

Regression discontinuity designs for time-to-event outcomes: An approach using accelerated failure time models

Mariam O. Adeleke¹ | Gianluca Baio¹  | Aidan G. O’Keeffe² 

¹University College London, London, UK

²University of Nottingham, Nottingham, UK

Correspondence

Aidan G. O’Keeffe, School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.

Email: aidan.o'keeffe@nottingham.ac.uk

Abstract

Regression discontinuity designs (RDDs) have been developed for the estimation of treatment effects using observational data, where a treatment is administered using an externally defined decision rule, linked to a continuous assignment variable. Typically, RDDs have been applied to situations where the outcome of interest is continuous and non-temporal. Conversely, RDDs for time-to-event outcomes have received less attention, despite such outcomes being common in many applications. We explore RDDs for a time-to-event outcome subject to right censoring. An accelerated failure time (AFT) approach is used to establish a treatment effect estimate for a fuzzy RDD (where treatment is not always strictly applied according to the decision rule). This estimation approach is robust to different levels of fuzziness and unobserved confounding, assessed using simulation studies and compares favourably to established structural AFT models. A motivating example is presented in which models are fitted to estimate the effect of metformin on mortality and cardiovascular disease rate using real observational data from UK Primary Care.

KEYWORDS

accelerated failure time, observational study, regression discontinuity design, time-to-event

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* published by John Wiley & Sons Ltd on behalf of Royal Statistical Society.

1 | INTRODUCTION

Regression discontinuity designs (RDDs) are an approach to the estimation of a treatment effect in observational studies. An RDD can be applied to situations where a treatment or intervention is allocated according to an externally imposed and pre-specified guideline that is linked to the value of a continuous assignment variable and a threshold. The assignment variable is measured for a subject and the subject should receive treatment if the value of their assignment variable is greater (or less) than or equal to the pre-specified threshold value. An RDD exploits the fact that, for subjects with assignment variables close to the threshold, the threshold can be seen as a randomising device, where groups of subjects whose assignment variable values lie ‘just above’ and ‘just below’ the threshold may, under certain assumptions, be seen as ‘exchangeable’ and balanced with regard to confounding variables.

As an example, which shall be explored in this paper, in the United Kingdom it has been recommended that metformin, a medicine used to reduce blood sugar level, be prescribed to patients who are at risk of diabetes whose haemoglobin A1c (HbA1c) level is at or above the level of 48 mmol/mol. A high level of HbA1c can be an indicator of uncontrolled diabetes which may lead to serious health complications or death if not treated. In this example, the ‘treatment’ is a prescription of metformin, the ‘assignment variable’ is a patient’s HbA1c level and the treatment ‘threshold’ is an HbA1c level of 48 mmol/mol or above. An effect that we might wish to estimate is that of metformin on all-cause mortality or perhaps on the occurrence of particular complications.

In general, a treatment guideline is not always adhered to strictly. As such, there may be some subjects whose assignment variable value lies below the threshold but who receive treatment (and vice-versa). In this case, the design is referred to as a fuzzy RDD. Otherwise, where there is strict adherence to the guideline, the design is known as a sharp RDD. In this work, we focus on fuzzy RDD designs, which are common in medical studies and for which treatment effect estimation may be more challenging in general. In the metformin example, it may be that some patients are counter-indicated for metformin, perhaps owing to side effects or co-morbidities. It may be that some patients would prefer to try other measures to reduce their blood glucose level, such as changing their diet or lifestyle choices, or that patients who are ‘just below’ the 48 mmol/mol HbA1c level are prescribed metformin as a precautionary measure.

The RDD was first introduced by Thistlethwaite and Campbell (1960) and has since been used widely in economics (Angrist et al., 1996; Hahn et al., 2001; Imbens & Angrist, 1994; Imbens & Kalyanaraman, 2012; Xu, 2017) and political science (Caughey & Sekhon, 2011; Erikson et al., 2015). Despite the obvious analogy between an RDD and a randomised controlled trial (with the threshold seen as a randomising device), RDDs have not been used extensively in medical studies until relatively recently. Most applications of the RDD to medicine have focused on studies where the outcome of interest has a continuous distribution, such as a normal distribution (see, e.g. Bor et al., 2014; Geneletti et al., 2015; Law et al., 2017; Linden et al., 2006; O’Keeffe et al., 2014). A few more recent applications of RDDs when the outcome is binary has also been explored (Bor et al., 2017; Geneletti et al., 2019; van Leeuwen et al., 2018). Time-to-event outcomes are important and ubiquitous measures in many medical studies. Despite this, the use of an RDD with a time-to-event or survival outcome has received little attention. Bor et al. (2014) fitted a model that used a log hazard ratio to estimate the effect of antiretroviral drugs on mortality in HIV patients, where antiretrovirals were prescribed according to a decision rule linked to a patient’s CD4 count. A scaling method was used for the model fit but this was only valid where the event of interest is rare and the follow-up time short.

In this paper, we explore approaches for the estimation of a treatment effect in an RDD where the outcome is a time-to-event. In particular, we focus on treatment effect estimation under the accelerated failure time (AFT) assumption, an alternative to the popular proportional hazard (PH) assumption for modelling time-to-event outcomes. An attractive property of the AFT assumption is that the treatment effect estimate can be interpreted directly in terms of the time-to-event and the AFT assumption is also more compatible with assumptions required for an RDD. An estimator is derived for an AFT treatment effect in an RDD, using the assumptions made for an RDD. In addition, we also demonstrate that a spline-based approach may be suitable for situations where the assignment variable–outcome relationship is non-linear around the threshold. These estimation approaches are compared to that from a structural AFT (S-AFT) model, commonly used in causal inference for time-to-event outcomes in observational studies (Clarke & Windmeijer, 2010; Didelez et al., 2010; Hernán & Robins, 2006). Using simulation studies, the RDD estimator is desirable in that it is less sensitive to confounding. An important assumption of the S-AFT model is that all potential confounders are observed or that unobserved confounders can be explained by the observed confounders. This is a reasonably strong assumption and, since it is untestable, the corresponding estimate might be misleading when the assumption is violated. The RDD approach does not include such a stringent assumption, particularly when the treatment guideline is adhered to.

This paper is organised as follows: in Section 2, the RDD and its assumptions are outlined. In Section 3, the RDD-AFT and S-AFT estimators are introduced and described. A simulation study is carried out in Section 4 to compare estimation approaches. The example on the prescription of metformin on mortality and cardiovascular disease is presented in Section 5 and a discussion and conclusions are provided in Section 6.

2 | THE REGRESSION DISCONTINUITY DESIGN

In this paper, motivated by the modelling of the effect of metformin on all-cause mortality for pre-diabetic patients (patients at risk of type 2 diabetes), we develop an RDD for a time-to-event outcome. We now describe the notation and the modelling assumptions for an RDD with a time-to-event outcome in detail.

2.1 | RDD: variables and assumptions

We define the following variables that will be used in the RDD. The index i denotes the i^{th} subject for the set of subjects $\{1, \dots, N\}$. In this work, we consider only right censoring for the time-to-event.

- T_i^* is the true time-to-event and C_i is the time at which right censoring occurs. In a given data set, the observed ‘time-to-event’ is denoted $T_i = \min\{T_i^*, C_i\}$.
- Since either T_i^* or C_i is observed (and not both), we define the event indicator $\delta_i = \mathbb{1}\{T_i^* < C_i\}$ which takes the value 1 if the i^{th} subject’s time-to-event is observed and 0 if the time-to-event is right censored.
- X_i is the continuous assignment variable and x_0 is the treatment threshold, such that the i^{th} subject should receive treatment if $X_i \geq x_0$.

- $Z_i = \mathbb{1}\{X_i \geq x_0\}$ is threshold indicator, which is equal to 1 (0) if the i^{th} subject's assignment variable lies above (below) the threshold.
- A_i is a treatment indicator, such that $A_i = 1$ if the i^{th} subject receives treatment and 0 otherwise.
- $C_i = \mathcal{O}_i \cup \mathcal{U}_i$ is a set of confounding variables both observed (\mathcal{O}_i) and unobserved (\mathcal{U}_i).

As noted previously, an RDD considers subjects whose assignment variable values lie 'just above' the threshold to be similar to those whose assignment variable values lie 'just below' the threshold. To quantify this notion, we define a bandwidth, h , such that only those subjects whose assignment variable values lie within the set $[x_0 - h, x_0 + h]$ are included in the RDD and used to estimate the treatment effect at the threshold. Some works have discussed bandwidth selection and, in particular, choosing the 'optimal' bandwidth for an RDD (Calonico et al., 2014; Imbens & Kalyanaraman, 2012; Ricciardi et al., 2020); however, in this paper, we refrain from a focus on an optimal bandwidth. In a medical context, it is likely that clinicians would have input into how similar subjects would be with regard to their assignment variable values, particularly where these values are clinical indicators or perhaps risk scores. As such, we shall consider RDDs for a variety of bandwidths and investigate the sensitivity of results to bandwidth choice in a pragmatic manner, similar to the approach taken in other papers (for example Geneletti et al., 2015; Geneletti et al., 2019; Li et al., 2015).

In an RDD, inference is typically focused on treatment effect estimation at the threshold. To identify an effect at the threshold, we make a number of assumptions that should hold. In this paper, we use a conditional independence framework (Dawid, 1979; Didelez et al., 2010) to outline these and details of these assumptions are provided in Appendix A. We now consider approaches for a time-to-event (survival) outcome.

3 | RDD FOR A TIME-TO-EVENT

We use an AFT assumption when comparing groups with a time-to-event outcome in an RDD. First, we outline the general AFT model for groups and will later adapt this for use in an RDD.

3.1 | Accelerated failure time assumption

An AFT model assumes that the time-to-event in the 'treated' group is accelerated (or decelerated) compared to that in the 'untreated' group. Formally, the survivor functions for the treated and untreated in a general AFT model may be expressed as

$$S(\phi t | A_i = 1) = S(t | A_i = 0).$$

Here ϕ is termed an 'acceleration factor' and it is assumed that, in general, the times taken to attain the same survival probability in the treated and untreated groups are ϕt and t respectively. As such, ϕ is a measure of the difference in the time-to-event between the two groups and, clearly, $\phi = 1$ would suggest no general difference. Typically, a log-linear form may be used where an AFT assumption is made, with a general model for the time-to-event, T_i , given by

$$\log T_i = \beta A_i + \psi_i. \quad (1)$$

Here ψ_i denotes the (natural) log event time for a subject who does not receive treatment ($A_i = 0$) and the acceleration factor is given by $\phi = \exp(\beta)$. We note that other covariates could be easily added into this model if desired. If the probability distribution of ψ_i is specified, the parameter β may be estimated parametrically. Alternatively, a semi-parametric AFT model may be used without specifying the distribution of ψ_i (Buckley & James, 1979; Louis, 1981; Wei, 1992). For a randomised controlled trial without non-compliance, or a sharp RDD with no unobserved confounding, an estimate of ϕ reflects the treatment effect at the threshold (subject to the validity of the AFT model and any distributional assumptions). However, in a fuzzy RDD (essentially similar to a randomised controlled trial with non-compliance), estimation of the treatment effect at the threshold is less straightforward.

Proportional hazards (PH) models are used commonly for inference in analyses involving time-to-event outcomes. However, PH models may have a limited interpretation when estimating causal effects, with or without confounding (Aalen et al., 2015; Hernán, 2010; Lange & Hansen, 2011; Martinussen et al., 2020; VanderWeele, 2011). In a PH model, marginal and conditional hazard ratio estimates may differ considerably and give different causal interpretations, owing to the non-additive form of the hazard function with respect to the intervention and potential confounding variables. This may be problematic and lead to the misinterpretation of a treatment effect (Janes et al., 2010; Martinussen & Vansteelandt, 2013; Sjölander et al., 2016; Sutradhar & Austin, 2018). Furthermore, for a Cox PH model, the construction of the likelihood function using risk sets implies that, if the risk is different for different levels of confounding variables, the independence of individuals who have not experienced the event of interest by a given time—but have different levels of confounding variables—may not hold if the confounding variables are not included in the model or are unknown. In contrast, the AFT model has been used for causal inference (Huling et al., 2019; Robins, 1992; Robins & Tsiatis, 1991). In particular, the Weibull AFT model, used in this paper, can be easily formulated as a location-scale model which yields a treatment effect estimate that is collapsible and can be interpreted causally. Furthermore, the AFT model is used to describe and compare the survivor experience of group(s) using the probability of survival, which may be more intuitive than comparing survivor experience using hazards.

3.2 | RDD-AFT estimator

We now consider an estimator for the acceleration factor at the threshold based on the RDD assumptions outlined in Appendix A. In RDDs where the outcome is continuous and not a time-to-event, the local average treatment effect (LATE) has been used extensively for effect estimation at the threshold (Hahn et al., 2001; Imbens & Lemieux, 2008). Essentially, the LATE involves fitting linear models for the outcome above and below the threshold to estimate the ‘jump’ in the outcome, using only data where the assignment variable, x_i lies in the range $[x_i - h, x_i + h]$ for a given bandwidth h . The fuzziness present in an RDD is accounted for by estimating the proportion of subjects who receive treatment above and below the threshold and then scaling the estimate of the ‘jump’ by the difference in these proportions. That is, the LATE at the threshold is given by

$$\text{LATE} = \frac{\alpha}{\pi_a - \pi_b}$$

where α is the difference in means for the linear models above and below the threshold and $\pi_a - \pi_b$ is the difference in the probability of receiving treatment above (a) and below (b) the threshold.

We note that the AFT model specified in Equation (1) is a linear model and methods used to estimate the LATE in a standard RDD may be adapted for use in an AFT, as we now outline. A model that estimates the discontinuity in the time-to-event at the threshold may be written

$$\log T_i = \alpha_1 Z_i + \alpha_2 X_i + \alpha_3 X_i Z_i + \psi_i. \quad (2)$$

The interaction term between the assignment variable and the threshold indicator may be dropped if it is assumed that slopes (on the log-scale) are the same above and below the threshold by setting $\alpha_3 = 0$. The discontinuity or ‘jump’ at the threshold is measured by α_1 . As mentioned previously, if the distribution of ψ_i is specified then model parameters may be estimated using maximum likelihood and accounting for right censoring, where appropriate, and these models may be fitted easily using standard statistical software (e.g. using the `survival` package in R (Therneau, 2021)).

In this paper, we shall assume a Weibull distribution for ψ_i and use maximum likelihood methods for parameter estimation. In practice, other distributions such as a log-normal or log-logistic distribution could be fitted and an optimality criteria based on the model residuals can be used to select the distribution that best fits a data set (Chan et al., 2018).

In a sharp RDD, the estimate of α_1 should be unbiased for the treatment effect at the threshold on the log-scale (assuming, of course, that the RDD assumptions are satisfied and that the model is correctly specified). In a fuzzy RDD, direct estimation of α_1 by fitting the model (2) would yield a biased estimate of the treatment effect at the threshold, owing to the fact that Z_i does not necessarily distinguish the treated from the untreated. As such, in a fuzzy RDD, the estimate of the treatment effect at the threshold on the log-scale may be given by

$$\beta_{\text{RDD}} = \frac{\alpha_1}{\pi_a - \pi_b}.$$

Here π_a and π_b denote the probability of treatment for subjects with assignment variable values above and below the threshold, respectively. A proof of this result is provided in Appendix B. The parameters π_a and π_b may be estimated using logistic regression or another appropriate method. We now consider a structural AFT model as a method for treatment effect estimation.

3.3 | Structural AFT model

Structural models have been considered widely for causal effect estimation in observational or non-randomised studies (Clarke & Windmeijer, 2010; Didelez et al., 2010; Hernán & Robins, 2006; Robins et al., 2000; VanderWeele, 2009; Vansteelandt & Goetghebeur, 2003). In particular, the structural AFT (S-AFT) model has been used for treatment effect estimation where the outcome is a time-to-event (Hernán et al., 2005). The S-AFT model may be defined easily using a potential outcomes framework which we now outline. We define a as an indicator such that $a = 1$ implies that treatment is assigned and $a = 0$ that treatment is not assigned. The potential outcomes are $T_i(1)$ (where $a = 1$) and $T_i(0)$ (where $a = 0$) and these represent the time-to-event for the i^{th} subject where the subject is treated and not treated, respectively. The treatment effect at the subject level would be ascertained by comparing $T_i(1)$ and $T_i(0)$. However, since these cannot both be observed, the treatment effect is typically estimated using G-estimation, which involves exploiting the conditional independence between the potential outcomes and the treatment indicator. Using the AFT assumption, it follows that

$$T_i(1) = \gamma T_i(0)$$

where γ is the acceleration factor. For the S-AFT model to be valid, the following assumptions must hold:

S1 There are no unobserved confounders and treatment allocation is strongly ignorable conditional on observed confounders. This assumption may be written formally as

$$T_i(1), T_i(0) \perp\!\!\!\perp A_i \mid \mathcal{O}_i.$$

In words, this assumption implies that the time-to-event outcome for the i^{th} subject, whether treated or not does not depend on whether or not the subject has actually received treatment (observed as treated).

S2 The probability of treatment is non-zero, conditional on observed variables. That is

$$\mathbb{P}(A_i = 1 \mid \mathcal{O}_i) > 0 \quad \text{for all } i.$$

The S-AFT model is not strictly an RDD approach. Instead, the S-AFT method compares the treated and untreated groups while correcting for effect of confounding using the observed confounders and relying on the assumptions presented above. In contrast, the RDD-AFT method compares the patients above and below the threshold, using the threshold as a quasi-randomising device, and a treatment effect at the threshold is estimated by adjusting the result of this comparison with the probability of compliance to the treatment guideline. Considering the estimates made by the RDD-AFT and S-AFT models (the RDD-AFT model compares outcomes for levels of Z and the S-AFT model compares outcomes for levels of A), we note that assumption **S1** is similar to **A5** in the RDD assumptions (Appendix A). In addition, assumption **S2** is similar to **A6**. The S-AFT model would be applicable for an RDD, assuming that there is no unmeasured confounding in the region around the threshold or, alternatively, if any confounding variables present in a region around the threshold were accounted-for in the S-AFT model. Where these assumptions are satisfied, the acceleration factor γ (i.e. the treatment effect at the threshold) may be estimated using G-estimation (Hernán et al., 2005) G-estimation is well known as a method that may overcome bias from time-varying confounders when estimating a treatment effect and has been used with AFT models. It is known that estimation problems may sometimes be encountered when fitting S-AFT models using G-estimation, owing to the ‘artificial censoring’ procedure used in the model fitting process (Joffe et al., 2012). An overview of the G-estimation procedure used for the fitting of the S-AFT model is provided in Appendix C.

We now aim to compare the S-AFT and RDD-AFT approaches using a variety of simulated scenarios, before a comparison and application using a real data set on the prescription of metformin and mortality in UK primary care.

4 | SIMULATION STUDY

We consider some scenarios using simulated data in which an RDD would apply, with different levels of confounding. In doing so, our aim is to explore and evaluate the use of the S-AFT and RDD-AFT models for a time-to-event outcome in a fuzzy RDD. The simulation algorithm and set-up is outlined below. First we shall consider a simple linear relationship between the

assignment variable and the outcome of interest. We shall then consider a simulation scenario with a non-linear assignment variable–outcome variable relationship where flexible functions of the assignment variable are considered in the model.

4.1 | Simulation algorithm and set-up

We describe the simulation algorithm where M data sets, each of size N , are simulated ($M \in \mathbb{N}$, $N \in \mathbb{N}$). As noted previously, the index i denotes a subject with $i \in \{1, \dots, N\}$.

Step 1: For each subject, an assignment variable, X_i , is simulated from a continuous uniform distribution.

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

Step 2: The threshold is set to be equal to 0.5 and we define the centred assignment variable, X_i^C , and threshold indicators, Z_i , as

$$\begin{aligned} X_i^C &= X_i - 0.5; \\ Z_i &= \mathbb{1}\{X_i^C \geq 0\}. \end{aligned}$$

Step 3: A confounding variable, W_i , is simulated from a standard normal distribution

$$W_i \sim \mathcal{N}(0, 1).$$

Step 4: The probability that the i^{th} subject receives treatment, p_i , is given by

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Z_i + \beta_2 X_i^C + \beta_3 W_i. \quad (3)$$

The parameters of this model are specified to reflect the simulated fuzziness of the design and the level of confounding with regard to treatment allocation.

Step 5: The treatment indicator, A_i , is simulated as follows

$$A_i \sim \text{Bernoulli}(p_i)$$

Step 6: The ‘true’ time-to-event is simulated as

$$\log T_i^* = \beta_4 A_i + \beta_5 W_i + \psi_i. \quad (4)$$

Here, β_4 defines the treatment effect (on the log-scale), β_5 reflects the correlation between the time-to-event outcome and the confounder W_i and ψ_i defines the probability distribution for $\log T_i^*$.

Step 7: To incorporate right censoring, a censoring time C_i is simulated. A censoring probability, p_c , and an ‘end-of-study’ time K are specified. A random variable is specified $U_i^C \sim \text{Uniform}(0, 1)$ and C_i is defined as follows

$$C_i = \begin{cases} K & \text{if } U_i^C > p_c \\ \sim \text{Uniform}(0, K) & \text{if } U_i^C \leq p_c \end{cases}$$

Step 8: The observed time-to-event and event indicators are defined as follows

$$T_i = \min(T_i^*, C_i)$$

$$\delta_i = \mathbb{1}\{T_i^* < C_i\}$$

Step 9: Steps 1–8 are repeated N times, to create a data set with N subjects.

Step 10: Steps 1–9 are repeated M times and M data sets, each with N subjects, are obtained.

In this paper, we set $\beta_0 = -2$ and $\beta_2 = 2$. We set $\beta_4 = \log 1.5$ so that the acceleration factor is 1.5. We choose the median survival time in the untreated group to be 7 years and, as such, we set

$$\exp(\psi_i) \sim \text{Weibull}\left(\frac{\log(2)}{7^2}, 2\right).$$

For the censoring mechanism, we set the maximum follow-up time as $K = 10$ and the probability of drop-out (right censoring) before 10 years as $p_c = 0.15$. As noted in the simulation algorithm, β_1 denotes the strength of the relationship between the threshold indicator and the probability of treatment (i.e. the fuzziness of the RDD). In addition, β_3 and β_5 reflect levels of unobserved confounding with treatment and the time-to-event, respectively. We varied these parameters to adjust the level of fuzziness and level of confounding, resulting in a range of simulated data sets using which the methods outlined in Section 3 may be compared. The type of simulation scenarios, to which we refer in our discussion of simulation results, are outlined with parameter choices in Table 1.

No confounding implies that there is no linear correlation between the confounding variable W_i and both the time-to-event T_i and treatment allocation A_i . The chosen values of β_3 and β_5 result in the Pearson correlation coefficients shown in the columns $\rho_{T,W}$ and $\rho_{A,W}$ which allows an assessment of the linear correlation level between variable pairs and, hence, the level of confounding.

We measure how well the treatment guideline is adhered to using the level of fuzziness of the design. Where there is a high level of compliance to the treatment guideline, the fuzziness is weak and a low compliance to the treatment guideline implies that fuzziness is strong. This is shown in the column P.C. of Table 1. In the case of weak fuzziness, about 80% to 90% of the patients complied to the treatment guideline, while the probability of compliance to the treatment guideline is about 55% where fuzziness is strong. We chose $N = 2000$ and $M = 1000$. Standard errors were estimated using bootstrapping, with a bootstrap sample size of 1000.

TABLE 1 Values of parameters in Equations (3) and (5) for the simulation scenarios with the corresponding probability of compliance (P.C.) and estimates of correlation coefficients between T and W ($\rho_{T,W}$) and A and W ($\rho_{A,W}$)

Scenario	Parameters			P.C.	$\rho_{T,W}$	$\rho_{A,W}$
Weak fuzziness, No Confounding	$\beta_1 = 10$	$\beta_3 = 0$	$\beta_5 = 0$	0.90	0.00	0.00
Weak fuzziness, Low Confounding	$\beta_1 = 10$	$\beta_3 = -1$	$\beta_5 = 0.3$	0.87	0.16	-0.10
Weak fuzziness, High Confounding	$\beta_1 = 10$	$\beta_3 = 2$	$\beta_5 = 0.7$	0.79	0.40	0.22
Strong fuzziness, No Confounding	$\beta_1 = 2.5$	$\beta_3 = 0$	$\beta_5 = 0$	0.57	0.00	0.00
Strong fuzziness, Low Confounding	$\beta_1 = 2.5$	$\beta_3 = -0.5$	$\beta_5 = 0.3$	0.55	0.16	-0.16
Strong fuzziness, High Confounding	$\beta_1 = 2.5$	$\beta_3 = 0.7$	$\beta_5 = 0.7$	0.54	0.41	0.22

4.2 | Simulation study—results

Simulations were carried out using the chosen parameter values stated in Section 4.1. We chose RDD bandwidths of 0.05, 0.1, 0.15 and 0.2 and fitted both the S-AFT and RDD-AFT models to estimate the treatment effect at the threshold. We note that the use of the bandwidths results in a reduction in sample size from $N = 2000$, with smaller bandwidths using smaller sample sizes. The probability of censoring (either by drop-out or by not experiencing the event before the simulated end point of 10 years) was approximately 0.47 for each simulated data set.

Figure 1 shows box plots that summarise the acceleration factor estimates for each bandwidth, estimation approach and confounding level where the level of fuzziness is weak. As would be expected, when there is no confounding both methods yield unbiased estimates of the acceleration factor across all bandwidths. Where there is a low level of unobserved confounding, the S-AFT method is generally biased for the acceleration factor whereas the RDD-AFT method is not, with the median RDD-AFT estimate lying close to the true acceleration factor value for all bandwidths. Similarly, for the scenario where the level of confounding is high, the RDD-AFT estimate is more desirable than the S-AFT estimate, generally producing an unbiased estimate of the acceleration factor for all bandwidths. The difference between the S-AFT and RDD-AFT estimates would be expected for scenarios where confounding occurs, because confounding variables should be accounted-for for the S-AFT model to be valid. To check the S-AFT model further, we repeated the simulation study but conditioned on confounders and found that the S-AFT estimates were less biased, as would be expected, shown by the boxplots for the adjusted S-AFT estimates in Figures 1 and 2. Numerical summaries of these adjusted S-AFT estimates (and all other simulations) are shown in Table 2.

Figure 2 shows box plots that summarise the acceleration factor estimates for each bandwidth, estimation approach and confounding level where the level of fuzziness is strong. Both methods produce unbiased estimates of the acceleration factor where there is no confounding. For both the low and high unobserved confounding levels, the S-AFT model yields a biased estimate of the acceleration factor. In contrast, the RDD-AFT model is less biased and generally provides a more accurate estimate of the acceleration factor (treatment effect).

In the simulation study, the time-to-event was drawn from a Weibull distribution, which has a monotonic hazard function. To investigate the sensitivity of the approach to the distributional assumption, we simulated data from a log-logistic distribution that has a non-monotonic hazard function and fitted the Weibull RDD-AFT model to the simulated data sets. Results are shown in Appendix D and these show some bias in the RDD-AFT estimates, which tend to underestimate the treatment effect at the threshold. However, the bias is smaller as the bandwidth reduces. From the simulation study, we can conclude that a smaller bandwidth not only reduces the effect of unobserved confounding, it may also reduce the effect of model mis-specification, subject to the validity of the RDD assumptions.

4.3 | Simulation study—flexible models

We consider a scenario where the relationship between the assignment variable and outcome is non-linear and a flexible model is used. The simulation algorithm is similar to that given in Section 4.1, but Step 6. is replaced with the step outlined below, to allow a non-linear relationship between the ‘true’ time-to-event and the assignment variable.

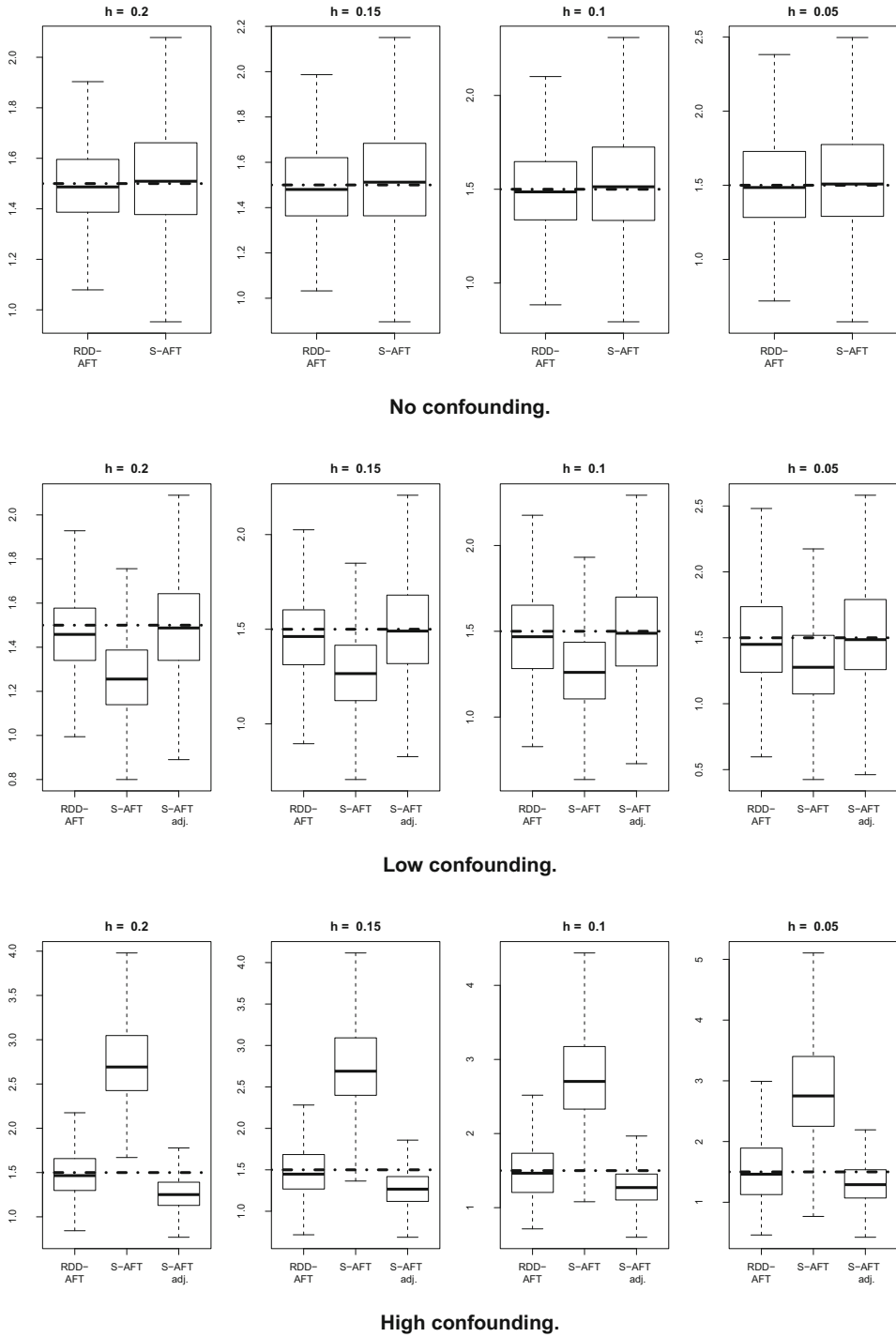


FIGURE 1 Boxplots showing the acceleration factor estimates from the simulation studies to compare performance of the RDD-AFT and S-AFT approaches for designs with weak fuzziness. The true value of the acceleration factor is 1.5 (dashed horizontal line). The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

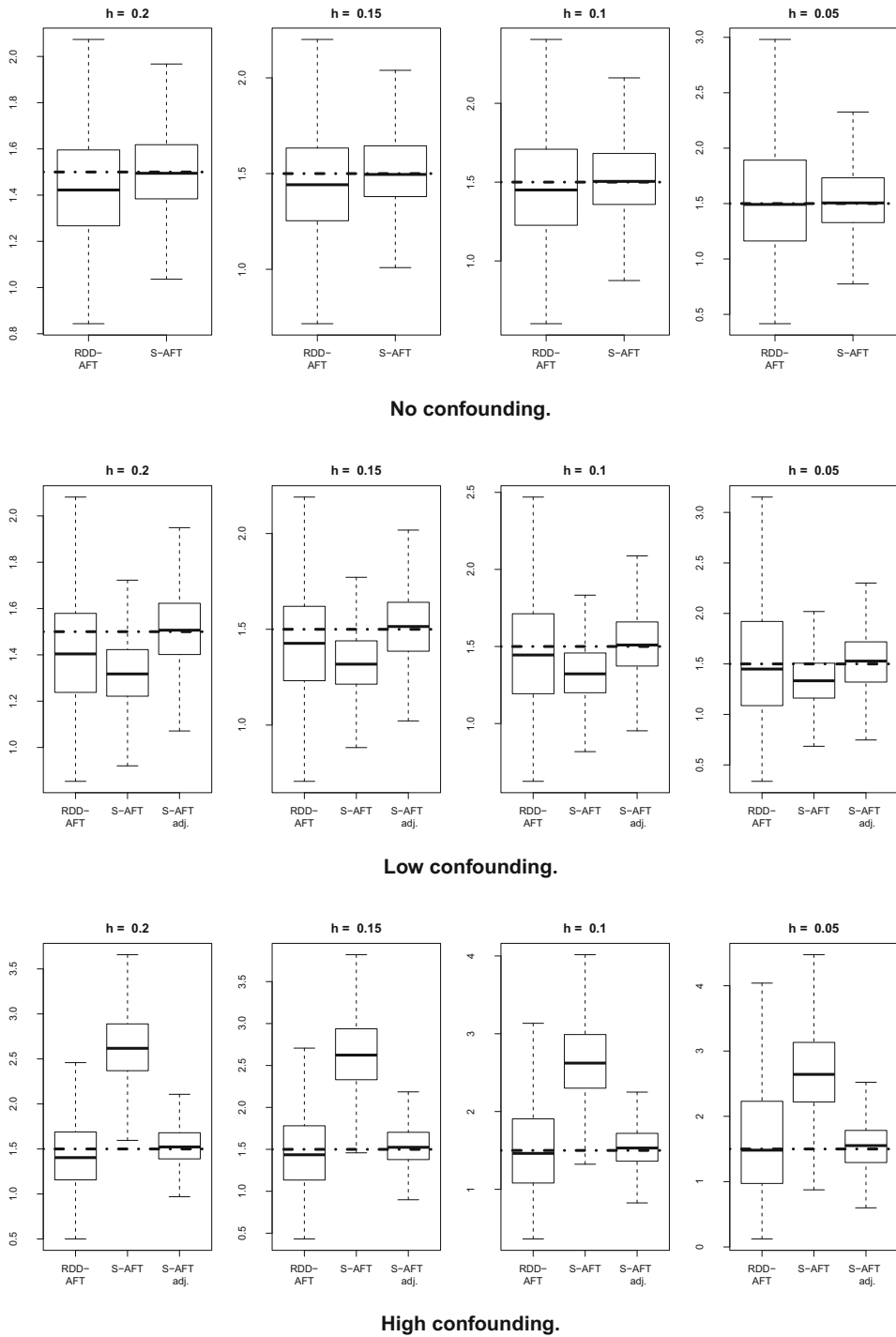


FIGURE 2 Boxplots showing the acceleration factor estimates from the simulation studies to compare performance of the RDD-AFT and S-AFT approaches for designs with strong fuzziness. The true value of the acceleration factor is 1.5 (dashed horizontal line). The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

TABLE 2 Estimates, biases, empirical standard errors (ESE), average standard errors (ASE) and 95% coverage for the log of the acceleration factor from simulation studies to compare performance of the RDD-AFT and S-AFT approaches. S-AFT denotes the S-AFT model not adjusted for confounders, S-AFT adj. denotes the S-AFT model adjusted for confounders. The true value of the log of the acceleration factor is $\log(1.5) = 0.405$. The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

Method	Weak fuzziness					Strong fuzziness				
	Estimate	Bias	ESE	ASE	95% Coverage	Estimate	Bias	ESE	ASE	95% Coverage
No confounding										
Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.40	0.01	0.11	0.11	94.0	0.35	0.05	0.17	0.17	93.3
S-AFT	0.41	-0.01	0.19	0.15	94.9	0.40	0.00	0.14	0.11	94.4
Bandwidth = 0.15, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.40	0.01	0.13	0.13	95.8	0.36	0.04	0.20	0.20	94.7
S-AFT	0.42	-0.01	0.19	0.16	95.9	0.41	0.00	0.16	0.13	94.5
Bandwidth = 0.1, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.40	0.01	0.16	0.16	95.9	0.37	0.04	0.25	0.26	95.1
S-AFT	0.42	-0.01	0.23	0.19	95.4	0.41	-0.01	0.19	0.15	93.6
Bandwidth = 0.05, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.40	0.01	0.23	0.24	96.8	0.40	0.01	0.39	0.40	95.2
S-AFT	0.42	-0.01	0.31	0.26	96.0	0.42	-0.02	0.25	0.19	94.6
Low confounding										
Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.03	0.12	0.12	94.5	0.34	0.07	0.18	0.19	94.7
S-AFT	0.23	0.17	0.18	0.15	71.8	0.27	0.13	0.15	0.11	75.0
S-AFT adj.	0.40	0.01	0.18	0.15	94.9	0.40	0.00	0.15	0.11	94.4
Bandwidth = 0.15, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.03	0.14	0.14	95.5	0.35	0.06	0.22	0.22	94.5
S-AFT	0.23	0.17	0.20	0.16	74.3	0.27	0.13	0.16	0.13	78.9
S-AFT adj.	0.40	0.01	0.21	0.16	95.6	0.40	0.00	0.16	0.13	94.3
Bandwidth = 0.1, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.03	0.18	0.18	95.9	0.37	0.04	0.28	0.29	95.6
S-AFT	0.23	0.17	0.24	0.19	80.6	0.27	0.13	0.19	0.15	83.5
S-AFT adj.	0.40	0.01	0.25	0.19	95.0	0.40	0.00	0.19	0.15	94.9
Bandwidth = 0.05, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.02	0.26	0.26	96.0	0.38	0.02	0.42	0.44	97.1
S-AFT	0.24	0.16	0.32	0.25	85.9	0.27	0.13	0.25	0.20	87.9
S-AFT adj.	0.41	0.00	0.32	0.25	94.1	0.41	0.00	0.25	0.19	95.2

(Continues)

TABLE 2 (Continued)

Method	Weak fuzziness					Strong fuzziness				
	Estimate	Bias	ESE	ASE	95%	Estimate	Bias	ESE	ASE	95%
					Coverage					Coverage
High confounding										
Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.02	0.18	0.18	95.1	0.34	0.06	0.28	0.28	95.1
S-AFT	1.01	-0.60	0.19	0.16	6.3	0.97	-0.56	0.18	0.15	6.9
S-AFT adj.	0.22	0.18	0.18	0.15	74.0	0.42	-0.02	0.17	0.13	93.0
Bandwidth = 0.15, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.02	0.21	0.21	95.4	0.35	0.05	0.34	0.34	94.7
S-AFT	1.01	-0.60	0.22	0.18	9.7	0.97	-0.56	0.20	0.16	11.0
S-AFT adj.	0.23	0.18	0.21	0.17	78.2	0.42	-0.02	0.19	0.15	92.0
Bandwidth = 0.1, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.03	0.26	0.26	95.5	0.37	0.04	0.45	0.44	94.9
S-AFT	1.00	-0.60	0.26	0.21	18.0	0.97	-0.56	0.22	0.19	16.7
S-AFT adj.	0.23	0.17	0.24	0.20	82.7	0.42	-0.01	0.21	0.17	93.4
Bandwidth = 0.05, Treatment effect = 0.405, Sample size = 2000										
RDD-AFT	0.38	0.02	0.38	0.38	95.5	0.38	0.02	0.67	0.68	95.5
S-AFT	1.02	-0.61	0.32	0.28	37.8	0.98	-0.57	0.30	0.25	31.8
S-AFT adj.	0.24	0.16	0.30	0.25	87.9	0.43	-0.02	0.29	0.23	92.4

Step 6 The ‘true’ time-to-event is simulated as

$$\log T_i^* = f(X_i^c, A_i) + \beta_5 W_i + \psi_i, \tag{5}$$

where

$$f(X_i^c, A_i) = \begin{cases} \frac{1}{240} \{ (4 - 20X_i^c)(1 - 40X_i^c)(2 - 200X_i^c) - 8 \} + \beta_4 & \text{if } A_i = 1 \\ -1.5 - 50 \{ (X_i^c)^2 - 0.1 \} \{ (X_i^c)^2 + 0.3 \} & \text{if } A_i = 0 \end{cases}$$

Here, β_4 defines the treatment effect (on the log-scale), β_5 reflects the correlation between the time-to-event outcome and the confounder W_i and ψ_i defines the probability distribution for $\log T_i^*$. Figure 3 shows a plot the functional relationship between the assignment variable and the outcome on the log-scale.

To fit a flexible model, we use a natural cubic spline (Hastie & Tibshirani, 1999) to model the relationship between the outcome and assignment variable (i.e. the numerator of the log (RDD-AFT) estimator). The number of knots in the spline (V) varied for different bandwidth sizes as follows: for $h = 0.2$, $V = 8$; for $h = 0.15$, $V = 6$; for $h = 0.1$, $V = 4$ and for $h = 0.05$, $V = 2$.

Table 3 shows numerical summaries of acceleration factor estimates at the threshold where data were simulated from the non-linear scenario for all levels of confounding and where

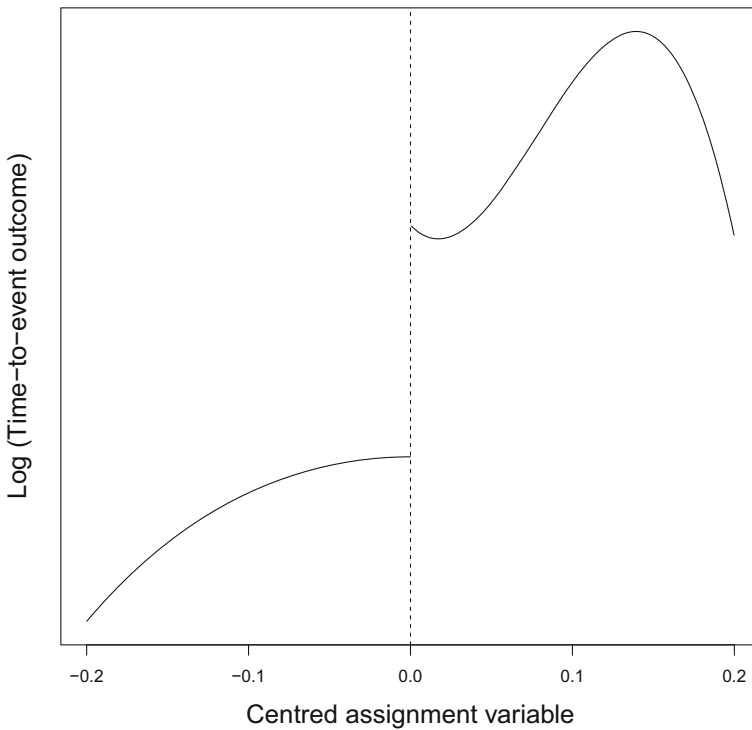


FIGURE 3 Plot showing the functional form of the non-linear relationship between the assignment variable and the outcome for the simulation study considered in Section 4.3

fuzziness is weak or strong. Figure 4 shows boxplots that summarise acceleration factor estimates for each bandwidth and where the fuzziness level is weak. Figure 5 shows similar boxplots but where the level of fuzziness is strong.

Examining these results and summaries, across all scenarios, the linear RDD-AFT approach does not accurately estimate the true acceleration factor as we should expect. In contrast, the spline-based RDD-AFT model appears to provide an unbiased estimate of the true acceleration factor across all bandwidths and confounding levels, where fuzziness is weak. Performance of the spline-based model is also reasonable where fuzziness is strong. However, across all scenarios, we note that the variability in the spline-based estimates is higher than that for the linear RDD-AFT estimates. We might expect the variability to be larger with a spline-based model owing to the possibility of a bias–variance trade-off commonly encountered when fitting splines. We note that the linear RDD-AFT method improves (in terms of a reduction in bias) as the bandwidth decreases. This would be expected as the functional form of the assignment variable–outcome variable relationship varies less over a smaller bandwidth. The linear S-AFT model performs reasonably well where the bandwidth is small and confounding is low. However, where confounding is high, the S-AFT model does not estimate the true acceleration factor accurately.

5 | EXAMPLE: METFORMIN IN UK PRIMARY CARE

In the United Kingdom, the National Institute for Health and Care Excellence (NICE) has recommended that metformin, a drug that reduces blood sugar level, is prescribed to adults at risk

TABLE 3 Estimates, biases, empirical standard errors (ESE), average standard errors (ASE) and 95% coverage for the log of the acceleration factor from simulation studies for the RDD-AFT estimators estimated using the previous approaches and a flexible model (Spline RDD-AFT). The true value of the log of the acceleration factor is $\log(1.5) = 0.405$. The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

Method	Weak fuzziness				Strong fuzziness				
	Estimate	Bias	ESE	ASE	Estimate	Bias	ESE	ASE	
No confounding									
Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000									
RDD-AFT	0.28	0.12	0.13	0.13	0.23	0.18	0.19	0.19	83.6
Spline RDD-AFT	0.41	0.00	0.27	0.28	0.36	0.05	0.43	0.43	95.3
S-AFT	0.83	-0.43	0.19	0.16	0.87	-0.46	0.16	0.13	7.5
Bandwidth = 0.15, Treatment effect = 0.405									
RDD-AFT	0.20	0.21	0.14	0.14	0.17	0.23	0.22	0.21	81.0
Spline RDD-AFT	0.41	0.00	0.28	0.29	0.35	0.05	0.45	0.45	95.7
S-AFT	0.78	-0.37	0.21	0.17	0.75	-0.34	0.17	0.14	27.5
Bandwidth = 0.1, Treatment effect = 0.405									
RDD-AFT	0.26	0.15	0.16	0.16	0.23	0.17	0.26	0.26	91.0
Spline RDD-AFT	0.41	0.00	0.29	0.30	0.36	0.04	0.49	0.49	96.1
S-AFT	0.70	-0.29	0.24	0.20	0.61	-0.20	0.18	0.15	72.3
Bandwidth = 0.05, Treatment effect = 0.405									
RDD-AFT	0.37	0.04	0.23	0.24	0.34	0.07	0.39	0.39	96.1
Spline RDD-AFT	0.42	-0.01	0.33	0.36	0.38	0.03	0.58	0.59	96.6
S-AFT	0.55	-0.15	0.31	0.26	0.47	-0.07	0.22	0.19	94.1

TABLE 3 (Continued)

Method	Weak fuzziness					Strong fuzziness				
	Estimate	Bias	ESE	ASE	95% Coverage	Estimate	Bias	ESE	ASE	95% Coverage
Low confounding										
	Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000									
RDD-AFT	0.24	0.16	0.14	0.14	78.5	0.21	0.20	0.21	0.21	84.2
Spline RDD-AFT	0.38	0.03	0.29	0.31	96.5	0.37	0.03	0.46	0.48	95.6
S-AFT	0.84	-0.44	0.18	0.16	17.2	0.75	-0.35	0.15	0.13	19.8
	Bandwidth = 0.15, Treatment effect = 0.405									
RDD-AFT	0.16	0.25	0.15	0.15	63.2	0.14	0.27	0.24	0.24	80.1
Spline RDD-AFT	0.37	0.03	0.30	0.32	96.9	0.37	0.03	0.48	0.50	95.8
S-AFT	0.75	-0.35	0.20	0.17	41.9	0.64	-0.23	0.17	0.14	55.9
	Bandwidth = 0.1, Treatment effect = 0.405									
RDD-AFT	0.23	0.18	0.18	0.18	81.4	0.21	0.19	0.29	0.30	90.7
Spline RDD-AFT	0.37	0.03	0.32	0.34	96.4	0.38	0.02	0.53	0.54	95.4
S-AFT	0.63	-0.22	0.23	0.20	77.0	0.50	-0.09	0.19	0.16	89.9
	Bandwidth = 0.05, Treatment effect = 0.405									
RDD-AFT	0.34	0.07	0.24	0.26	95.0	0.34	0.06	0.42	0.44	96.6
Spline RDD-AFT	0.38	0.03	0.39	0.40	96.4	0.41	0.00	0.63	0.66	96.5
S-AFT	0.42	-0.01	0.30	0.25	96.0	0.34	0.07	0.24	0.20	93.5

(Continues)

TABLE 3 (Continued)

Method	Weak fuzziness					Strong fuzziness				
	Estimate	Bias	ESE	ASE	95% Coverage	Estimate	Bias	ESE	ASE	95% Coverage
High confounding										
	Bandwidth = 0.2, Treatment effect = 0.405, Sample size = 2000									
RDD-AFT	0.27	0.13	0.20	0.20	90.7	0.20	0.20	0.31	0.32	91.9
Spline RDD-AFT	0.37	0.03	0.47	0.46	94.4	0.37	0.04	0.72	0.73	95.1
S-AFT	1.36	-0.96	0.20	0.17	1.2	1.42	-1.02	0.20	0.16	1.3
	Bandwidth = 0.15, Treatment effect = 0.405									
RDD-AFT	0.15	0.26	0.22	0.22	78.6	0.13	0.27	0.36	0.36	88.8
Spline RDD-AFT	0.37	0.04	0.48	0.47	94.4	0.37	0.04	0.75	0.77	96.2
S-AFT	1.30	-0.89	0.23	0.19	2.3	1.30	-0.90	0.22	0.18	2.8
	Bandwidth = 0.1, Treatment effect = 0.405									
RDD-AFT	0.22	0.19	0.27	0.27	88.6	0.20	0.21	0.44	0.45	93.9
Spline RDD-AFT	0.37	0.04	0.50	0.49	94.2	0.37	0.03	0.81	0.83	96.6
S-AFT	1.21	-0.81	0.26	0.22	6.4	1.15	-0.75	0.25	0.20	7.3
	Bandwidth = 0.05, Treatment effect = 0.405									
RDD-AFT	0.32	0.08	0.39	0.38	93.7	0.34	0.07	0.67	0.68	96.9
Spline RDD-AFT	0.39	0.01	0.56	0.58	96.0	0.37	0.03	0.99	1.00	94.9
S-AFT	1.11	-0.70	0.34	0.28	25.9	1.01	-0.61	0.29	0.25	27.7

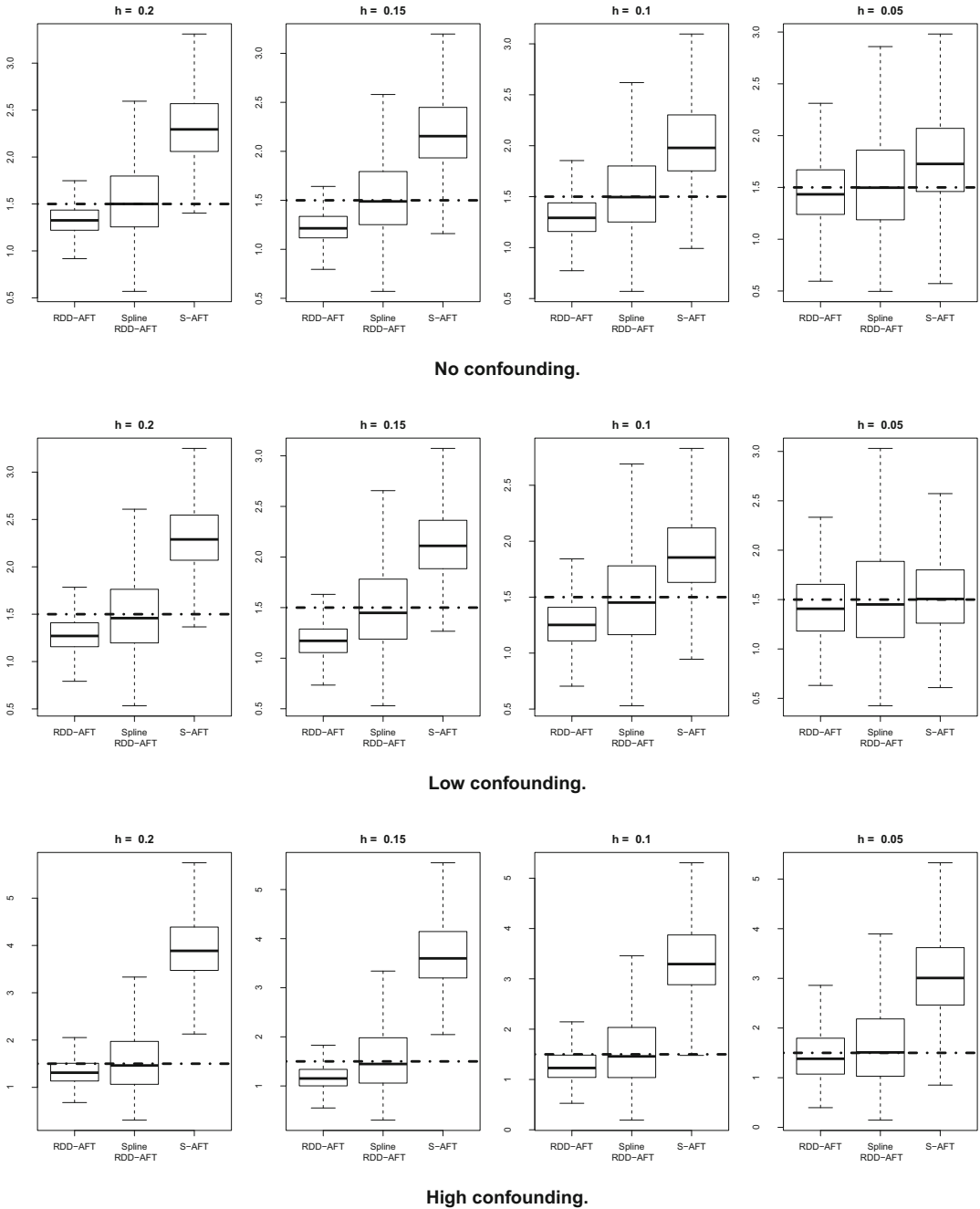


FIGURE 4 Boxplots showing the acceleration factor estimates from the simulation studies to compare performance of the standard RDD-AFT and spline-based RDD-AFT approaches for designs with weak fuzziness. The true value of the acceleration factor is 1.5 (dashed horizontal line). The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

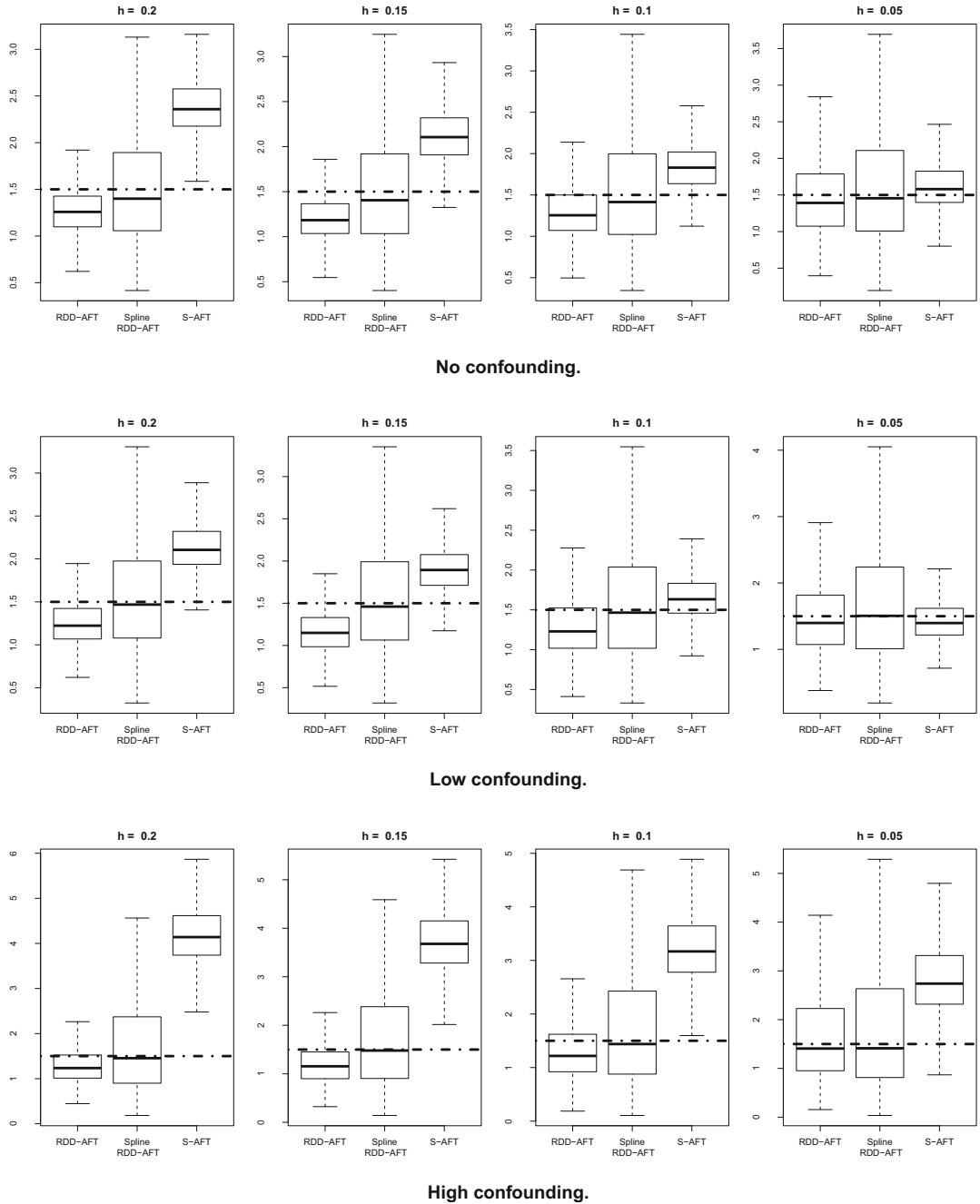


FIGURE 5 Boxplots showing the acceleration factor estimates from the simulation studies to compare performance of the standard RDD-AFT and spline-based RDD-AFT approaches for designs with strong fuzziness. The true value of the acceleration factor is 1.5 (dashed horizontal line). The sample size was 2000 in each simulated data set and simulations were repeated 1000 times

of developing type 2 diabetes. Specifically, a NICE guideline states that metformin should be considered as a treatment for patients whose HbA1c level is greater than or equal to 48 mmol/mol (NICE, 2015). Here, HbA1c is the continuous assignment variable, the threshold is an HbA1c level of 48mmol/mol and the treatment is metformin. The treatment effect of metformin is to reduce blood sugar which, in turn, reduces complications that may arise. We aim to explore the effect of metformin on survival (time to death) and time to a cardiovascular event (myocardial infarction or stroke) for patients who are at risk of type 2 diabetes (i.e. patients with a high HbA1c level). We consider a subset of patients from a large data set of patients in UK Primary Care, known as The Health Improvement Network (THIN) database. The THIN database contains anonymised data collected at over 500 UK general practices (family doctors) and is representative of the general UK population (Blak et al., 2011; Bourke et al., 2004). We use data from 4532 male patients aged between 40 and 80 years who had their first HbA1c measurement in 2010, who had not been diagnosed with diabetes previously and whose body mass index (BMI) was less than 30 kg/m². The time origin is the time of HbA1c measurement.

Of these 4532 patients, 643 patients had HbA1c values above the threshold and 3889 patients had HbA1c values below the threshold. Of those patients with HbA1c values above the threshold, 453 (70%) were prescribed metformin whereas, for those with HbA1c values below the threshold, 27 (1%) were prescribed metformin. Hence, it seems plausible that the probability of treatment is very different above and below the threshold, strengthening the case for the use of an RDD with these data.

For bandwidth selection, we consider the distributions of potential confounding variables for patients above and below the threshold. Since the bandwidth should reflect the ‘similarity’ of patients above and below the threshold, our aim is that, within a given bandwidth, distributions of confounding variables appear similar. The confounders that we consider, after discussion with epidemiologists, are: age at origin, BMI, low-density lipoprotein (LDL) cholesterol level and high-density lipoprotein (HDL) cholesterol level. Table 4 shows means and standard deviations for each of these variables for patients above and below the threshold, of both raw and standardised variables. Examining Table 4, we see that BMI values and LDL and HDL cholesterol values are broadly similar for all bandwidths, when comparing groups. For age, patients above the threshold appear to be slightly older in general than those below the threshold. As such, we will incorporate age at origin as a covariate when fitting models. In doing this, we are assuming that the relationship between the treatment and age has been correctly specified—we assume a simple linear relationship here.

Empirical probabilities of the events of interest are given Table 5. The probability of dying is similar in patients above and below the threshold across the bandwidths. Table 5 also shows median time-to-events for patients that experienced the events of interest. For mortality, empirical probabilities are very similar, above and below the threshold, for all bandwidths. Median times to death are broadly similar but perhaps slightly higher for subjects below the threshold, especially as the bandwidth increases. For CVD, the probability of a CVD event appears to be higher overall for subjects below the threshold. The median time-to-event is also higher for subjects below the threshold across all bandwidths. This may suggest that CVD event rates and times may differ for subjects above and below the threshold, perhaps implying that there is a discontinuity in the CVD event rate at the threshold.

Finally, we apply the methods discussed in Section 3 to the data. Since age appears to be unbalanced for patients above and below the threshold, age was included as a covariate in the methods for modelling both the LATE numerator and denominator. In comparing values of potential confounders above and below the threshold, we are seeking to verify that groups of patients above

TABLE 4 Sample means and standard deviations (SD) for potential confounding variables (both raw and standardised) above ($Z = 1$) and below ($Z = 0$) the threshold, for various HbA1c bandwidths (h)

Factors		$h = 10$		$h = 8$		$h = 6$		$h = 5$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at origin (years)	$Z = 0$	60.47	10.27	61.10	10.22	61.46	10.19	61.70	10.22
	$Z = 1$	62.75	9.80	63.04	9.63	62.90	9.85	63.41	9.83
Standardised age at origin	$Z = 0$	-0.02	1.00	-0.03	1.01	-0.02	1.00	-0.03	1.01
	$Z = 1$	0.20	0.96	0.17	0.95	0.12	0.97	0.14	0.97
BMI (kg/m^2)	$Z = 0$	25.98	2.66	26.14	2.63	26.23	2.59	26.29	2.62
	$Z = 1$	26.50	2.42	26.50	2.46	26.44	2.56	26.47	2.60
Standardised BMI	$Z = 0$	-0.02	1.01	-0.02	1.01	-0.01	1.00	-0.01	1.00
	$Z = 1$	0.18	0.92	0.12	0.94	0.07	0.99	0.06	0.99
LDL Cholesterol (mmol/L)	$Z = 0$	3.20	0.93	3.18	0.94	3.14	0.92	3.15	0.92
	$Z = 1$	3.15	0.87	3.17	0.85	3.19	0.85	3.19	0.86
Standardised LDL cholesterol	$Z = 0$	0.00	0.29	0.00	0.29	0.00	0.29	0.00	0.29
	$Z = 1$	-0.02	0.27	0.00	0.27	0.01	0.27	0.01	0.27
HDL Cholesterol (mmol/L)	$Z = 0$	1.30	0.37	1.27	0.35	1.25	0.36	1.24	0.36
	$Z = 1$	1.20	0.37	1.20	0.37	1.19	0.35	1.20	0.36
Standardised HDL cholesterol	$Z = 0$	0.01	0.29	0.01	0.28	0.01	0.29	0.01	0.29
	$Z = 1$	-0.07	0.29	-0.05	0.30	-0.04	0.28	-0.03	0.29

TABLE 5 Empirical probabilities of death and a CVD event, together with sample median time-to-event (in years) for a variety of bandwidths

Bandwidth:		$h = 10$		$h = 8$		$h = 6$		$h = 5$	
Threshold indicator Z		0	1	0	1	0	1	0	1
No. of patients n		1956	202	1172	184	860	155	626	133
Death	Probability	0.06	0.06	0.07	0.07	0.08	0.08	0.08	0.09
	Median time-to-event (years)	2.97	2.31	2.92	2.31	2.99	2.31	2.72	2.31
CVD Event	Probability	0.07	0.04	0.08	0.04	0.07	0.03	0.07	0.03
	Median time-to-event (years)	2.58	1.34	2.71	1.18	2.38	1.34	2.38	1.62

and below the threshold are balanced with respect to these variables, as would be expected when comparing randomised groups in a randomised controlled trial, in line with assumption A5 in Appendix A. Implicitly, we assume that patients either side of the threshold are exchangeable and that there is a degree of randomness with regard to where patients' HbA1c values lie (above or below the threshold).

Estimates and associated 95% confidence intervals of the acceleration factor for the effect of metformin prescription on mortality and CVD are presented in Table 6. Standard errors that were

TABLE 6 Estimates and 95% confidence intervals for the acceleration factor across varying bandwidths (h) for the RDD-AFT and S-AFT approaches. n denotes the sample size for the corresponding model

Event	Method	$h = 10$ $n = 2158$	$h = 8$ $n = 1356$	$h = 6$ $n = 1015$	$h = 5$ $n = 759$
All-cause mortality	RDD-AFT	2.37 (0.65, 8.60)	2.09 (0.42, 10.31)	1.32 (0.31, 5.70)	1.21 (0.22, 6.72)
	S-AFT	1.11 (0.16, 7.93)	1.22 (0.20, 7.40)	2.59 (0.43, 15.44)	2.59 (0.41, 16.17)
CVD event	RDD-AFT	2.47 (0.59, 10.27)	1.25 (0.26, 6.07)	1.12 (0.11, 11.76)	1.22 (0.07, 19.71)
	S-AFT	2.51 (0.34, 18.72)	2.69 (0.34, 21.16)	6.53 (0.92, 46.25)	3.51 (1.25, 9.88)

used to construct the confidence intervals for the RDD-AFT and S-AFT models were computed using bootstrapping, with a bootstrap sample size of 1000.

Acceleration factor estimates for the two methods are all greater than 1 and this may suggest that metformin prescription is beneficial for the events considered. That is, patients that receive metformin prescription have a higher median time-to-event (mortality and CVD) compared to patients that did not receive the prescription. However, these estimates are not statistically significant as the confidence bounds include 1 for the two methods. The 95% confidence intervals are generally wide and this may be because of a reduced sample size when the data are sub-sampled so that only patients whose HbA1c values lie close to treatment threshold are included in the RDD. In addition, the number of events observed is low for both deaths and CVD events which is likely to reduce the precision when estimating parameters of interest.

6 | DISCUSSION

In this work, we have focused on methods for an RDD with a time-to-event (survival) outcome using an AFT approach. Time-to-event outcomes are common in medical scenarios and, in addition, treatments are increasingly prescribed or administered according to pre-defined, external guidelines. As such, an RDD approach for time-to-event outcomes is appealing. We have shown that the assumptions made for a standard RDD may be adapted for use with a time-to-event outcome using an AFT model. The RDD-AFT model is intuitive in that the treatment effect estimate obtained may be interpreted directly in terms of the time-to-event.

We have shown that the RDD-AFT approach performs reasonably well, typically yielding an unbiased treatment effect estimate at the threshold, even when the RDD is fuzzy and there is confounding. We compared the approach to the commonly used S-AFT model and saw that, as would be expected, the S-AFT model did not provide an accurate or precise estimate of the treatment effect where confounding was ignored. When confounding variables were conditioned on in the simulation study, the S-AFT model produced more precise and unbiased estimates of the treatment effect at the threshold. We note that this relied on conditioning on the correct, and known, confounding variables. This is straightforward in a simulation study but perhaps not so in real data sets in which confounders may be unmeasured or unknown. In contrast, the RDD-AFT method did not rely on the inclusion of confounding variables in the fitted model, even for simulated scenarios in which a high level of confounding was incorporated, and the model treatment effect estimates were still, generally, unbiased. Despite this, it is still possible to include confounding variables in the RDD-AFT model if desired. In addition, we showed that the RDD-AFT model could be adapted to handle situations where the assignment variable–outcome variable relationship is non-linear, using a spline-based approach to estimation.

In the metformin prescription example, we compared the RDD-AFT and S-AFT models, controlling for baseline age as confounder. Although estimates from both approaches were imprecise, perhaps owing to the sample sizes, the point estimate of the acceleration factor was greater than one, for both time to death and time to a CVD event. The two modelling approaches agreed on the 'direction' of the acceleration factor estimate and yielded estimates reasonably close to one-another for all bandwidths. This may suggest that the RDD-AFT model would serve as a useful check in situations where an S-AFT model has been fitted.

The RDD-AFT model is parametric and we examined examples using a Weibull and a log-logistic distribution. This could be a drawback of the RDD-AFT approach, when compared to the semi-parametric S-AFT model which requires fewer distributional assumptions. However, the choice between parametric and non-parametric survival models may be often debated, even in a non-RDD setting. As with any statistical model, we would recommend that distributional assumptions are justified or checked before fitting a parametric RDD-AFT model. In addition, checking that the RDD assumptions are valid is also important.

Overall, we recommend that RDDs be considered more often for time-to-event outcomes, particularly in medical studies although methods presented may be easily adapted to other fields. The RDD-AFT method is an approach that is easy to fit using standard statistical software and that may produce a useful and interpretable estimate of the difference, if any, in the time-to-event between treated and untreated groups using observational data and a fuzzy RDD.

ACKNOWLEDGEMENTS

We thank the Associate Editor and reviewers, whose comments have helped to improve this paper. Mariam Adeleke received funding from a UCL Overseas Graduate Research Scholarship. Aidan O'Keefe received funding from the UCL/UCLH National Institute for Health Research Biomedical Research Centre. Approval for the use of the THIN data was granted by the Cegedim SA Scientific Review Committee in June 2019 (ref. 19THIN026). We thank Prof. Irene Petersen, Department of Primary Care and Population Health, UCL, for advice on the use of the THIN data.

ORCID

Gianluca Baio  <https://orcid.org/0000-0003-4314-2570>

Aidan G. O'Keefe  <http://orcid.org/0000-0002-0776-0868>

REFERENCES

- Aalen, O.O., Cook, R.J. & Røysland, K. (2015) Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis*, 21, 579–593.
- Angrist, J.D., Imbens, G.W. & Rubin, D.B. (1996) Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434), 444–455.
- Blak, B.T., Thompson, M., Dattani, H. & Bourke, A. (2011) Generalisability of the health improvement network (thin) database: demographics, chronic disease prevalence and mortality rates. *Informatics in Primary Care*, 19, 251–255.
- Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L. & Barnighausen, T. (2014) Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*, 25, 729–737.
- Bor, J., Fox, M.P., Rosen, S., Venkataramani, A., Tanser, F., Pillay, D. & Barnighausen, T. (2017) Treatment eligibility and retention in clinical HIV care: a regression discontinuity study in South Africa. *PLOS Medicine*, 14, e1002463.
- Bourke, A., Dattani, H. & Robinson, M. (2004) Feasibility study and methodology to create a quality-evaluated database of primary care data. *Informatics in Primary Care*, 12, 171–177.
- Buckley, J. & James, I. (1979) Linear regression with censored data. *Biometrika*, 66(3), 429–436.

- Calonico, S., Cattaneo, M.D. & Titiunik, R. (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82, 2295–2326.
- Caughey, D. & Sekhon, J.S. (2011) Elections and the regression discontinuity design: lessons from close U.S. house races, 1942–2008. *Political Analysis*, 19, 385–408.
- Chan, P.H., Xu, R. & Chambers, C.D. (2018) A study of R2 measure under the accelerated failure time models. *Communications in Statistics – Simulation and Computation*, 47, 380–391.
- Clarke, P.S. & Windmeijer, F. (2010) Identification of causal effects on binary outcomes using structural mean models. *Biostatistics*, 11, 756–770.
- Dawid, A.P. (1979) Conditional independence in statistical theory. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(1), 1–31.
- Didelez, V., Meng, S. & Sheehan, N.A. (2010) Assumptions of IV methods for observational epidemiology. *Statistical Science*, 25(1), 22–40.
- Erikson, R.S., Folke, O. & Snyder, J.M. (2015) A gubernatorial helping hand? How governors affect presidential elections. *The Journal of Politics*, 77(2), 491–504.
- Geneletti, S., O’Keeffe, A.G., Sharples, L.D., Richardson, S. & Baio, G. (2015) Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, 34, 2334–2352.
- Geneletti, S., Ricciardi, F., O’Keeffe, A.G. & Baio, G. (2019) Bayesian modelling for binary outcomes in the regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(3), 983–1002.
- Hahn, J., Todd, P. & der Klaauw, W.V. (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1), 201–209.
- Hastie, T.J. & Tibshirani, R.J. (1999) *Generalized additive models*. London: Chapman and Hall.
- Hernán, M.A. (2010) The hazards of hazard ratios. *Epidemiology*, 21, 13–15.
- Hernán, M.A. & Robins, J.M. (2006) Instruments for causal inference: an epidemiologist’s dream? *Epidemiology*, 4, 360–372.
- Hernán, M.A., Cole, S.R., Margolick, J.B., Cohen, M.H. & Robins, J.M. (2005) Structural accelerated failure time models for survival analysis in studies with timevarying treatments. *Pharmacoepidemiology and Drug Safety*, 14(7), 477–491.
- Huling, J.D., Yu, M. & O’Malley, A.J. (2019) Instrumental variable based estimation under the semiparametric accelerated failure time model. *Biometrics*, 75, 516–527.
- Imbens, G.W. & Angrist, J.D. (1994) Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475.
- Imbens, G. & Kalyanaraman K. (2012) Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, 79(3), 933–959. Available from: <https://doi.org/10.1093/restud/rdr043>
- Imbens, G. & Lemieux, T. (2008) Regression discontinuity designs: a guide to practice. *Journal of Econometrics*, 142, 615–635.
- Janes, H., Dominici, F. & Zeger, S. (2010) On quantifying the magnitude of confounding. *Biostatistics*, 11, 572–582.
- Joffe, M.M., Yang, W.P. & Feldman, H. (2012) G-estimation and artificial censoring: problems, challenges, and applications. *Biometrics*, 68, 275–286.
- Lange, T. & Hansen, J.V. (2011) Direct and indirect effects in a survival context. *Epidemiology*, 22, 575–581.
- Law, M.R., Cheng, L., Worthington, H., Mamdani, M., McGrail, K.M., Chan, F.K.I. & Majumdar, S.R. (2017) Impact of income-based deductibles on drug use and health care utilization among older adults. *CMAJ*, 189, E690–E696.
- van Leeuwen, N., Lingsma, H.F., Mooijaart, S.P., Nieboer, D., Trompet, S. & Steyerberg, E.W. (2018) Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials. *Journal of Clinical Epidemiology*, 98, 70–79.
- Li, F., Mattei, A. & Mealli, F. (2015) Evaluating the causal effect of University grant on student dropout: evidence from a regression discontinuity design using principal stratification. *Annals of Applied Statistics*, 9(4), 1906–1931.
- Linden, A., Adams, J.L. & Roberts, N. (2006) Evaluating disease management programme effectiveness: an introduction to the regression discontinuity design. *Journal of Evaluation in Clinical Practice*, 12, 124–131.
- Louis, T.A. (1981) Nonparametric analysis of an accelerated failure time model. *Biometrika*, 68(2), 381–390.

- Martinussen, T. & Vansteelandt, S. (2013) On collapsibility and confounding bias in Cox and Aalen regression models. *Lifetime Data Analysis*, 19, 279–296.
- Martinussen, T., Vansteelandt, S. & Andersen, P.K. (2020) Subtleties in the interpretation of hazard contrasts. *Lifetime Data Analysis*, 26(4), 833–855. Available from: <https://doi.org/10.1007/s10985-020-09501-5>
- National Institute for Health and Care Excellence (NICE). (2015) [NG28] Type 2 diabetes in adults: management. Available from: www.nice.org.uk/guidance/ng28
- O’Keeffe, A.G., Geneletti, S., Baio, G., Sharples, L.D., Nazareth, I. & Petersen, I. (2014) Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. *BMJ*, 349, g5293.
- R Core Development Team. (2019) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available from: <https://www.R-project.org>
- Ricciardi, F., Liverani, S. & Baio, G. (2020) Dirichlet process mixture models for regression discontinuity designs. arXiv.
- Robins, J. (1992) Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika*, 72(2), 321–334.
- Robins, J.M. & Tsiatis, A.A. (1991) Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in Statistics - Theory and Methods*, 20(8), 2609–2631. Available from: <https://doi.org/10.1080/03610929108830654>
- Robins, J.M., Hernán, M.A. & Brumback, B. (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5), 550–560.
- Sjölander, A., Dahlqvist, E. & Zetterqvist, J. (2016) A note on the noncollapsibility of rate differences and rate ratios. *Epidemiology*, 27, 356–359.
- Sutradhar, R. & Austin, P.C. (2018) Relative rates not relative risks: addressing a widespread misinterpretation of hazard ratios. *Annals of Epidemiology*, 28, 54–57.
- Therneau, T.M. (2021) Survival: A Package for Survival Analysis in R. Available from: <https://CRAN.R-project.org/package=survival>. R package version 3.2-13
- Thistlethwaite, D.L. & Campbell, D.T. (1960) Regression-discontinuity analysis: an alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51, 309–317.
- VanderWeele, T.J. (2009) Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, 20, 18–26.
- VanderWeele, T.J. (2011) Causal mediation analysis with survival data. *Epidemiology*, 22, 582–585.
- Vansteelandt, S. & Goetghebuer, E. (2003) Causal inference with generalized structural mean models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(4), 817–835.
- Wei, L.J. (1992) The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871–1879.
- Xu, K.-L. (2017) Regression discontinuity with categorical outcomes. *Journal of Econometrics*, 201(1), 1–18.

How to cite this article: Adeleke, M.O., Baio, G. & O’Keeffe, A.G. (2022) Regression discontinuity designs for time-to-event outcomes: An approach using accelerated failure time models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 185(3), 1216–1246. Available from: <https://doi.org/10.1111/rssa.12812>

APPENDIX A. RDD ASSUMPTIONS

Within the conditional independence framework, the notation used is such that if a random variable A is independent of another random variable, B , we write $A \perp\!\!\!\perp B$ and, where A and B are not independent, we write $A \not\perp\!\!\!\perp B$. We note that the RDD assumptions may also be expressed using a potential outcomes framework (Hahn et al., 2001; Imbens & Lemieux, 2008).

A1: The threshold and treatment indicators are not independent.

$$A_i \not\perp Z_i.$$

This assumption implies that the threshold rule and the allocation to treatment must be associated. For a sharp design, this relationship would be deterministic. For a fuzzy design, there should be an association between A_i and Z_i which should be of sufficient strength for an RDD approach to be valid.

A2: The probability of receiving treatment is discontinuous at the threshold. That is

$$\lim_{x \downarrow x_0} \mathbb{P}(A_i = 1 | X_i = x) \neq \lim_{x \uparrow x_0} \mathbb{P}(A_i = 1 | X_i = x).$$

This implies that the threshold signifies a separation in the probability of treatment (above and below) and, coupled with A1, ensures that the strength of association between treatment and threshold allows an RDD to be used.

A3:

$$f(t | X_i = x, Z_i, A = a) \text{ is continuous at } x = x_0 \text{ for } a \in \{0, 1\},$$

where $f(\cdot)$ is the distribution function of the time-to-event. If a treatment effect on the time-to-event occurs then we should expect a difference in the distribution of the time-to-event above and below the threshold. Assumption A3 ensures that any ‘jump’ or difference in the time-to-event distribution at the threshold is due to the treatment and not any other variable.

A4: The threshold indicator is independent of confounders conditional on the assignment variable.

$$Z_i \perp C_i | X_i.$$

This assumption implies that the treatment guideline (threshold rule) cannot be adapted for individual subjects and we note that this should hold at least within the region of the threshold.

A5: The threshold indicator and the time-to-event outcome variable are independent conditional on the other variables.

$$T_i \perp Z_i | (X_i, A_i, C_i).$$

This assumption ensures that, within a group of exchangeable subjects with assignment variable values close to the threshold, subjects ought to fall randomly above and below the treatment threshold. The threshold represents an externally defined decision rule and does not change based on subject characteristics. We should expect confounding variables to be distributed similarly amongst groups of subjects above and below the threshold, in an analogous manner to randomised groups in a randomised controlled trial. In this sense, the threshold acts as a quasi-randomisation device for a group of subjects whose assignment variable values lie close enough to the threshold for these subjects to exhibit similar characteristics. In practice, this assumption is likely to hold if there is an association between

the threshold and treatment assignment and as long as patients are not prescribed (or prescribed) treatment in a systematic manner that ignores the threshold–assignment variable decision rule. In addition, this assumption implies that the treatment guideline (threshold) is externally defined and would imply that subjects cannot manipulate their outcome variable so that they would lie above (or below) the treatment threshold. As the outcome is a time-to-event (and especially in the metformin example where the outcome is all-cause mortality) this example is likely to hold in many scenarios.

- A6:** We assume that there is no systematic non-adherence to the treatment guideline. This assumption implies that there are no healthcare practitioners who would systematically behave in the opposite way suggested by the threshold rule. In other words, that there is not a practitioner who would deliberately prescribe treatment to *all* subjects with assignment variables below the threshold and withhold treatment from subjects with assignment variable values above the threshold. This assumption would, in general, seem plausible and especially so in the metformin example discussed in this paper.
- A7:** Censoring is uninformative. We assume that right censoring would occur at similar rates (and for similar reasons) above and below the threshold (or, at least, in the subset of subjects whose assignment variables lie close enough to the threshold to be included in an RDD).

APPENDIX B. DERIVATION OF THE RDD-AFT ESTIMATOR

Consider the model:

$$\log T = \beta A + \varepsilon \quad (\text{A1})$$

β is the treatment effect of interest. Here, the expectation of ε is not necessarily zero as it represents the expected value of $\log T$ when treatment is not received.

Fitting the marginal model in Equation (A1) may result in a biased estimate of the treatment effect, owing to confounding variables. Therefore, we estimate the treatment effect by exploiting the fact that we have partial information on treatment allocation.

We can represent

$$\begin{aligned} \lim_{x \downarrow x_0} \mathbb{E}(\log T | X = x) &= \lim_{x \downarrow x_0} \mathbb{E}(\log T | X = x, Z = 1) \text{ and} \\ \lim_{x \uparrow x_0} \mathbb{E}(\log T | X = x) &= \lim_{x \uparrow x_0} \mathbb{E}(\log T | X = x, Z = 0). \end{aligned}$$

In subsequent derivations, the limits will be dropped for simplicity: $\mathbb{E}(\log T | Z = 1)$ will appear instead of $\lim_{x \downarrow x_0} \mathbb{E}(\log T | X = x)$ and $\mathbb{E}(\log T | Z = 0)$ instead of $\lim_{x \uparrow x_0} \mathbb{E}(\log T | X = x)$. It follows that

$$\begin{aligned} \mathbb{E}(\log T | Z = 1) - \mathbb{E}(\log T | Z = 0) &= \mathbb{E}(\beta A + \varepsilon | Z = 1) - \mathbb{E}(\beta A + \varepsilon | Z = 0) \\ &= \beta [\mathbb{E}(A | Z = 1) - \mathbb{E}(A | Z = 0)] \\ &\quad - [\mathbb{E}(\varepsilon | Z = 1) - \mathbb{E}(\varepsilon | Z = 0)] \end{aligned}$$

Substitution of $\varepsilon = \mathbb{E}(\log T | A = 0)$ yields

$$\begin{aligned} \mathbb{E}(\log T | Z = 1) - \mathbb{E}(\log T | Z = 0) &= \beta [\mathbb{E}(A | Z = 1) - \mathbb{E}(A | Z = 0)] \\ &\quad - [\mathbb{E}(\log T | A = 0, Z = 1) - \mathbb{E}(\log T | A = 0, Z = 0)] \end{aligned}$$

Using Assumption 3, we have

$$\begin{aligned} \mathbb{E}(\log T | A = 0, Z = 1) - \mathbb{E}(\log T | A = 0, Z = 0) &= 0 \\ \implies \mathbb{E}(\log T | Z = 1) - \mathbb{E}(\log T | Z = 0) &= \beta[\mathbb{E}(A | Z = 1) - \mathbb{E}(A | Z = 0)]. \end{aligned}$$

Therefore, reverting to the original notation including limits,

$$\beta = \frac{\lim_{x \downarrow x_0} \mathbb{E}(\log T | X = x) - \lim_{x \uparrow x_0} \mathbb{E}(\log T | X = x)}{\mathbb{E}(A | Z = 1) - \mathbb{E}(A | Z = 0)}.$$

The numerator term is equivalent to the estimate of α_1 in Equation (2) while the denominator term is the probability of compliance.

APPENDIX C. G-ESTIMATION: OVERVIEW

We define the counterfactual outcomes of receiving and not receiving treatment as T_i^1 and T_i^0 , respectively. The counterfactual outcome of not receiving the treatment is estimated from the S-AFT model using the relationship:

$$T_i^0 = \exp(-\gamma)T_i^1. \quad (\text{A2})$$

We note that if the i^{th} subject did not receive the treatment, then $T_i = T_i^0$ whereas $T_i = T_i^1$ if the i^{th} subject received the treatment, where T_i is the observed outcome. Hence Equation (A2) may be rewritten as:

$$T_i^0(\gamma) = \exp(-\gamma A_i)T_i$$

and $T_i^0 = T_i$ for subjects who do not receive treatment and $T_i^0 = \exp(-\gamma)T_i$ for those who receive treatment. We consider T_i^0 as a function and write $T_i^0 = T_i^0(\gamma)$; the parameter of interest, to be estimated using G-estimation, is γ .

A logistic regression model for the probability of treatment, conditional on observed confounders and $T_i^0 = T_i^0(\gamma)$ may be written

$$\text{logit}\{\mathbb{P}(A_i = 1 | \mathcal{O}_i, T_i^0(\gamma))\} = \theta_0 \mathcal{O}_i + \theta_1 T_i^0(\gamma). \quad (\text{A3})$$

When assumption S1 holds, that is, $T_i^0(\gamma)$ is independent of treatment conditional on \mathcal{O}_i , then $\theta_1 = 0$. Therefore the G-estimate of γ is the value of γ (say, γ^*) that leads to the failure to reject the null hypothesis that $\theta_1 = 0$. The G-estimate can be obtained by minimising the score test statistic for $\theta_1 = 0$ with respect to γ^* . This is equivalent to finding the solution to the estimating equation (Hernán et al., 2005):

$$U(\gamma^*) = 0,$$

where

$$U(\gamma^*) = \sum_i T_i^0(\gamma^*) [A_i - \mathbb{P}(A_i = 1 | \mathcal{O}_i)]. \quad (\text{A4})$$

Hence, the value γ^* that solves the estimating equation above is the G-estimate of γ . The solution to the estimating in Equation (A4) is computed using the `uniroot` function in R.

To handle right-censoring using G-estimation, the type of censoring should be considered. Specifically, we should consider whether censoring occurs owing to a subject dropping out prior to the end of a study's follow-up period or if censoring happens because the event of interest has not occurred by the end of follow-up (administrative censoring). When censoring occurs because of drop out, inverse probability weighting is used to account for right-censoring. We define an indicator variable D_i , where $D_i = 0$ if the i^{th} subject drops out before the end of the study and $D_i = 1$ otherwise. The inverse probability weight for censoring is calculated as

$$W_i^D = \frac{D_i}{\mathbb{P}(D_i = 1 | X_i, A_i)}$$

with $W_i^D = 0$ if $D_i = 0$. The score function (Equation A4) is then adjusted to account for the inverse probability weight.

For administrative censoring, we specify K to be the end of follow-up time for the study, such that any subject for whom $T_i^0(\gamma) \geq K$ would have a time-to-event that is right-censored. We define $K(\gamma)$ as

$$K(\gamma) = \begin{cases} K & \text{if } \gamma \leq 0; \\ K \exp(-\gamma) & \text{if } \gamma > 0. \end{cases}$$

Using $K(\gamma)$, we define $\Delta_i(\gamma) = \mathbb{1}(T_i^0(\gamma) \geq K(\gamma))$. For any subject where $\Delta_i(\gamma) = 1$, $T_i^0(\gamma)$ is replaced with zero in the estimating equation below (Equation A5). Further explanation regarding the intuition behind this approach can be found in Hernán et al. (2005)

$$U(\gamma^*) = \sum_i W_i^D \Delta_i(\gamma^*) [A_i - \mathbb{P}(A_i = 1 | \mathcal{O}_i)]. \quad (\text{A5})$$

The G-estimate of γ is then the γ^* that solves the estimating in Equation (A5) and, typically, this equation will be solved numerically (for example, using `uniroot` function in R (R Core Development Team, 2019)).

APPENDIX D. SIMULATION STUDY RESULTS WHERE THE HAZARD FUNCTION IS NON-MONOTONIC

The distribution of (ψ_i) in Step 6 of the simulation study presented in Section 4 is simulated from a logistic distribution:

$$\psi_i \sim \text{Logistic}[\log(7), 0.15].$$

Therefore, T^* comes from a log-logistic distribution with scale parameter of 7 and shape parameter of $\frac{1}{0.15}$. Since the value of the shape parameter is greater than 1, the hazard function of T^* is clearly non-monotonic, which implies the Weibull distribution is not appropriate. Figure A1 shows boxplots that summarise simulation study estimates.

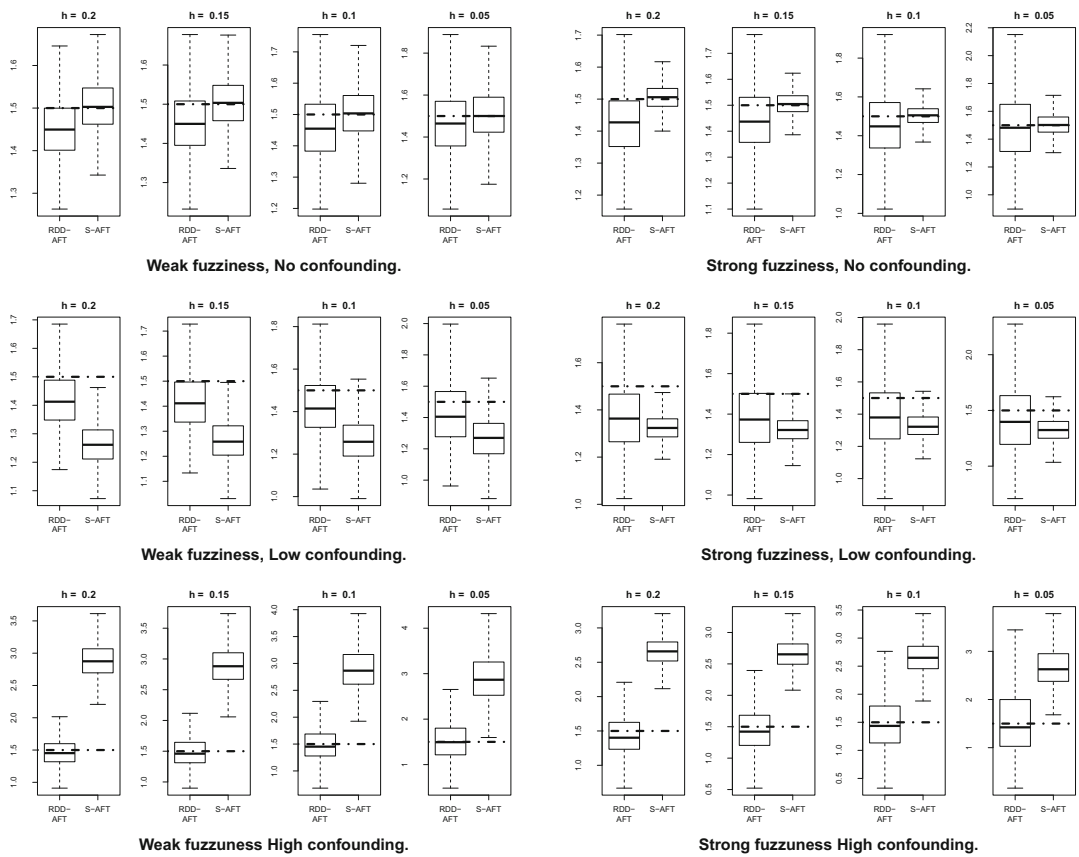


FIGURE A1 Boxplots showing the acceleration factor estimates from the simulation studies to compare performance of the RDD-AFT and S-AFT approaches. The true value of the acceleration factor is 1.5 (dashed horizontal line). The sample size was 2000 in each simulated data set and simulations were repeated 1000 times