example, with the value of β_4 obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n} \text{expit}\left(\beta_4 + \beta_5 T_i + \beta_6 U_i\right) = 0.44.$$

We set β_5 to be 1.5 which reflects the treatment effect (risk ratio). For each simulation run, the probability of complying with the treatment guideline was calculated as

$$\mathbb{P}(T_i = 1 \mid Z_i = 1) - \mathbb{P}(T_i = 1 \mid Z_i = 0)$$

and the mean of these probabilities is reported as the probability of compliance (P.C.) in Table 1. In addition, Table 1 shows sample correlations between the confounder U_i and each of the treatment indicator, T_i, and outcome Y_i.

Table 1: Parameter values for the simulation scenarios with the corresponding probability of compliance (P.C.) and estimates of correlations between Y_i and U_i ($\rho_{Y,U}$) and T_i and U_i ($\rho_{T,U}$)

Scenario	Parameters			P.C.	$\rho_{Y,U}$	$\rho_{T,U}$
Weak Fuzziness, No Confounding	$\beta_1 = 6$	$\beta_2 = 0$	$\beta_6 = 0$	0.91	0.00	0.00
Weak Fuzziness, Low Confounding	$\beta_1 = 8$	$\beta_2 = 6.5$	$\beta_6 = 1$	0.90	0.16	0.12
Weak Fuzziness, High Confounding	$\beta_1 = 8$	$\beta_2 = -9$	$\beta_6 = 2.5$	0.81	0.29	-0.24
Strong Fuzziness, No Confounding	$\beta_1 = 2$	$\beta_2 = 0$	$\beta_6 = 0$	0.53	0.00	0.00
Strong Fuzziness, Low Confounding	$\beta_1 = 2$	$\beta_2 = 1.5$	$\beta_6 = 1$	0.52	0.16	0.15
Strong Fuzziness, High Confounding	$\beta_1 = 2.5$	$\beta_2 = -3.5$	$\beta_6 = 2$	0.53	0.21	-0.31

For each simulation scenario we estimated the risk ratio using the three approaches outlined in Section 3: the Wald estimator (WALD-RR), the multiplicative structural mean model estimator (MSMM-RR) and the proposed RD design method (RDD-RR). A total of 2000 datasets were simulated for each simulation scenario and results are reported for bandwidths of 0.05, 0.1, 0.15 and 0.2.

4.3 Simulation Study: Results

For each simulation scenario, and where the true risk ratio is equal to 1.5, Table 2 shows numerical summaries of simulation study results with box plots of risk ratio estimates shown in Figure 1 (weak fuzziness) and Figure 2 (strong fuzziness).

Examining Table 2 and Figures 1 and 2, where there is no confounding all methods produce unbiased estimates of the log risk ratio with similar standard error estimates. Standard error estimates are generally larger where fuzziness is stronger and the bandwidth decreases, which would be expected. For the low confounding scenario the three approaches produce generally similar estimates with a similar pattern regarding standard errors which increase as the bandwidth - and thus the sample size - decreases. For the high confounding scenario the MSMM-RR approach is more biased than either the WALD-RR or RDD-RR methods for all bandwidths. The RDD-RR and WALD-RR approaches produce similar results and, as with the low and no confounding scenarios, standard error estimates generally increase as the fuzziness increases and as the bandwidth decreases.

That the WALD-RR and RDD-RR methods yield similar results is not surprising as each of the assumptions required for the WALD-RR estimator to be valid, outlined in Section 3.1, holds here. Conversely, the MSMM-RR method may be less desirable in that the estimates can be biased where confounding is high and, in addition, the time taken to produce estimates was slightly longer than either the WALD-RR or RDD-RR approaches. One possible drawback of the WALD-RR method is that we require the treatment effect to be reasonably small and the event of interest to be rare. As a result we ran further simulations where the true risk ratio was increased to 4.

Table 3 shows numerical summaries of simulation results where the true risk ratio is 4, with box plots of risk ratio estimates shown in Figures 3 and 4. We see that, across all bandwidths, confounding levels and fuzziness scenarios the WALD-

Table 2: Estimates, Biases, empirical standard errors (ESE), average standard errors (ASE) and 95% coverage (95% Cov.) of the log risk ratio. The true value of the log of the risk ratio is log(1.5) = 0.405. The original sample size was 1384 in each simulated dataset and simulations were repeated 2000 times.

			Weak fuzziness					*Strong fuzziness*				
	Bandwidth	Method	Estimate	Bias	ESE	ASE	95% Cov.	Estimate	Bias	ESE	ASE	95% Cov.
No confounding	0.2	RDD-RR	0.40	0.01	0.13	0.13	94.9	0.38	0.02	0.19	0.20	96.2
		MSMM-RR	0.41	-0.01	0.15	0.15	95.6	0.42	-0.02	0.22	0.24	97.2
		WALD-RR	0.41	-0.01	0.14	0.14	94.8	0.39	0.02	0.20	0.20	95.7
	0.15	RDD-RR	0.40	0.01	0.14	0.14	95.0	0.38	0.02	0.20	0.20	95.8
		MSMM-RR	0.41	-0.01	0.15	0.16	95.9	0.42	-0.01	0.23	0.25	97.1
		WALD-RR	0.41	-0.01	0.15	0.15	95.2	0.39	0.02	0.20	0.20	95.4
	0.1	RDD-RR	0.40	0.00	0.16	0.16	95.2	0.39	0.02	0.23	0.24	96.2
		MSMM-RR	0.41	-0.01	0.18	0.18	96.3	0.42	-0.02	0.26	0.30	97.6
		WALD-RR	0.42	-0.01	0.17	0.17	95.0	0.40	0.01	0.23	0.24	95.5
	0.05	RDD-RR	0.41	0.00	0.22	0.22	95.5	0.40	0.01	0.31	0.34	98.2
		MSMM-RR	0.43	-0.02	0.25	0.26	97.0	0.44	-0.03	0.38	0.44	98.1
		WALD-RR	0.42	-0.02	0.23	0.23	95.4	0.40	0.00	0.32	0.33	96.7
Low confounding	0.2	RDD-RR	0.40	0.00	0.13	0.13	95.5	0.38	0.03	0.20	0.20	95.2
		MSMM-RR	0.40	0.01	0.14	0.14	95.3	0.41	-0.01	0.23	0.24	96.8
		WALD-RR	0.41	-0.01	0.14	0.14	95.5	0.38	0.02	0.20	0.20	94.7
	0.15	RDD-RR	0.40	0.00	0.14	0.14	95.6	0.38	0.02	0.21	0.21	95.3
		MSMM-RR	0.40	0.01	0.14	0.15	95.7	0.41	-0.01	0.24	0.25	96.5
		WALD-RR	0.41	-0.01	0.14	0.15	96.0	0.39	0.02	0.21	0.21	94.8
	0.1	RDD-RR	0.40	0.00	0.15	0.16	95.9	0.39	0.02	0.24	0.24	95.7
		MSMM-RR	0.40	0.00	0.16	0.17	96.1	0.42	-0.01	0.28	0.30	97.1
		WALD-RR	0.42	-0.01	0.16	0.17	95.9	0.39	0.01	0.24	0.24	94.6
	0.05	RDD-RR	0.40	0.00	0.22	0.22	95.0	0.40	0.01	0.33	0.35	98.2
		MSMM-RR	0.41	0.00	0.23	0.24	96.3	0.44	-0.03	0.40	0.46	98.5
		WALD-RR	0.42	-0.02	0.23	0.23	94.9	0.40	0.00	0.33	0.34	96.2
High confounding	0.2	RDD-RR	0.40	0.01	0.14	0.14	94.3	0.39	0.02	0.21	0.21	95.3
		MSMM-RR	0.45	-0.05	0.17	0.17	96.5	0.44	-0.04	0.27	0.29	97.5
		WALD-RR	0.41	-0.01	0.14	0.14	94.5	0.39	0.02	0.21	0.21	94.7
	0.15	RDD-RR	0.40	0.00	0.14	0.14	94.5	0.39	0.02	0.22	0.22	96.0
		MSMM-RR	0.45	-0.05	0.18	0.18	96.7	0.44	-0.04	0.28	0.30	97.9
		WALD-RR	0.41	-0.01	0.15	0.15	94.5	0.39	0.01	0.22	0.22	95.2
	0.1	RDD-RR	0.40	0.00	0.16	0.16	95.7	0.39	0.01	0.25	0.26	97.0
		MSMM-RR	0.46	-0.05	0.20	0.21	97.8	0.45	-0.04	0.31	0.36	98.3
		WALD-RR	0.42	-0.01	0.17	0.17	95.8	0.40	0.01	0.25	0.26	96.0
	0.05	RDD-RR	0.41	0.00	0.22	0.22	95.6	0.41	0.00	0.35	0.38	98.6
		MSMM-RR	0.47	-0.07	0.28	0.31	97.9	0.47	-0.06	0.44	0.53	98.9
		WALD-RR	0.42	-0.02	0.23	0.23	95.5	0.41	-0.01	0.35	0.36	96.2

RR estimates are biased. Conversely, the RDD-RR approach appears to produce unbiased estimates when fuzziness is weak, even when confounding is high. For the

strong fuzziness scenarios the RDD-RR approach performs reasonably well when the bandwidth is small but becomes more biased as the bandwidth increases. This is not uncommon in RD designs, where careful consideration should be given to a suitable bandwidth choice and, generally - when using the local randomisation approach - a smaller bandwidth with a reasonable sample size (i.e. that which would yield a precise risk ratio estimate) would be desirable. Overall, we see that the RDD-RR approach may be a more flexible estimation approach than the WALD-RR method and is able to produce an unbiased risk ratio estimate in a wider variety of scenarios.

Table 3: Estimates, Biases, empirical standard (ESE), average standard errors (ASE) and 95% coverage (95% Cov.) of the log of the risk ratio. The true value of the log of the risk ratio is log(4) = 1.387. The sample size was 1384 in each simulated dataset and simulations were repeated 2000 times.

			Weak fuzziness					*Strong fuzziness*				
	Bandwidth	Method	Estimate	Bias	ESE	ASE	95% Cov.	Estimate	Bias	ESE	ASE	95% Cov.
No confounding	0.2	RDD-RR	1.38	0.01	0.18	0.18	94.2	1.30	0.09	0.27	0.27	91.8
		WALD-RR	1.50	-0.11	0.21	0.20	92.9	1.27	0.12	0.23	0.23	90.6
	0.15	RDD-RR	1.38	0.00	0.18	0.18	94.3	1.30	0.09	0.27	0.28	92.7
		WALD-RR	1.50	-0.11	0.21	0.21	92.2	1.27	0.11	0.24	0.24	91.4
	0.1	RDD-RR	1.39	0.00	0.21	0.20	94.5	1.32	0.07	0.31	0.33	93.4
		WALD-RR	1.51	-0.13	0.24	0.24	93.0	1.29	0.10	0.27	0.27	92.6
	0.05	RDD-RR	1.41	-0.02	0.29	0.29	95.0	1.37	0.01	0.47	0.50	94.5
		WALD-RR	1.54	-0.15	0.34	0.33	94.5	1.32	0.06	0.38	0.38	94.0
Low confounding	0.2	RDD-RR	1.38	0.01	0.18	0.18	94.7	1.29	0.10	0.27	0.28	91.8
		WALD-RR	1.50	-0.11	0.20	0.20	93.0	1.25	0.13	0.23	0.23	89.7
	0.15	RDD-RR	1.38	0.00	0.18	0.18	94.9	1.29	0.10	0.27	0.29	92.6
		WALD-RR	1.50	-0.11	0.21	0.21	93.6	1.26	0.13	0.23	0.24	90.5
	0.1	RDD-RR	1.39	0.00	0.20	0.20	95.0	1.32	0.07	0.32	0.34	94.1
		WALD-RR	1.51	-0.12	0.23	0.23	93.8	1.28	0.11	0.27	0.28	93.1
	0.05	RDD-RR	1.40	-0.01	0.28	0.29	95.7	1.39	-0.01	0.51	0.52	94.2
		WALD-RR	1.52	-0.14	0.32	0.33	96.0	1.32	0.06	0.39	0.39	94.0
High confounding	0.2	RDD-RR	1.38	0.01	0.18	0.18	95.4	1.33	0.06	0.29	0.32	94.1
		WALD-RR	1.49	-0.10	0.20	0.21	94.2	1.28	0.11	0.24	0.25	92.6
	0.15	RDD-RR	1.38	0.01	0.18	0.18	95.8	1.33	0.05	0.30	0.32	94.3
		WALD-RR	1.49	-0.11	0.20	0.21	94.3	1.28	0.11	0.25	0.25	92.6
	0.1	RDD-RR	1.38	0.00	0.21	0.21	95.1	1.36	0.02	0.36	0.39	95.2
		WALD-RR	1.50	-0.11	0.24	0.24	93.9	1.30	0.09	0.29	0.29	94.0
	0.05	RDD-RR	1.39	-0.01	0.29	0.29	95.0	1.45	-0.06	0.56	0.57	94.9
		WALD-RR	1.51	-0.12	0.33	0.33	95.2	1.35	0.04	0.41	0.42	95.3

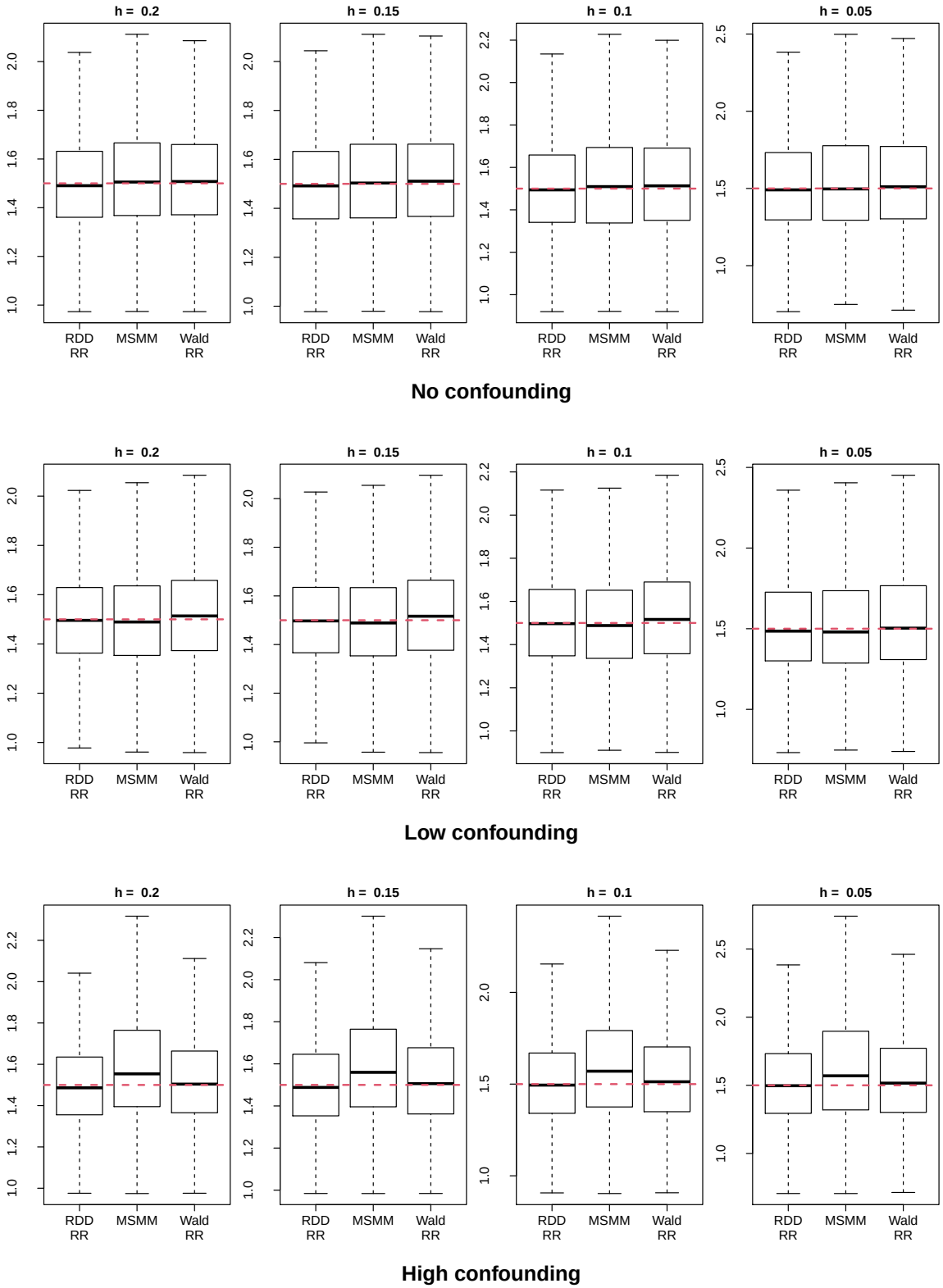

Figure 1: Boxplots showing log risk ratio estimates for the RDD-RR, MSMM-RR and WALD-RR approaches for different bandwidths (h) and confounding levels where the true risk ratio is 1.5. Simulations performed using a weak fuzziness scenario.

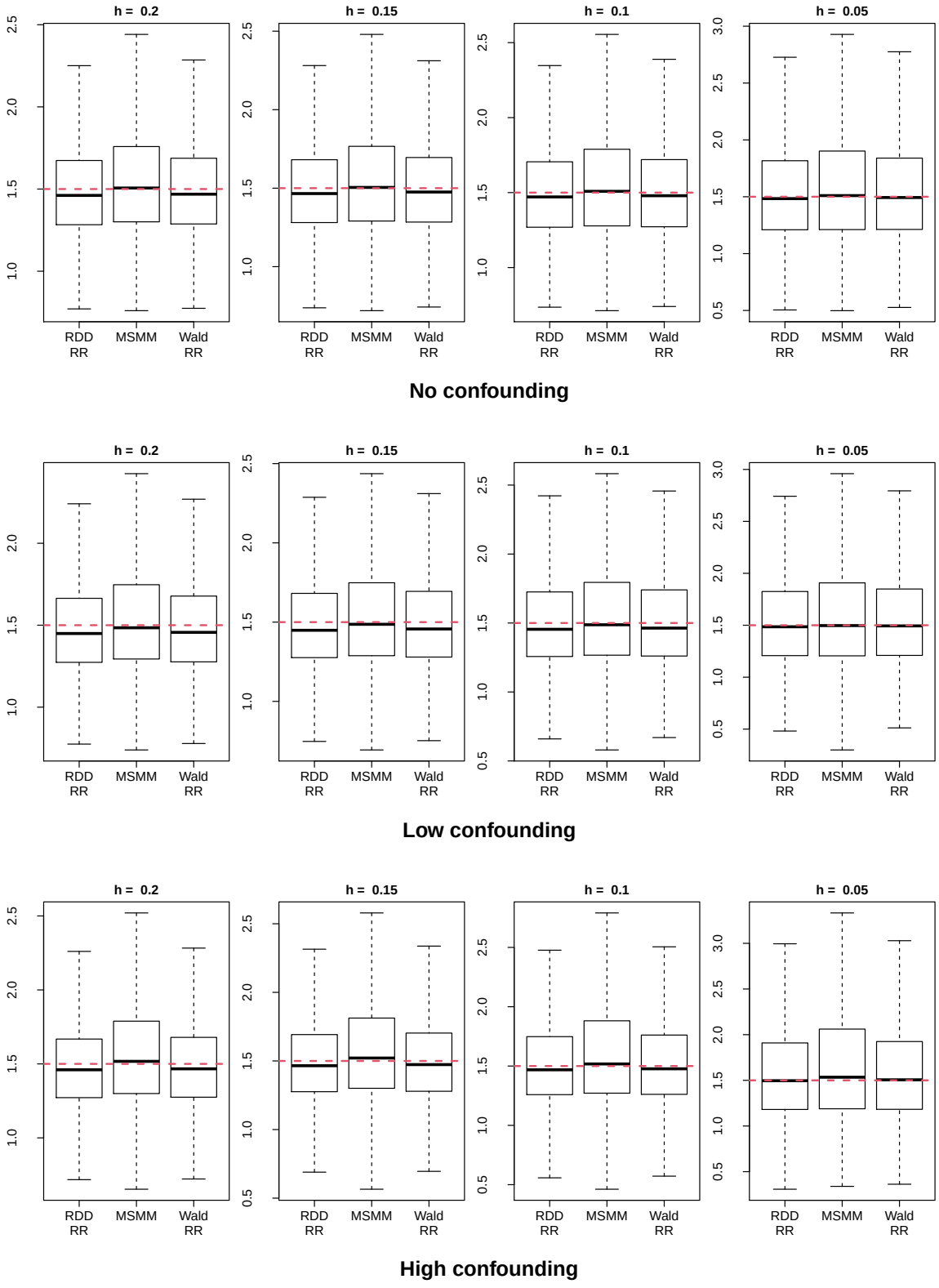

Figure 2: Boxplots showing risk ratio estimates for the RDD-RR, MSMM-RR and WALD-RR approaches for different bandwidths (h) and confounding levels where the true risk ratio is 1.5. Simulations were performed using a strong fuzziness scenario.

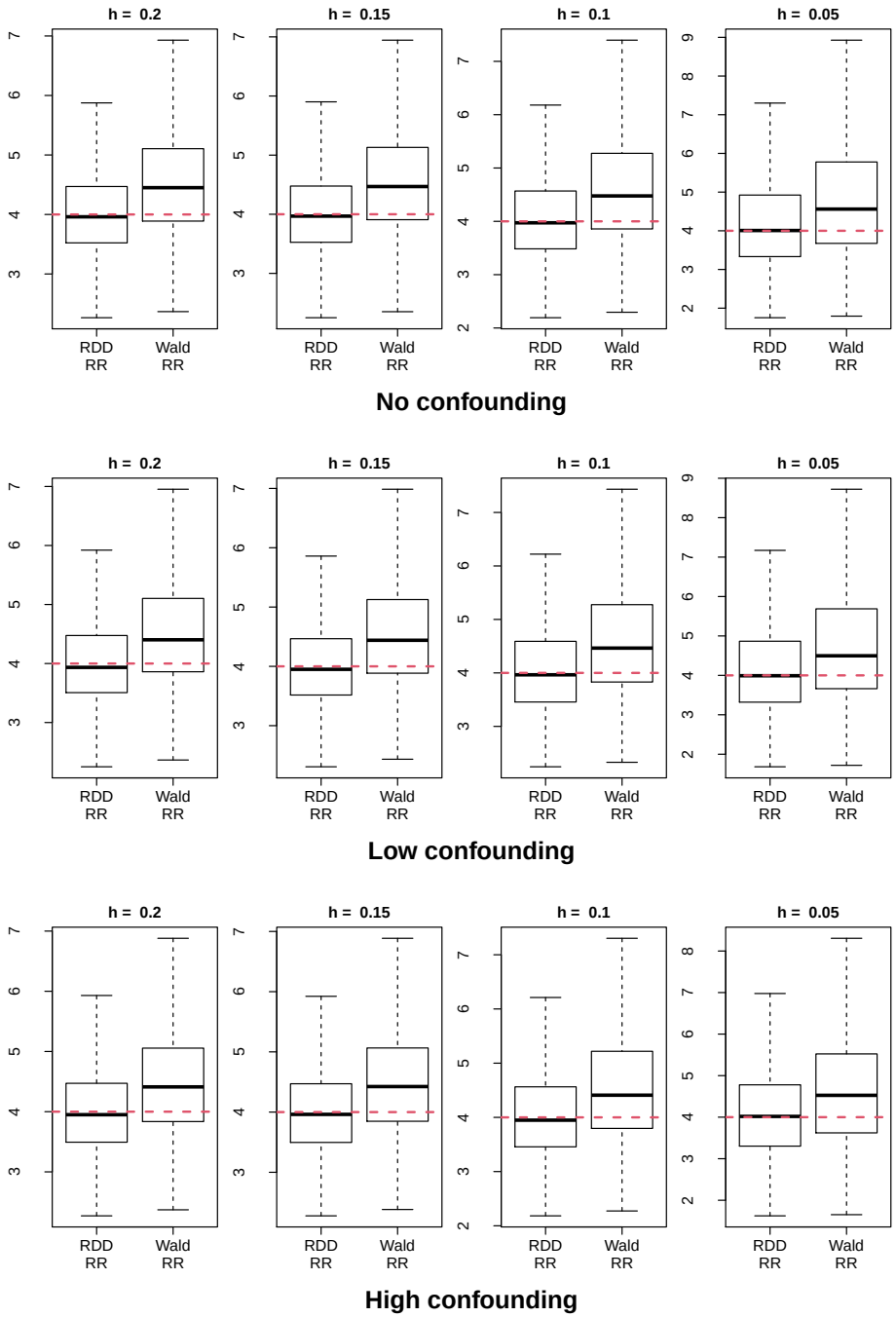

Figure 3: Boxplots showing log risk ratio estimates for the RDD-RR, MSMM-RR and WALD-RR approaches for different bandwidths (h) and confounding levels where the true risk ratio is 4. Simulations performed using a weak fuzziness scenario.

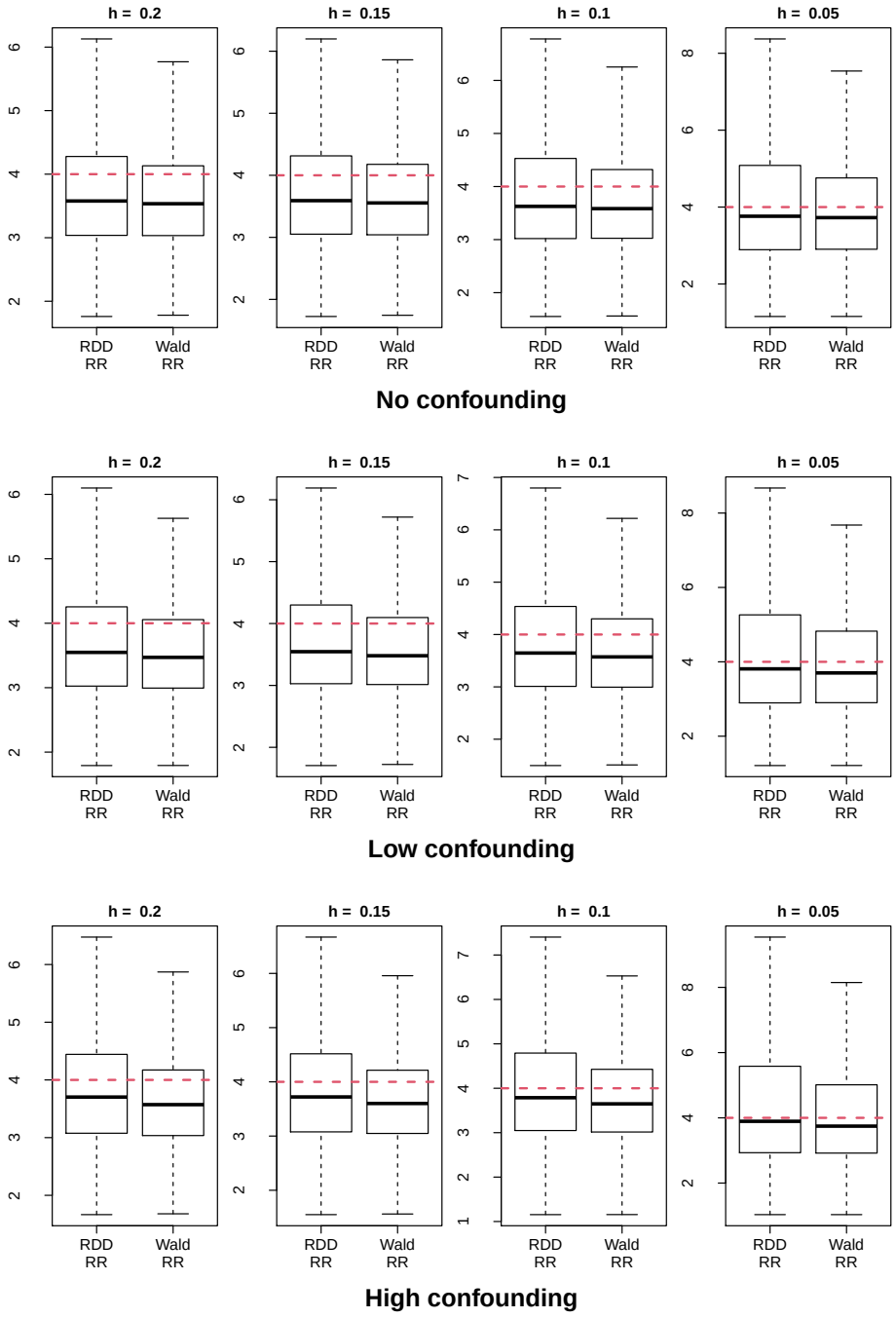

Figure 4: Boxplots showing log risk ratio estimates for the RDD-RR, MSMM-RR and WALD-RR approaches for different bandwidths (h) and confounding levels where the true risk ratio is 4. Simulations performed using a strong fuzziness scenario.

5 Example: Prescription of Statins in UK Primary Care

Having described the dataset on the prescription of statins in Section 4.1 we apply the three risk ratio estimation methods to these data to estimate the risk ratio for the prescription of stations on the event: a 1mmol/L or greater reduction in LDL cholesterol, using an RD design. Here the assignment variable is the 10-year CVD risk score and the treatment threshold is a 10-year CVD risk score of 0.2. Of the 1384 patients in the dataset, 705 patients (51%) were prescribed statins. 830 patients (60%) have 10-year CVD risk scores ≥ 0.2 and the empirical probabilities of receiving statins for patients with risk scores above and below the 0.2 threshold are 0.73 and 0.18, respectively.

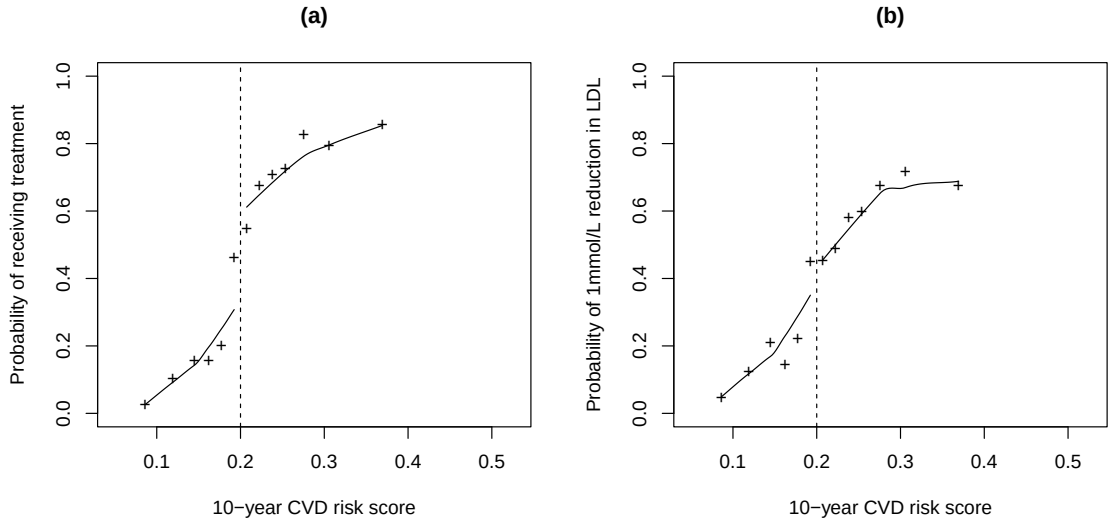

Figure 5: Plots of (a) the treatment indicator and (b) the indicator of 1 mmol/L reduction in LDL cholesterol level against the risk of developing CVD in 10 years. The black crosses and lines are the expected probabilities calculated in bins.

Figure 5 shows plots of the probability of receiving statins (a) and the probability of a 1mmol/L reduction in LDL cholesterol (b) for groups of patients with similar 10-year CVD risk score values. Looking at these plots, we see that the probability of receiving a statin prescription changes rapidly around the risk score threshold of

0.2. There is also some visual evidence to suggest that the probability of a reduction of at least 1mmol/L in LDL cholesterol also increases as the 10-year CVD risk score exceeds 0.2 but whether or not this increase is a true discontinuity is less obvious.

The three estimation methods described in Section 3 were used to estimate the risk ratio with regard to the effect of statins on attaining a reduction of 1mmol/L or more in the LDL cholesterol level. As in the simulation studies, RD design bandwidths of 0.2, 0.15, 0.1 and 0.05 were used. In practice, we might pre-select an appropriate bandwidth before undertaking an RD design - perhaps using clinical or other expert knowledge to elicit an acceptable region in which me might expect groups of patients to be balanced with respect to confounding variables. In this analysis we compare estimates obtained using different bandwidths for demonstrative purposes. Table 4 shows risk ratio estimates together with associated 95% confidence intervals for each of the considered bandwidths.

Table 4: Risk ratio estimates and associated 95% confidence intervals for the effect of statins on a reduction of 1mmol/L or more in LDL cholesterol.

Method	Risk ratio estimate	95% CI
	Bandwidth = 0.2	
RDD-RR	1.43	(0.91, 2.23)
MSMM-RR	2.43	(0.92, 6.41)
WALD-RR	1.43	(0.91, 2.25)
	Bandwidth = 0.15	
RDD-RR	1.26	(0.79, 2.00)
MSMM-RR	1.81	(0.52, 6.31)
WALD-RR	1.26	(0.79, 2.01)
	Bandwidth = 0.1	
RDD-RR	1.17	(0.65, 2.11)
MSMM-RR	1.51	(0.32, 7.18)
WALD-RR	1.17	(0.65, 2.10)
	Bandwidth = 0.05	
RDD-RR	0.71	(0.24, 2.08)
MSMM-RR	0.38	(0.04, 3.78)
WALD-RR	0.72	(0.29, 1.80)

Examining these results we see that the RDD-RR and WALD-RR approaches produce similar estimates of the risk ratio for each bandwidth. In addition, the MSMM-

RR estimate differs substantially from the RDD-RR and WALD-RR estimates. Considering differences in estimates between bandwidths, for bandwidths of 0.2, 0.15 and 0.1, risk ratio estimates are greater than 1 but differ since risk ratio estimates decrease as the bandwidth decreases. For the bandwidth of 0.05, all risk ratio estimates are less than 1, although the precision of these estimates is lower than those for higher bandwidths because of a reduction in sample size. All 95% confidence intervals include 1 and we conclude that there is insufficient evidence to suggest a significant effect of statins on the occurrence of a reduction of 1mmol/L or more in LDL cholesterol in these data. The differing results as the bandwidth changes imply that the choice of bandwidth to reflect suitable, balanced groups above and below the threshold would be of importance when applying the RD design here.

We note that, for this demonstrative example, the continuous outcome: LDL cholesterol level has been dichotomised which is not necessarily the best way to assess the effect of statins on LDL cholesterol level owing to a loss of information [36], although this depends on the precise, clinically important reasons for a particular study or estimand of interest.

6 Discussion

Our focus was on the estimation of the treatment effect as a risk ratio in an RD design when the outcome of interest is binary and a local randomisation RD design approach is taken. Specifically, we developed an approach to estimate the risk ratio at the threshold with regard to binary event of interest using the standard RD design assumptions (known as the RDD-RR estimator) and compared this to the established Wald estimator from the IV literature (WALD-RR) and that from a multiplicative structural mean model (MSMM-RR).

The Wald estimator is known to be consistent for the risk ratio when the treatment effect is small and the event of interest is rare whereas the MSMM-RR approach

requires additional assumptions and may produce negative estimates of the risk ratio in some cases. In contrast, the RDD-RR approach taken in this work may be more stable and can be used for a variety of treatment effects. It is based on the RD design assumptions and, as such, fits naturally with other RD design approaches. Simulation studies showed that this estimator performs well when the RD design has weak fuzziness but less so when the fuzziness is stronger. In practice, if fuzziness is fairly strong then all RD design methods should be used with caution because, in the presence of strong fuzziness, the relationship between the threshold treatment rule and actual treatment allocation is weakened which may violate assumptions A1 and A2 of the RD design, outlined in Section 2.1. The performance of the estimators was investigated when the risk ratio is large (here, set to be equal to 4) and, in this case, it was observed that the Wald estimator becomes biased even when there is no confounding.

The three methods for estimating risk ratio were applied to a real dataset from UK primary care to explore the effect of statins on a clinically important reduction in LDL cholesterol. Estimates varied for different bandwidths which highlights the importance of selecting and justifying an appropriate RD design bandwidth when using an RD design with a local randomisation approach. None of the estimates was statistically significant at the 5% level but we noted that the Wald and RDD-RR approaches provided similar estimates across all bandwidths. This might suggest that the RDD-RR approach outlined in this work may be a useful alternative to the Wald method in general and that applying both methods to a given set of data may be of interest as a sensitivity analysis.

Based on the results from the simulation studies, the proposed RDD-RR approach appears to be a sound alternative to two existing methods in literature and is set up to fit naturally with RD design assumptions. We recommend its use should be considered when applying an RD design where the outcome of interest is binary.

Acknowledgements

Approval for use of the THIN data was obtained from the THIN Scientific Review Committee in August 2014.

References

[1] Thistlethwaite DL and Campbell DT. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 1960; 51(6): 309–317.

[2] Hahn J, Todd P and der Klaauw WV. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 2001; 69(1): 201–209.

[3] Imbens G and Lemieux T. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 2008; 142: 615–635.

[4] Mattei A and Mealli F. Regression discontinuity designs as local randomized experiments. *Observational Studies* 2017; 3: 156–173.

[5] Cattaneo M, Frandsen BR and Titiunik R. Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate. *Journal of Causal Inference* 2015; 3(1): 1—24.

[6] Li F, Mattei A and Mealli F. Bayesian inference for regression discontinuity designs with application to the evaluation of Italian university grants. *Annals of Applied Statistics* 2015; 9(4): 1906–1931.

[7] Cattaneo MD, Titiunik R and Vazquez-Bare G. Inference in regression discontinuity designs under local randomization. *The Stata Journal* 2016; 16(2): 331–367.

[8] Sekhon JS and Titiunik R. On interpreting the regression discontinuity design as a local experiment. In Cattaneo MD and Escanciano JC (eds.) *Regression Discontinuity Designs: Theory and Applications*, chapter 1. Emerald Publishing Ltd., 2017. pp. 1–28.

[9] Rutter M. Epidemiological methods to tackle causal questions. *International Journal of Epidemiology* 2009; 38(1): 3–6.

[10] Bor J, Moscoe E, Mutevedzi P et al. Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology* 2014; 25: 729–737.

[11] Geneletti S, O'Keeffe AG, Sharples LD et al. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine* 2015; 34(15): 2334–2352.

[12] Moscoe E, Bor J and Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *Journal of Clinical Epidemiology* 2015; 68(2): 122–133.

[13] Lesik SA. Applying the regression-discontinuity design to infer causality with non-random assignment. *The Review of Higher Education* 2006; 30(1).

[14] Palmer TM, Sterne JAC, Harbord RM et al. Instrumental variable estimation of causal risk ratios and causal odds ratios in Mendelian randomization analyses. *American Journal of Epidemiology* 2011; 173(12): 1392–1403.

[15] Xu KL. Regression discontinuity with categorical outcomes. *Journal of Econometrics* 2017; 201: 1–18.

[16] van Leeuwen N, Lingsma HF, Mooijaart SP et al. Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. *Journal of Clinical Epidemiology* 2018; 98: 70–79.

[17] Bor J, Fox MP, Rosen S et al. Treatment eligibility and retention in clinical HIV care: A regression discontinuity study in South Africa. *PLOS Medicine* 2017; 14(11): e1002463.

[18] Geneletti S, Ricciardi F, O'Keeffe AG et al. Bayesian modelling for binary outcomes in the regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2019; 182(3): 983–1002.

[19] Wilson PWF, D'Agostino RB, Belanger AM et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998; 97: 1837–1847.

[20] Hippisley-Cox J, Coupland C, Vinogradova Y et al. Predicting cardiovascular risk in england and wales: prospective derivation and validation of QRISK2. *BMJ* 2008; 336(7659): 1475–1482. Doi: 10.1136/bmj.39609.449676.25.

[21] Collins R, Reith C, Emberson J et al. Interpretation of the evidence for the efficacy and safety of statin therapy. *Lancet* 2016; 388(10059): 2532–2561.

[22] Bourke A, Dattani H and Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Journal of Innovation in Health Informatics* 2004; 12: 171—177.

[23] Blak BT, Thompson M, Dattani H et al. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Journal of Innovation in Health Informatics* 2011; 19: 251—255.

[24] Imbens G and Kalyanaraman K. Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies* 2012; 79(3): 933–959.

[25] Calonico S, Cattaneo MD and Titiunik R. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 2014; 82(6): 2295–2326.

[26] Ricciardi F, Liverani S and Baio G. Dirichlet process mixture models for regression discontinuity designs. *Statistical Methods in Medical Research* 2023; 23(1): 55–70.

[27] Imbens GW and Angrist JD. Identification and estimation of local average treatment effects. *Econometrica* 1994; 62(2): 467–475.

[28] Dawid AP. Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B (Methodological)* 1979; 41(1): 1–31.

[29] Lee DS and Lemieux T. Regression discontinuity designs in economics. *Journal of Economic Literature* 2010; 48: 281–355.

[30] Burgess S. Estimating and contextualizing the attenuation of odds ratios due to non collapsibility. *Communications in Statistics - Theory and Methods* 2017; 46(2): 786–804.

[31] Wald A. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 1940; 11(3): 284–300.

[32] Didelez V, Meng S and Sheehan NA. Assumptions of iv methods for observational epidemiology. *Statistical Science* 2010; 25(1): 22–40.

[33] Robins JM, Hernán MA and Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11: 550–560.

[34] Hernán MA and Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* 2006; 4(17): 360–372.

[35] Clarke PS and Windmeijer F. Identification of causal effects on binary outcomes using structural mean models. *Biostatistics* 2010; 11(4): 756–770.

[36] Altman DG and Royston P. The cost of dichotomising continuous variables. *BMJ* 2006; 332(7549): 1080. Doi: `10.1136/bmj.332.7549.1080`.

A Regression Discontinuity Design Assumptions

We provide interpretations of the RD design assumptions stated in Section 2.1.

A1 The probability of receiving treatment, conditional on the assignment variable, is discontinuous at the threshold:

$$\lim_{x \to x_0^-} \mathbb{P}(T_i = 1 | X_i = x) \neq \lim_{x \to x_0^+} \mathbb{P}(T_i = 1 | X_i = x).$$

This assumption states that the probability that a patient whose 10-year CVD risk score lies below the 20% threshold receives statins should differ from the probability that a patient whose 10-year CVD risk score lies above the threshold receives statins. This is a fundamental assumption for an RD design to be valid - that the probability of receiving treatment changes abruptly at the decision threshold. We should expect this assumption to hold in the statins example because if GPs follow the NICE 20% 10-year CVD risk guideline then there should be significantly more patients who are prescribed statins with 10-year CVD risk scores above the threshold than amongst patients with 10-year CVD risk scores below the threshold.

A2 The threshold indicator and treatment indicator are not independent:

$$T_i \not\perp\!\!\!\perp Z_i.$$

This assumption expresses the relationship between the threshold rule and treatment status. In the statins example it implies that the decision to treat a patient (prescribe statins) is associated with the 20% 10-year CVD risk score threshold. As discussed, typically there will be instances within a given dataset where a GP does not necessarily follow the 20% threshold rule. However, this does not imply that an RD is inappropriate - as long as there is a reasonably strong association between the threshold and treatment allocation then a fuzzy

RD design may be applied (subject to other assumptions being valid).

A3 The threshold indicator is independent from confounding variables, conditional on the assignment variable:

$$Z_i \perp\!\!\!\perp \mathcal{C}_i | X_i.$$

In the statins example, this assumption implies that the 20% threshold guideline (Z_i) depends on the 10-year CVD risk score only (X_i). Although this assumption is untestable, because the statins treatment guideline is set externally by NICE it would be unlikely that the threshold indicator would be influenced by any other patient-level variables. However, the assumption may be weakened if desired whereby the threshold indicator is assumed not to depend on unobserved confounders, given some relevant observed confounding variables. The design would then require either stratification by observed confounding variables (if appropriate) or conditioning on such variables in the modelling for treatment effect estimation to be valid.

A4 The expectation of the binary outcome Y_i, conditional on the assignment variable, is continuous at the threshold given treatment allocation:

$$\mathbb{E}(Y_i | X_i = x, T_i = t) \text{ is continuous at } x = x_0 \text{ for each of } t \in \{0, 1\}.$$

In words, this assumption implies that - conditional on treatment being fixed and not changing - the distribution of the outcome of interest does not exhibit a discontinuity at the threshold. This ensures that only a change in treatment would be responsible for any discontinuity in the distribution of Y_i at the threshold and not any other variables. In the statins example, this assumption implies that any change in the LDL cholesterol level in a region close to the threshold is because of the prescription of statins.

A5 The binary outcome is independent of threshold indicator, given the other

variables. That is:

$$Y_i \perp\!\!\!\perp Z_i | (T_i, X_i, \mathcal{C}_i).$$

This assumption ensures that the treatment threshold decision rule is a valid randomisation device - at least in a region close to the threshold - and that groups of patients above and below the threshold will be balanced with respect to confounding variables. In the statins example, it implies that a patient cannot manipulate their outcome (LDL cholesterol level) in order for their 10-year CVD risk score to lie above or below the treatment threshold and ensures that there is randomness with regard to where a patient's 10-year risk score lies in relation to the threshold. Although this assumption is untestable, in the local randomisation RD design framework, once the bandwidth (h) has been selected and prior to estimating the treatment effect at the threshold, it is usual to produce summary statistics for potentially important confounding variables separately for groups of patients with assignment variable values in $[x_0 - h, x_0)$ and $[x_0, x_0 + h]$. These summary statistics are then compared to ensure that there is balance between groups, in a similar way to the checking of patient variables at baseline in a two-group individually randomised controlled trial.

A6 No defiers/monotonicity assumption.

This assumes that there are no GPs (or patients) that will intentionally do the opposite of the treatment guideline. For instance, that no GPs would prescribe statins to all patients with risk scores below 20% and withhold statins from patients with risk scores above 20%. This behaviour is unlikely to happen in practice.

B Assumptions and Derivation of the Wald Risk Ratio Estimator

Assumption **W1** states that $\log[\mathbb{E}(Y_i|T_i = t)]$ is linear in t and this implies that

$$\begin{aligned} &\log[\mathbb{E}(Y_i|T_i = t)] = \alpha + \phi t \\ \implies &\mathbb{E}(Y_i|T_i = t) = e^{\alpha}e^{\phi t} \end{aligned} \tag{2}$$

for parameters α and ϕ.

We note that, under this assumption, the risk ratio can be written

$$\frac{\mathbb{E}(Y_i|T_i = 1)}{\mathbb{E}(Y_i|T_i = 0)} = e^{\phi}$$

and, hence, we seek to estimate ϕ.

Using assumption **W2** ($\mathbb{E}(T_i|Z_i = z)$ is linear in z) we write

$$T_i = \beta + \gamma Z_i + \epsilon_i \tag{3}$$

where ϵ_i is an error term such that $\mathbb{E}(\epsilon_i) = 0$ and ϵ_i and Z_i are independent.

Now we consider

$$\begin{aligned} \mathbb{E}(Y_i|Z_i = z) &= \mathbb{E}_{T_i|Z_i=z}\left[\mathbb{E}(Y_i|T_i)\right] \\ &= \mathbb{E}_{T_i|Z_i=z}\left[e^{\alpha}e^{\phi T_i}\right] \quad \text{(using (2))} \\ &= \mathbb{E}_{\epsilon_i|Z_i=z}\left[e^{\alpha}e^{\phi(\beta+\gamma z+\epsilon_i)}\right] \quad \text{(using (3))} \\ &= e^{\alpha}e^{\phi(\beta+\gamma z)}\mathbb{E}_{\epsilon_i|Z_i=z}(e^{\phi\epsilon_i}). \end{aligned}$$

Since ϵ_i and Z_i are independent it follows that $\mathbb{E}_{\epsilon_i|Z_i=z}(e^{\phi\epsilon_i})$ does not depend on z.

Hence we write

$$\mathbb{E}(Y_i|Z_i = z) = Ce^{\phi\gamma z}$$

where

$$C = e^{\alpha+\phi\beta}\mathbb{E}_{\epsilon_i|Z_i=z}(e^{\phi\epsilon_i}).$$

Therefore

$$\begin{aligned}\log\left[\mathbb{E}(Y_i|Z_i = 1)\right] - \log\left[\mathbb{E}(Y_i|Z_i = 0)\right] &= \log C + \phi\gamma - \log C \\ &= \phi\gamma.\end{aligned}$$

Finally, using assumption **W2**,

$$\begin{aligned}\mathbb{E}(T_i|Z_i = 1) &= \beta + \gamma \\ \mathbb{E}(T_i|Z_i = 0) &= \beta\end{aligned}$$

and hence

$$\frac{\log\left[\mathbb{E}(Y_i|Z_i = 1)\right] - \log\left[\mathbb{E}(Y_i|Z_i = 0)\right]}{\mathbb{E}(T_i|Z_i = 1) - \mathbb{E}(T_i|Z_i = 0)} = \frac{\phi\gamma}{\gamma} = \phi.$$

C Proof: Logistic regression model satisfies Assumption M2

We prove the result that, for a binary treatment, a logistic regression model satisfies Assumption **M2**.

Assumption **M2** implies that

$\log \mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z)$ and $\log \mathbb{E}(Y_i^{(0)}|T_i = t, Z_i = z)$ are both linear in t.

When fitting a logistic regression model, we note that

$\log \mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z) - \log(1 - \mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z))$ and

$\log \mathbb{E}(Y_i^{(0)}|T_i = t, Z_i = z) - \log(1 - \mathbb{E}(Y_i^{(0)}|T_i = t, Z_i = z))$ are linear in t.

Considering the above we may write

$$\log \mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z)) - \log(1 - \mathbb{E}(Y_i^{(1)}|T_i = t, Z_i = z)) = \alpha + \beta t.$$

for some parameters α and β. We deduce that

$$\begin{aligned} \mathbb{E}\left(Y_i^{(1)} \mid T_i = t, Z_i = z\right) &= \frac{\exp(\alpha + \beta t)}{1 + \exp(\alpha + \beta t)} \\ \implies \log\left[\mathbb{E}\left(Y_i^{(1)} \mid T_i = t, Z_i = z\right)\right] &= \alpha + \beta t - \log[1 + \exp(\alpha + \beta t)] \end{aligned}$$

In general, the expression above is not linear in t. However, because $t \in \{0, 1\}$ the expression can be re-written as

$$\log\left[\mathbb{E}\left(Y_i^{(1)} \mid T_i = t, Z_i = z\right)\right] = \begin{cases} \alpha - \log(1 + e^{\alpha}) & t = 0; \\ \alpha + \beta - \log(1 + e^{\alpha+\beta}) & t = 1. \end{cases}$$

we see that

$$\log\left[\mathbb{E}\left(Y_i^{(1)} \mid T_i = t, Z_i = z\right)\right] = \alpha^* + \beta^* t$$

where

$$\begin{aligned} \alpha^* &= \alpha - \log(1 + e^{\alpha}); \\ \beta^* &= \beta - \log\left(\frac{1 + e^{\alpha+\beta}}{1 + e^{\alpha}}\right) \end{aligned}$$

A similar expression can also be derived for $\mathbb{E}(Y_i^{(0)}|Z_i = z, T_i = t)$ and we see that, for a binary treatment, the logistic regression model satisfies Assumption **M2** with respect to the MSMM.

D Derivation of the RDD-RR Estimator

We assume that the bandwidth, h, is chosen such that the subjects included in the data are balanced with regard to confounders.

Using Assumption A3 the threshold indicator Z_i is independent of confounders conditional on X_i, we can obtain unbiased estimates of

$$\lim_{x \to x_0} \mathbb{E}(Y_i | Z_i = z, X_i = x) \text{ for } z \in \{0, 1\}$$

by fitting logistic regression models for Y_i separately for subjects with risk scores above and below the threshold x_0.

For simplicity, we drop $\lim_{x \to x_0}$ and X_i, then

$$\lim_{x \to x_0} \mathbb{E}(Y_i | Z_i = z, X_i = x) \equiv \mathbb{E}(Y_i | Z_i = z) \text{ for } z \in \{0, 1\}$$

at the threshold. Using the law of total probability:

$$\begin{aligned}\mathbb{E}(Y_i | Z_i = z) &= \mathbb{E}(\mathbb{E}(Y_i | Z_i = z) | T_i) \\ &= \mathbb{E}(Y_i | Z_i = z, T_i = 1)\mathbb{P}(T_i = 1 | Z_i = z) + \mathbb{E}(Y_i | Z_i = z, T_i = 0)\mathbb{P}(T_i = 0 | Z_i = z)\end{aligned}$$

Applying Assumption 5 (conditional independence of Y_i and Z_i) we have

$$\mathbb{E}(Y_i | Z_i = 1) = \mathbb{E}(Y_i | T_i = 1)\mathbb{P}(T_i = 1 | Z_i = 1) + \mathbb{E}(Y_i | T_i = 0)\mathbb{P}(T_i = 0 | Z_i = 1) \quad (4)$$

and

$$\mathbb{E}(Y_i | Z_i = 0) = \mathbb{E}(Y_i | T_i = 1)\mathbb{P}(T_i = 1 | Z_i = 0) + \mathbb{E}(Y_i | T_i = 0)\mathbb{P}(T_i = 0 | Z_i = 0) \quad (5)$$

Solving Equations 4 and 5 simultaneously yields

$$\mathbb{E}(Y_i|T_i = 1) = \frac{\mathbb{E}(Y_i|Z_i = 1)\mathbb{P}(T_i = 0|Z_i = 0) - \mathbb{E}(Y_i|Z_i = 0)\mathbb{P}(T_i = 0|Z_i = 1)}{\mathbb{P}(T_i = 1|Z_i = 1) - \mathbb{P}(T_i = 1|Z_i = 0)} \quad (6)$$

and

$$\mathbb{E}(Y_i|T_i = 0) = \frac{\mathbb{E}(Y_i|Z_i = 0)\mathbb{P}(T_i = 1|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 1)\mathbb{P}(T_i = 1|Z_i = 0)}{\mathbb{P}(T_i = 1|Z_i = 1) - \mathbb{P}(T_i = 1|Z_i = 0)} \quad (7)$$

As a result, using Equations 6 and 7, an estimator for the risk ratio is given by

$$\text{RDD-RR} = \frac{\mathbb{E}(Y_i|Z_i = 1)\mathbb{P}(T_i = 0|Z_i = 0) - \mathbb{E}(Y_i|Z_i = 0)\mathbb{P}(T_i = 0|Z_i = 1)}{\mathbb{E}(Y_i|Z_i = 0)\mathbb{P}(T_i = 1|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 1)\mathbb{P}(T_i = 1|Z_i = 0)}.$$

By substituting $\mathbb{P}(T_i = 0|Z_i = z) = 1 - \mathbb{P}(T_i = 1|Z_i = z)$ into the above and re-introducing $\lim_{x \to x_0}$ and X_i, we obtain

$$\text{RDD-RR} = 1 - \lim_{x \to x_0} \frac{\mathbb{E}(Y_i|Z_i = 1, X_i = x) - \mathbb{E}(Y_i|Z_i = 0, X_i = x)}{\mathbb{E}(Y_i|Z_i = 1, X_i = x)\mathbb{E}(T_i|Z_i = 0) - \mathbb{E}(Y_i|Z_i = 0, X_i = x)\mathbb{E}(T_i|Z_i = 1)}.$$