Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes

Nikki van Leeuwen
Hester F Lingsma
Anton J de Craen
Daan Nieboer
Simon Mooijaart
Edo Richard
Ewout W Steyerberg

Epidemiology 2016

ABSTRACT

Introduction

In epidemiology the regression discontinuity design has received increasing attention recently and might be an alternative to a randomized controlled trial (RCT) to evaluate treatment effects. In RD treatment is assigned above a certain threshold of an assignment variable for which the treatment effect is adjusted in the analysis.

Methods

We performed simulations and a validation study in which we used treatment effect estimations from an RCT as the reference for a prospectively performed regression discontinuity study. We estimated the treatment effect using linear regression with adjustment for the assignment variable both as linear terms and restricted cubic spline and using local linear regression models.

Results

In the first validation study the estimated treatment effect from a cardiovascular RCT was -4.0 mmHg blood pressure (95%CI -5.4;-2.6) at two years after inclusion. The estimated effect in regression discontinuity was -5.9 mmHg (95%CI -10.8; -1.0) with restricted cubic spline adjustment. Regression discontinuity showed different, local effects when analyzed with local linear regression. In the second RCT, regression discontinuity treatment effect estimates on total cholesterol level at three months after inclusion were similar to RCT estimates, but at least 6 times less precise.

Conclusion

Concluding, regression discontinuity may provide similar estimates of treatment effects to RCT estimates, but requires the assumption of a global treatment effect over the range of the assignment variable. In addition to a risk of bias due to wrong assumptions, researchers need to weigh better recruitment against the substantial loss in precision when considering a study with regression discontinuity versus RCT design.

INTRODUCTION

Randomized controlled trials (RCTs) are the reference standard to demonstrate the efficacy of medical interventions. However, recruitment of sufficient numbers of patients is a challenge in RCTs, and becomes increasingly difficult due to regulatory requirements. Including participants is particularly challenging when the effect of an intervention on an outcome is of interest, but withholding treatment is considered unethical. Also, patients may not want to be randomized(1, 2) or physicians are not willing to include patients.(3) It is estimated that between only 3% and 5% of all eligible adult cancer patients in the United Stated and the United Kingdom enroll in RCTs, partly due to dislike of the concept of trials and the idea of randomization of both patient and clinician.(4) Low recruitment rates in trials are also common in other fields, especially in surgery(5-6) and elderly(7, 8) research. Failure to achieve recruitment goals limits statistical precision, leads to an increase of costs, and decreases the efficiency of a RCT.(9) Even when investigators enroll an adequate number of participants, they rarely do so on schedule.(3, 10, 11)

Second, low recruitment rates threaten the generalizability of the findings in RCTs. Patients enrolled in trials may poorly represent the population of interest.(8, 12) Mostly women and elderly are underrepresented in RCTs.(8, 13)

An alternative epidemiologic design to assess effectiveness of medical treatment is the quasi-experimental "regression discontinuity" design. This design is common in the social sciences, and was introduced in public health and medicine in 1996.(14) Although in other fields regression discontinuity has been evaluated(15-20), recently Vandenbroucke et al.(21), Bor et al.(22) and O'Keeffe et al.(23) noted the importance of studying the feasibility and robustness of this design in clinical settings. In the regression discontinuity design, treatment is not allocated randomly, but is assigned to a subset of patients, based on a threshold of a baseline characteristic. The control group consists of a subset of patients below the threshold, not receiving treatment. The threshold variable does not necessarily have to be prognostic for the outcome measure assessed in the study. E.g. all patients with a baseline blood pressure (BP) over 140 mmHg may receive treatment (intervention group) and patients with a baseline BP below 140 mmHg do not receive treatment (control group). Such treatment allocation closely resembles clinical practice and may thus facilitate easier recruitment of participants into a prospective, comparative study. Due to the assignment rule, regression discontinuity designs can achieve balance on unobserved factors, just like in an RCT. When estimating the treatment effect, a regression analysis compares treated to control patients, while adjusting for the assignment variable, in this example baseline BP. Regression discontinuity provides a possible opportunity for obtaining unbiased causal effect estimates, when experiments are not feasible or when we want to evaluate interventions under "real-life" circumstances.(24)

The regression discontinuity design as a prospective quasi-experimental study might be attractive because the challenges of the randomization process are avoided. However, the estimates from a quasi-experimental regression discontinuity design might be different and substantially less efficient compared to an RCT. We aimed to assess validity of this design as a prospective quasi-experimental design compared to a traditional RCT, since this has not been discussed in detail in the epidemiologic literature. In this study we perform simulation studies and analyze data from two large cardiovascular RCTs as validation studies.

METHODS

Monte Carlo Simulations

We used Monte Carlo simulations to compare regression discontinuity and RCT. One hundred patients were simulated with a hypothetical mean prognostic measurement of 10 and a standard deviation (SD) of 2. The mean outcome was 90 (SD 20) and was normally distributed. A treatment effect of -10 was simulated. Linear correlations of the prognostic measurement with outcome were varied (R^2 0.0, 0.2, 0.5 and 0.8) to assess the importance of the prognostic strength of the adjustment model in the regression discontinuity design.

For the RCT, treatment was allocated at random, and sample size was 100 patients in total (50 per arm). In the regression discontinuity setting, treatment was assigned to 50 patients with a prognostic measurement above 10; 50 patients with a prognostic measurement below 10 were used as controls. For both the RCT and regression discontinuity settings, linear regression models were used to estimate the treatment effect, adjusted for the baseline variable both in a linear term and a restricted cubic spline term. The regression discontinuity setting was also analyzed with local linear regression analysis. In local linear regression, only patients around the threshold are used in the analysis to estimate a *local* treatment effect while normal regression uses all patients, resulting in a *global,* or average, treatment effect estimate. Treatment effect estimates were compared in terms of bias (expressed as mean estimated treatment effect) and precision (expressed as mean squared error of the treatment effect estimate). The simulation code is provided in the Appendix.

Validation study

Two cardiovascular trials were used to validate the regression discontinuity design in empirical data. The "Prevention of Dementia by Intensive Vascular Care" study (preDIVA) is an ongoing cluster-randomized trial to assess the efficacy of a multicomponent, nurse-led intervention targeting all cardiovascular risk factors in an elderly population (70-78

years). The study protocol has been approved by the medical ethical committee of the Academic Medical Centre.(25) The primary outcome of this RCT is incident dementia during 6 years of follow-up. Of 3533 patients enrolled, 1894 are in the intervention and 1639 in the control group. As this RCT is ongoing we randomly sampled 3000 patients from the enrolled patients.

The second RCT was the "PROspective Study of Pravastatin in elderly (70 to 82 years) individuals at risk of vascular disease" (PROSPER). It was conducted between December 1997 and May 1999 and enrolled 5804 patients, who were assigned to pravastatin (n=2891) or placebo (n=2913) to reduce the risk of coronary disease in elderly individuals.(26)

To validate the regression discontinuity design we used continuous measures collected during follow-up, which were not the primary endpoints of the trials, as the outcome variable. To evaluate the influence of the choice of the assignment variable on the estimates we did multiple analyses using two different baseline measures. For preDIVA both age and the blood pressure (BP) at baseline were used as the assignment variable and the BP at 2-year intermediate follow-up as outcome. Both BP measures were calculated as the mean of two systolic blood pressure measurements during a visit (expressed in mmHg). BP data at 2 years were unavailable for part of the 3000 randomly selected patients, mostly because they reached a clinical endpoint before 2 years, or missed one study visit. We could hence include 2346 patients for analysis. For PROSPER, we considered total cholesterol level measured three months after inclusion (expressed in mmol/L) as the outcome and both age and total cholesterol at baseline as the treatment assignment variables. After exclusion of patients with missing 3-month total cholesterol we were able to analyze 5581 patients from PROSPER.

Statistical analysis

Baseline characteristics were described with standard descriptive statistics; median and interquartile range for continuous variables and frequencies and percentages for categorical variables.

To analyze the data as an RD design, we selected those patients with a value of the assignment variable above a certain threshold treated as the intervention group, and those with a value below that threshold and not treated as control group. This thus led to a sample of approximately half the trial population. In both trials we used a threshold of the assignment variable known to be used in clinical practice(27, 28) (e.g. BP > 140 mmHg), or, if not available, a hypothetical threshold (e.g. age > 72 years). Histograms of the assignment variables in both preDIVA and PROSPER are shown in the eAppendix 1. In this patient selection the treatment effect was estimated using a linear regression model, with adjustment for the assignment variable (Y ~ Tx + Baseline assignment variable). We further analyzed a hypothetical different cut-off to assess the robustness of the

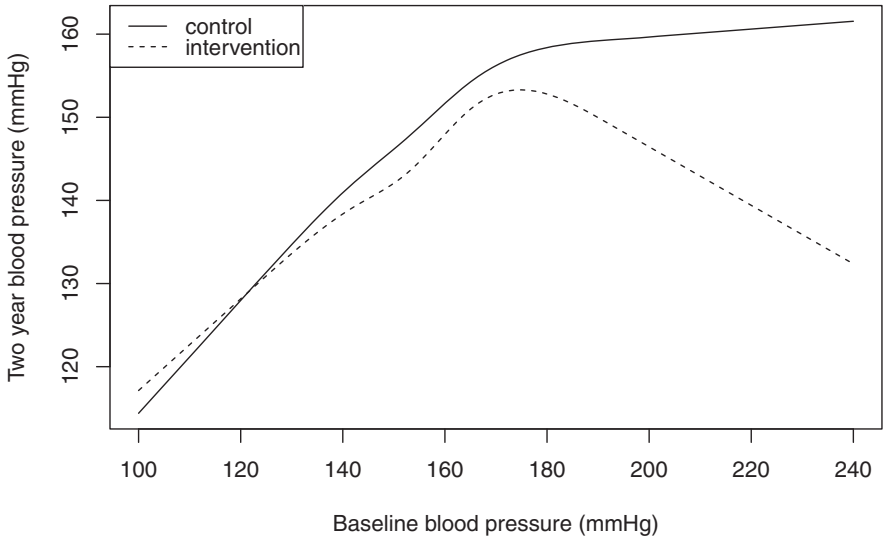
treatment effect estimate to the chosen cut-off. In other words, whether the treatment effect was global or local. Usually this would not be possible in regression discontinuity as the cut-off is determined in advance.

Analysis on these sets of patients was compared to a random sample (repeated 5,000 times) of half of the of the RCT patients. To compare relative efficiency in terms of required sample size between the different approaches we used standard errors (SEs) of the estimated treatment effects in the following formula: $(SE_{RD} / SE_{RCT})^2$. This is the ratio of variances in both designs.

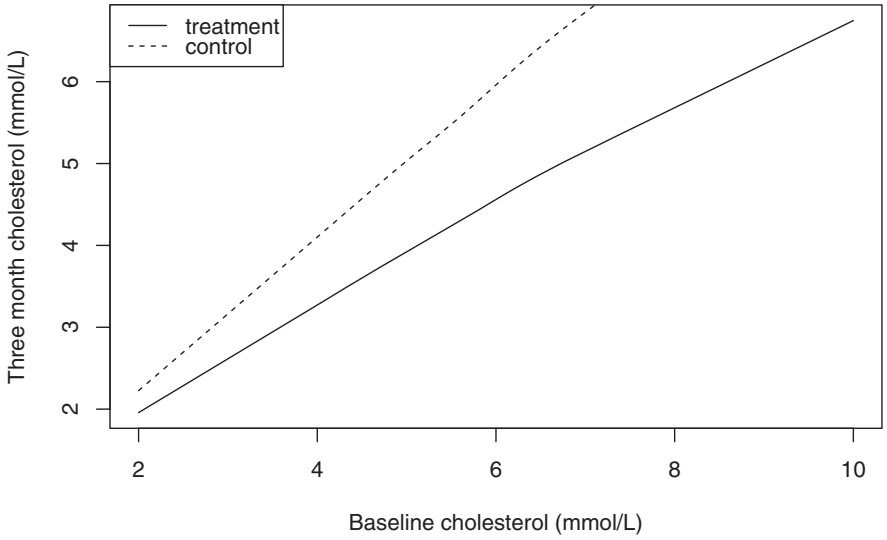
As the validity of the regression discontinuity design is dependent on proper adjustment for the assignment variable we explored the relation between the assignment variable and outcome in detail. We assessed non-linearity with non-linear restricted cubic spline functions and interaction between the baseline assignment variable and the treatment effect. Both were presented graphically (Figure 1a and 1b). A restricted cubic spline function is a smooth function that consists of pieced-together cubic splines that are restricted to be linear in the tails.(29) We used the default setting for flexibility of the model with five knots.(30) Consequently we used the restricted cubic spline of the assignment variable in the adjustment model to obtain the optimal model fit for adjustment. This method was compared to local linear regression models for the adjustment of the baseline variable. All analyses (RCT and regression discontinuity) were adjusted for the assignment measurement that was used to assign treatment; both age and baseline BP in preDIVA and both age and baseline cholesterol in the PROSPER trial. R^2 statistics were calculated to indicate the explained variance of the assignment model.

We further explored different assumptions on interaction between the assignment variable and treatment. We assumed no interaction between age and treatment in both studies. Therefore, we considered the treatment effect estimates in the regression discontinuity studies in which treatment was assigned on age, global treatment effects. We compared these treatment effect estimates to the global effect estimated in the RCT. The treatment effect estimates from the regression discontinuity studies, where treatment was assigned on baseline BP and baseline total cholesterol level, could be considered as local treatment effects, since we assumed interaction between treatment over both baseline BP and baseline cholesterol level. Therefore we compared these estimates to the local effects in the RCT, estimated with restricted cubic spline adjustment. These estimates are the differences between the treated and untreated lines in figures 1a and 1b. Treatment effects were estimated using linear regression and expressed as regression coefficients with 95% confidence Intervals (95% CIs) for blood pressure or cholesterol level in the treatment group compared to the controls.

All statistical analyses were performed in R statistical software version 2.15.3 (R Foundation for Statistical Computation, Vienna, Austria) using the rdd package and Harrell's rms package.

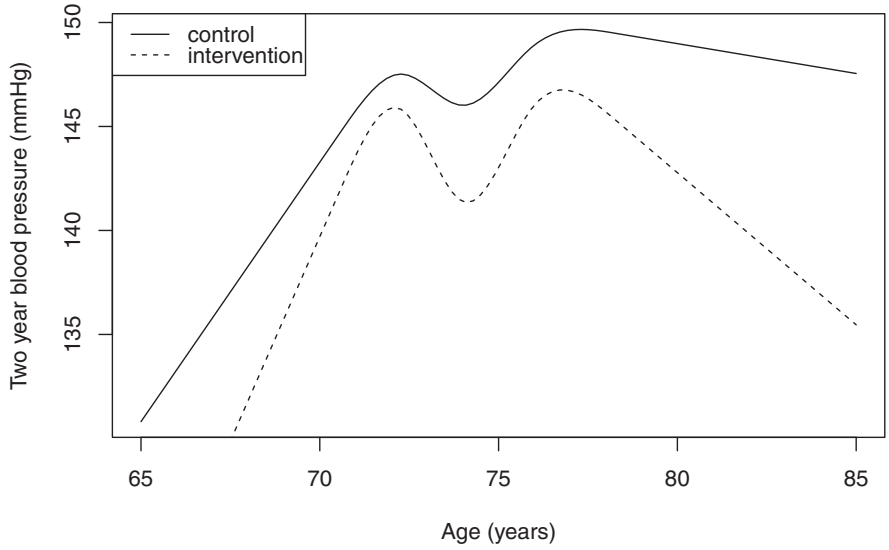


a) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the intervention effect over baseline blood pressure in the preDIVA study. ** The function fitted is: Two year blood pressure ~ Intervention * rcs(Baseline blood pressure).*

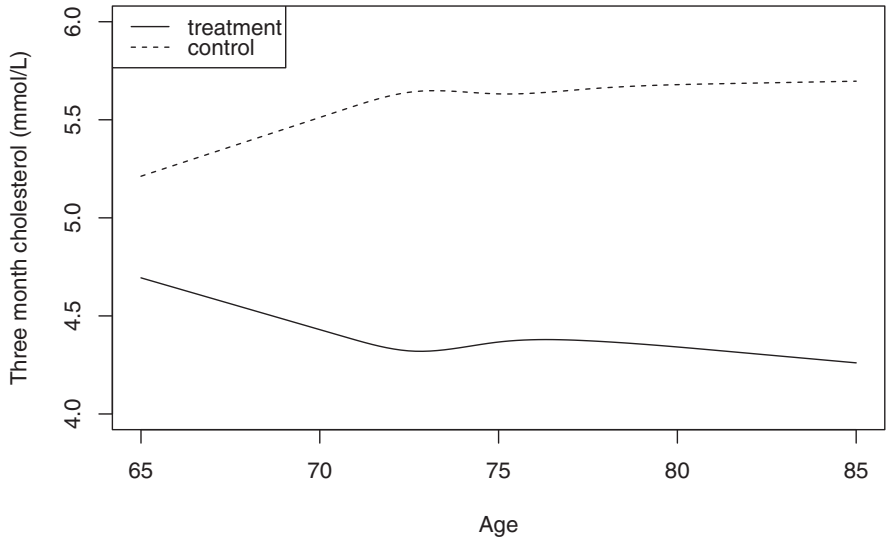


b) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the treatment effect over baseline total cholesterol in the PROSPER trial. ** The function fitted is: Three month cholesterol ~ Treatment * rcs(Baseline cholesterol).*

Figure 1. Non-linear restricted cubic spline functions of the interaction of the intervention effects over the assignment variables in both the preDIVA study (n = 2346) and the PROSPER trial (n = 5581).



c) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the intervention effect over baseline age in the preDIVA study. ** The function fitted is: Two year blood pressure ~ Intervention * rcs(Baseline age).*



d) Non-linear restricted cubic spline *(rcs)* function* of the interaction of the treatment effect over baseline age in the PROSPER trial. ** The function fitted is: Three month cholesterol ~ Treatment * rcs(Baseline age).*

Figure 1. Non-linear restricted cubic spline functions of the interaction of the intervention effects over the assignment variables in both the preDIVA study (n = 2346) and the PROSPER trial (n = 5581).

RESULTS

Simulations

Simulations showed that under the ideal situation of a linear relation between the assignment variable and the outcome, no unmeasured confounders, and no treatment effect interaction, regression discontinuity provided unbiased treatment effect estimates all scenarios (Table 1). However, RD with linear adjustment resulted in substantially less precise effect estimates compared to the RCT. For instance, in the hypothetical setting with an R^2 of 20% for the correlation of the assignment measurement with outcome, the mean squared error of the estimated treatment effect estimate in an RCT was 3.2 compared to 9.0 in the RD design. In this example this means that when the regression discontinuity design is used and the analysis matches the underlying true model, triple the number of patients is required to obtain the same statistical precision as when using an RCT. This magnitude of (in)efficiency was consistent for all simulated correlations between the assignment measurement and outcome. Regression discontinuity estimates adjusted with restricted cubic splines were 7 times less efficient than the RCT estimates analyzed with restricted cubic splines. When analyzing regression discontinuity with local linear regression, the estimated were on average 1.4 times less efficient than analyzing regression discontinuity with restricted cub spline adjustment (Table 1).

Table 1. Simulations treatment effect over baseline in randomized control trial and regression discontinuity design, analyzed with linear regression, restricted cubic spline functions and local linear regression.

		Randomized controlled trial				Regression discontinuity design			
	R^2 (%)	0	20	50	80	0	20	50	80
Linear regression	**Mean squared error**	4.2	3.2	2.0	0.8	11.1	9.0	5.6	2.3
Restricted cubic spline adjustment	**Mean squared error**	4.3	3.3	2.1	0.8	29.5	23.3	14.9	6.0
Local linear regression	**Mean squared error**	NA	NA	NA	NA	39.8	32.6	20.7	7.9

Validation study

In the validation study we assessed blood pressure in 2346 patients from the preDIVA trial. The median age was 74 years (IQR: 72, 76) and the median BP at baseline was 153 mmHg (IQR: 140, 168). We analyzed 5581 patients from the PROSPER trial, who had a median age of 74 years (IQR: 72 - 77 years) and a median total cholesterol level of 5.6 mmol/L (IQR: 5.0, 6.3) at baseline (Table 2).

In the RCT design, the treatment effect estimate on BP at two years adjusted for baseline BP was -4.0 mmHg (95% CI -5.4; -2.6) (Table 3).

Table 2. Patient characteristics of preDIVA (n=2346) and PROSPER (n=5581). Numbers are the median (IQR) or frequency (%)

Characteristic	preDIVA	PROSPER
Age, years	74 (72, 76)	74 (72, 77)
Sex, male	1071 (46)	2708 (49)
Baseline blood pressure in mmHg	153 (140, 168)	-
2-year blood pressure in mmHg	148 (136, 162)	-
Baseline cholesterol in mmol/L	-	5.6 (5.0, 6.3)
Three month cholesterol in mmol/L	-	4.9 (4.2, 5.7)

The adjusted estimated effect was -5.9 mmHg (95% CI -10.8; -1.0) with the hypothetical regression discontinuity design, adjusted for BP as restricted cubic spline variable. Here, patients with a BP over a baseline BP of 140 mmHg would receive treatment and patients with a lower BP would not receive treatment. An additional analysis of a hypothetical RD design in which patients with a BP of 160 mmHg would receive treatment and the higher baseline BP patients not, adjusted with a restricted cubic spline showed a treatment effect estimate of -5.9 mmHg (95% CI -11.4; -0.3) (Table 3).

In the preDIVA trial, there appeared to be a different treatment effect for patients with relative low baseline BP compared to high baseline BP, indicating statistical interaction (Figure 1a). This explains the different effect estimates in the regression discontinuity setting with treated high-risk patients and low-risk controls compared to the RCT estimate. When the treatment effect was analyzed with local linear regression, the intervention effect estimates in the two regression discontinuity designs were more different than the RCT estimate: -9.3 mmHg (95% CI -18.5; -0.1) for the setting with the cut-off at 140 mmHg and -10.2 mmHg (95% CI -21.0; 0.6) for the setting with the cut-off at 160 mmHg (Table 3).

In Figure 1c shows interaction between treatment and age. A global effect of treatment over the age range was assumed. The estimates from the regression discontinuity design analyzed with restricted cubic splines showed more similar estimates for the different cut-offs and these were closer to the estimate from the RCT (-0.66 (-6.44; 5.12) -2.71 (-7.16; 1.74)).

In the PROSPER trial, the estimated treatment effect on total cholesterol level at three months in the RCT, adjusted for baseline total cholesterol level was -1.31 mmol/L (95% CI -1.35; -1.27). In a hypothetical regression discontinuity, we used a threshold of 5.0 mmol/L for treatment assignment. The treatment effect estimate in PROSPER was beneficial over the whole range of baseline total cholesterol level, but differed in magnitude (Figure 1b). Analysis with local linear regression showed different treatment effect estimates: -1.04 mmol/L (95% CI -1.16; -0.93) with the cut-off at baseline BP of 5.0 mmol/L and -1.29 mmol/L (95% CI -1.40; -1.18) with the cut-off at baseline BP of 5.5 mmol/L, Table 4).

Table 3. RCT and RD analyses in the preDIVA study (n = 2346)

Analysis	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for two year blood pressure	Standard error (SE) of treatment effect estimate	Design and patients	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for two year blood pressure	Standard error (SE) of treatment effect estimate
		RCT						**RCT**			
Linear adjustment	1173	-	25	-4.0; (-5.4, -2.6)	0.7	*Linear adjustment*	1173	-	1	-3.2 (-5.5, -0.9)	1.2
		RD – assignment on blood pressure[1]						**RD – assignment on age[3]**			
Linear adjustment	1218	965 / 253	21	3.3 (0.1, 6.5)	1.6	*Linear adjustment*	1246	278 / 968	0	-2.4 (-6.0, 1.2)	1.8
RCS adjustment	1218	965 / 253	23	-5.9 (-10.8, -1.0)	2.5	*RCS adjustment*	1246	278 / 968	0	-0.7 (-6.4, 5.1)	3.0
Local linear regression	1218	965 / 253	NA	-9.3 (-18.5, -0.1)	4.7	*Local linear regression*	1246	278 / 968	NA	2.4 (-3.7, 8.5)	3.1
		RD – assignment on blood pressure[2]						**RD – assignment on age[4]**			
Linear adjustment	1201	505 / 696	18	-3.3 (-6.7, 0.2)	1.8	*Linear adjustment*	1176	547 / 629	0	-2.7 (-7.2, 1.7)	2.3
RCS adjustment	1201	505 / 696	22	-5.9 (-11.4, -0.3)	2.8	*RCS adjustment*	1176	547 / 629	0	-2.5 (-9.2, 4.2)	3.4
Local linear regression	1201	505 / 696	NA	-10.2 (-21.0, 0.6)	5.5	*Local linear regression*	1176	547 / 629	NA	-2.8 (-8.2, 2.7)	2.8

Patient selections:

[1] BP ≤ 140 receiving no treatment and BP > 140 receiving treatment

[2] BP ≤ 160 receiving no treatment and BP > 160 receiving treatment

[3] Age ≤ 72 receiving no treatment and age > 72 receiving treatment

[4] Age ≤ 74 receiving no treatment and age > 74 receiving treatment

Table 4. RCT and RD analyses in the PROSPER trial (n=5581)

Analysis	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for three month cholesterol level	Standard error (SE) of treatment effect estimate	Design and patients	N total	N treatment group / control group	R2 (%)	Treatment effect estimate (95% CI) for three month cholesterol level	Standard error (SE) of treatment effect estimate
		RCT						**RCT**			
Linear adjustment	2790	-	75	-1.31 (-1.35, -1.27)	0.021	*Linear adjustment*	2790	-	35	-1.28 (-1.35, -1.22)	0.033
		RD – assignment on cholesterol level[1]						**RD – assignment on age[3]**			
Linear adjustment	2787	2066 / 721	34	-0.99 (-1.06, -0.92)	0.034	*Linear adjustment*	2792	586 / 2206	26	-1.24 (-1.34, -1.13)	0.051
RCS adjustment	2787	2066 / 721	35	-1.14 (-1.25, -1.03)	0.057	*RCS adjustment*	2792	586 / 2206	26	-1.27 (-1.47, -1.08)	0.099
Local linear regression	2787	2066 / 721	NA	-1.04 (-1.16, -0.93)	0.058	*Local linear regression*	2792	586 / 2206	NA	-1.16 (-1.39, -0.93)	0.117
		RD – assignment on cholesterol level[2]						**RD – assignment on age[4]**			
Linear adjustment	2824	1532 / 1292	32	-1.14 (-1.21, -1.07)	0.034	*Linear adjustment*	2818	1175 / 1643	33	-1.22 (-1.33, -1.11)	0.057
RCS adjustment	2824	1532 / 1292	34	-1.28 (-1.40, -1.17)	0.065	*RCS adjustment*	2818	1175 / 1643	33	-1.27 (-1.49, -1.05)	0.112
Local linear regression	2824	1532 / 1292	NA	-1.29 (-1.40, -1.18)	0.055	*Local linear regression*	2818	1175 / 1643	NA	-1.26 (-1.45, -1.07)	0.097

Patient selections:

[1] Cholesterol ≤ 5.0 receiving no treatment and cholesterol > 5.0 receiving treatment

[2] Cholesterol ≤ 5.5 receiving no treatment and cholesterol > 5.5 receiving treatment

[3] Age ≤ 72 receiving no treatment and age > 72 receiving treatment

[4] Age ≤ 74 receiving no treatment and age > 74 receiving treatment

Table 5a. Relative efficiency of global treatment effect estimates in RD design in terms of required sample size (compared to global treatment effect estimate in RCT design) for different validation studies in preDIVA and PROSPER*.

	preDIVA	PROSPER
RCT (linear adjustment) vs RD (RCS adjustment)	6.25 [1]	9.0 [1]
RCT (linear adjustment) vs RD (RCS adjustment)	8.45 [2]	11.52 [2]

**Formula: $(SE_{RD} / SE_{RCT})^2$*

[1] Patient selection age ≤ 72 Tx- and BP > 72 Tx+

[2] Patient selection age ≤ 74 Tx- and BP > 74 Tx+

Table 5b. Relative efficiency of local treatment effect estimates in RD design in terms of required sample size (compared to local treatment effect estimate in RCT design) for different validation studies in preDIVA and PROSPER*.

	preDIVA	PROSPER
RCT (RCS adjustment) vs RD (local linear regression)	3.56 [1]	1.04 [3]
RCT (RCS adjustment) vs RD (local linear regression)	3.78 [2]	0.72 [4]

**Formula: $(SE_{RD} / SE_{RCT})^2$*

[1] Patient selection BP ≤ 140 Tx- and BP > 140 Tx+

[2] Patient selection BP ≤ 160 Tx- and BP > 160 Tx+

[3] Patient selection cholesterol ≤ 5.0 Tx- and cholesterol > 5.0 Tx+

[4] Patient selection cholesterol ≤ 5.5 Tx- and cholesterol > 5.5 Tx+

Analysis with restricted cubic spline adjustment in regression discontinuity in PROSPER showed only slightly different treatment effect estimates compared to the results from the regression discontinuity setting analyzed with local linear regression (Table 4).

In the hypothetical regression discontinuity design in PROSPER with assignment on the age variable, both analyses with local linear regression and normal linear regression with adjustment for age as a restricted cubic spline variable showed similar treatment effect estimates compared to the estimates from the RCT (Table 4). This is consistent with Figure 1d, which shows a constant effect of treatment over the whole range of age.

In terms of statistical precision, the regression discontinuity with restricted cubic spline adjustment was 1 to 4 times less efficient than the adjusted RCT for the local effects estimated, and 6 to 12 times less efficient for the global effects estimated in regression discontinuity compared to the RCT. (Table 5)

DISCUSSION

In this study we performed simulation studies and analyzed data from two large cardiovascular RCTs with the aim to assess the validity of the regression discontinuity design as a prospective quasi-experimental design compared to a traditional RCT. In the

simulations studies we found unbiased but substantially less precise treatment effect estimates from an regression discontinuity design compared to an RCT design.

In the two validation studies we found somewhat different results. In the case of the treatment assignment on baseline BP and baseline cholesterol level we assumed treatment effect heterogeneity over the baseline assignment variable. Therefore we assumed the treatment effect estimates in these regression discontinuity designs to be local. The estimates from the analyses with local linear regression are not comparable with the RCT estimates in terms of bias since these effects are local, in contrast to the global RCT estimates. In terms of statistical precision, the regression discontinuity with restriction cubic spline adjustment was 1 to 4 times less efficient than the adjusted RCT for the local effects estimated.

When treatment was assigned on the age variable, we assumed no interaction between age and treatment. Therefore, the treatment effect estimate for the regression discontinuity analysis was assumed to be a global treatment effect. Estimates from the analyses with flexible functions (restricted cubic splines) to obtain an optimal fit of the adjustment model were comparable to the RCT estimate. In PROSPER the treatment effect estimates were consistent over the different cut-offs, which confirms the assumption of no interaction. In preDIVA, the estimates were quite different over the different cut-offs. This gives the impression that there is interaction and that the treatment effects are local over the baseline age range. Further, we found that sample size needs to be at least 6 times larger to make regression discontinuity as precise as an RCT to estimate global treatment effect estimates.

Position of regression discontinuity design in epidemiology

The main goal of our study was to assess the regression discontinuity design as a prospective design, and compare it with an RCT design. The regression discontinuity design could be implemented as an observational study design, but we focused on the situation of prospective enrollment of patients, with a predefined cut-off. While estimation of treatment effects with retrospectively collected data is could be hampered by selection bias and confounding by indication, which is difficult to fully control for since unmeasured confounding cannot be accounted for ('residual confounding'), in the prospective application of the regression discontinuity design the confounding variable is measured and controlled for by design.

Local or global treatment effect and model specification

A very important question in designing a regression discontinuity study is whether a global or a local treatment effect should be assumed. The assignment variable in an RCT is random allocation and therefor would not interact with treatment in an RCT. This results in a global or average treatment effect. In contrast, the assignment variable in

regression discontinuity could interact with treatment. Therefor the estimated effect is not always a global effect, but a local treatment effect estimate. Only when the assumption of no interaction between the baseline assignment variable and treatment can be made, the estimated effect of regression discontinuity can be interpreted as a global treatment effect. Our analyses with normal linear regression models with restricted cubic spline adjustment assume such a global treatment effect. In our RCT data we could test the assumption of no interaction as we had treatment and control patients across the complete range of the assignment variable. In practice, however, this assumption cannot be tested in regression discontinuity designs, as treatment and control patients have no overlap in the assignment variable.

When the aim is to estimate the treatment effect on a certain threshold value of the assignment variable, local linear regression should be used. This estimate will be unbiased for that certain cut-off, but it is only valid for a small subset of patient around the cut-off and is not generalizable for the whole population.

Thus, if one aims to estimate an overall treatment effect from RD, using a normal regression models is preferred over local linear regression, as supported by our finding that the treatment effect estimates from regression discontinuity with restricted cubic splines more were more similar to the RCT estimate than the local linear regression estimates. However, this could be expected because local linear regression uses only data around the cut-off while normal regression uses all data, as happens in the RCT. In fact the local linear regression and restricted cubic splines estimates cannot be judged as more or less biased. They estimate a different effect (global or local) and their validity is dependent on the assumptions made.

In our case of blood pressure the assumption of no interaction was not met: graphical inspection showed qualitative treatment effect heterogeneity (Figure 1a). The intervention is a multi-component intervention tailored according to patients' risk profile, and the intensity of the intervention (both medical and lifestyle) is thus higher for participants with a higher risk profile.(25) This difference in intensity might explain the interaction of treatment with baseline risk for outcome, i.e. a stronger effect in patients with higher baseline risk. This led to a different treatment effect when we misspecified the adjustment model.

Optimal modeling of the effect of the assignment variable is extremely important in regression discontinuity. We used restricted cubic spline functions and local linear regression. Restricted cubic spline functions are attractive for their flexibility with low degrees of freedom. They and not driven by extreme values in both ends of the fit, which is an advantage over ordinary cubic spline functions.(29) We suggest that flexible functions should be used for optimal adjustment since this function accounts for differential effects of the baseline variable that is used as the assignment variable, on outcome.

Choice of regression discontinuity versus randomized clinical trials

Requirements for informed consent for clinical trials are often more stringent than for treatment outside of the setting of an RCT.(31) Patient enrollment may hence be easier due to avoidance of the randomization process. More lenient inclusion criteria and easier enrollment when using the RD design most likely will result in more representative cohorts for analysis.

A regression discontinuity design is attractive when random assignment of treatment is problematic or not possible. This may occur when a medical intervention is already standard practice for a part of the patients in clinic but the effectiveness has not yet been assessed. For instance, the effect of blood pressure and cholesterol lowering on incident dementia has only been studied in randomized controlled trials using dementia as a secondary outcome and with inconclusive results.(32) At present performing such RCTs is no longer considered ethical, as there is a clinical imperative to treat those with high blood pressure and cholesterol. However, there is circumstantial evidence that there is a beneficial effect of BP reduction on dementia risk, which is not translated to clinical guidelines in the absence of evidence to substantiate this claim. In this situation RD could be a solution to assess a treatment effect using less affected patients for whom an intervention is not deemed indicated as control patients. In fact, because this strategy already closely resembles clinical practice, it may well be feasible to include a number of participants in such a trial that is 6-12 fold higher than in a classical RCT. Adherence to assignment of treatment according to the threshold value is crucial, and both participating clinicians and patients should therefore be well aware that they are participating in a comparative study. A possible threat when using a regression discontinuity design is selection bias near the threshold value. When physicians selectively treat patients just below the threshold value and vice versa, selection bias occurs due to confounding by indication. It is thus very important to avoid such protocol violations in a prospective regression discontinuity study.

Strengths and limitations

A limitation of this study is that in the simulation study we assumed the ideal condition of a linear treatment effect and no residual confounding, which may not reflect real life practice. However, we showed that even in such an ideal setting regression discontinuity is less efficient than RCT. Further, we only studied continuous outcomes and therefore cannot draw conclusions on the performance of the regression discontinuity design for dichotomous outcome parameters. The relative inefficiency may be different for such settings. Furthermore, we assessed the regression discontinuity design in RCT data, in which we artificially set the threshold. This results in perfect adherence to the defined threshold, which is unlikely to occur in real life.

On the other hand, the possibility to change the cut-off was a major strength in our study. We were able to study interaction by varying the cut-off, which would be impossible in a 'real' regression discontinuity design. Moreover, with the RCT data we were able to study the effect of different hypothetical variables to assign treatment, which is a unique feature of this study.

There might be measurement error in the assignment variable. Many claim that regression to the mean caused by such measurement error is a possible threat to the validity of regression discontinuity.(14). A high baseline measurement will on expectation regress down to a lower value and a low baseline measurement will on expectation regress up to a higher value. However, as this will occur equally on both sides of the cut-off in the assignment variable, the measurement error will in the end be irrelevant for the correct estimation of the treatment effect.(21)

Conclusion

We conclude that the regression discontinuity design has perfect theoretical validity and may have reasonable validity in real life situations compared to RCT. Regression discontinuity may provide similar estimates of treatment effects to RCT estimates, but requires the assumption of a global treatment effect over the range of the assignment variable. Controlling for the assignment variable is essential and may be achieved by an optimal fit of the adjustment model, with for example restricted cubic splines when the assumption of a global treatment effect over the range of the assignment variable can be made. Regression discontinuity is, however inefficient, requiring sample sizes which are over 6 times higher than for conventional RCTs to obtain the same statistical precision for a global treatment effect estimate. When considering a study with regression discontinuity versus RCT design, in addition to a risk of bias due to wrong assumptions, researchers need to weigh better recruitment against the substantial loss in precision.

REFERENCES

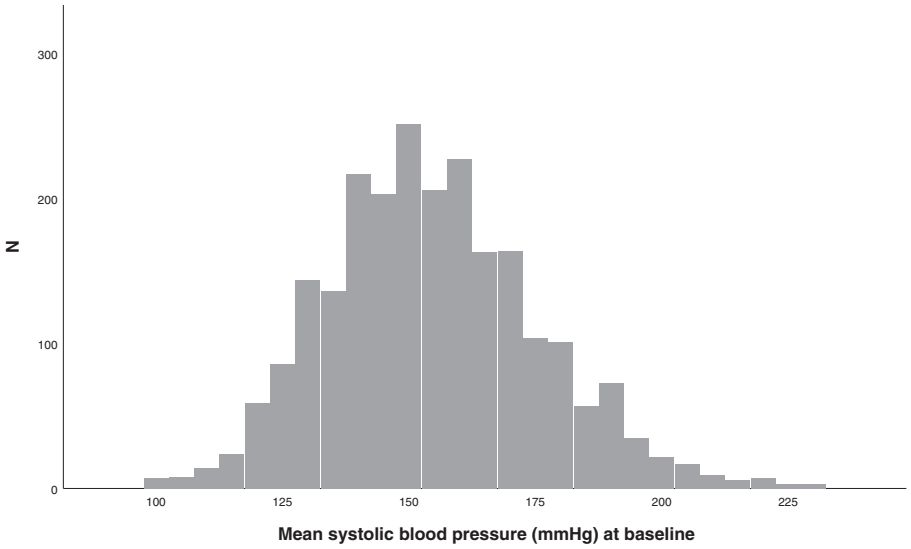
1. Wheeler JL, Segal AD, Ledoux WR, et al. Patient willingness to enroll in a randomized surgical trial. Presented at the 25th Annual Summer Meeting of the American Orthopaedic Foot and Ankle Society. July 15-18, 2009. Vancouver, British Columbia.
2. Jenkins V, Fallowfield L. Reasons for accepting or declining to participate in randomized clinical trials for cancer therapy. Br J Cancer. 2000 Jun;82(11):1783-8.
3. Gotay CC. Accrual to cancer clinical trials: directions from the research literature. Soc Sci Med. 1991;33(5):569-77.
4. Harrison JD, Solomon MJ, Young JM, Meagher A, Hruby G, Salkeld G, et al. Surgical and oncology trials for rectal cancer: who will participate? Surgery. 2007 Jul;142(1):94-101.
5. Abraham NS, Young JM, Solomon MJ. A systematic review of reasons for nonentry of eligible patients into surgical randomized controlled trials. Surgery. 2006 Apr;139(4):469-83.
6. Creel AH, Losina E, Mandl LA, Marx RJ, Mahomed NN, Martin SD, et al. An assessment of willingness to participate in a randomized trial of arthroscopic knee surgery in patients with osteoarthritis. Contemp Clin Trials. 2005 Apr;26(2):169-78.
7. Chang BH, Hendricks AM, Slawsky MT, Locastro JS. Patient recruitment to a randomized clinical trial of behavioral therapy for chronic heart failure. BMC Med Res Methodol. 2004 Apr 17;4:8.
8. Martinson BC, Crain AL, Sherwood NE, Hayes MG, Pronk NP, O'Connor PJ. Population reach and recruitment bias in a maintenance RCT in physically active older adults. J Phys Act Health. 2010 Jan;7(1):127-35.
9. Vozdolska R, Sano M, Aisen P, Edland SD. The net effect of alternative allocation ratios on recruitment time and trial cost. Clin Trials. 2009 Apr;6(2):126-32.
10. Halpern SD. Prospective preference assessment: a method to enhance the ethics and efficiency of randomized controlled trials. Control Clin Trials. 2002 Jun;23(3):274-88.
11. Institute of Medicine Forum on Drug Discovery D, Translation. 2010 Opportunities: Workshop Summary.
12. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? Ann Intern Med. 2002 Jul 2;137(1):10-6.
13. Beckie TM, Mendonca MA, Fletcher GF, Schocken DD, Evans ME, Banks SM. Examining the challenges of recruiting women into a cardiac rehabilitation clinical trial. J Cardiopulm Rehabil Prev. 2009 Jan-Feb;29(1):13-21; quiz 2-3.
14. Shadish WR, Cook TD, Campbell DT. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton-Mifflin; 2002. p 229.
15. Cook TD, Shadish WR, Wong VC. Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. J Policy Anal Manag. 2008 Fal;27(4):724-50.
16. Wong VC, Cook TD, Barnett WS, Jung K. An effectiveness-based evaluation of five state prekindergarten programs. J Policy Anal Manag. 2008 Win;27(1):122-54.
17. Aiken LS, West SG, Schwalm DE, Carroll JL, Hsiung S. Comparison of a randomized and two quasi-experimental designs in a single outcome evaluation - Efficacy of a university-level remedial writing program. Evaluation Rev. 1998 Apr;22(2):207-44.
18. Goldberger A.S. Selection bias in evaluating treatment effects: Some formal illustrations. Working Paper, Economics Department, University of Wisconsin. 1972.
19. Buddelmeyer H, Skoufias E. An evaluation of the Performance of Regression Discontinuity Design on PROGRESA. World Bank Policy Research. Aug 2004.

20. Black D, Galdo J, Smith J. Estimating the selection bias of the regression discontinuity design using experimental data. Working paper, University of Chicago. 2007.
21. Vandenbroucke JP, le Cessie S. Commentary: regression discontinuity design: let's give it a try to evaluate medical and public health interventions. Epidemiology. 2014 Sep;25(5):738-41.
22. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014 Sep;25(5):729-37.
23. O'Keeffe AG, Geneletti S, Baio G, Sharples LD, Nazareth I, Petersen I. Regression discontinuity designs: an approach to the evaluation of treatment efficacy in primary care using observational data. BMJ. 2014;349:g5293.
24. Moscoe E, Bor J, Barnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015 Feb;68(2):122-33.
25. Richard E, Van den Heuvel E, Moll van Charante EP, Achthoven L, Vermeulen M, Bindels PJ, et al. Prevention of dementia by intensive vascular care (PreDIVA): a cluster-randomized trial in progress. Alzheimer Dis Assoc Disord. 2009 Jul-Sep;23(3):198-204.
26. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. Lancet. 2002 Nov 23;360(9346):1623-30.
27. White HD, Simes RJ, Anderson NE, Hankey GJ, Watson JD, Hunt D, et al. Pravastatin therapy and the risk of stroke. N Engl J Med. 2000 Aug 3;343(5):317-26.
28. Port S, Demer L, Jennrich R, Walter D, Garfinkel A. Systolic blood pressure and mortality. Lancet. 2000 Jan 15;355(9199):175-80.
29. Steyerberg EW. Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.
30. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst. 1988 Oct 5;80(15):1198-202.
31. Ellis PM, Butow PN, Tattersall MH. Informing breast cancer patients about clinical trials: a randomized clinical trial of an educational booklet. Ann Oncol. 2002 Sep;13(9):1414-23.
32. Peters R, Beckett N, Forette F, Tuomilehto J, Clarke R, Ritchie C, et al. Incident dementia and blood pressure lowering in the Hypertension in the Very Elderly Trial cognitive function assessment (HYVET-COG): a double-blind, placebo controlled trial. Lancet Neurol. 2008 Aug;7(8):683-9.

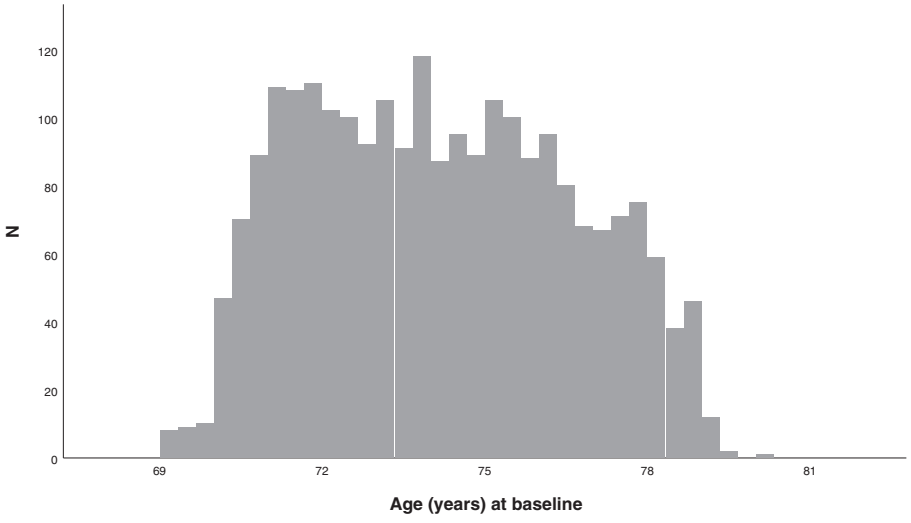
APPENDIX 1: R CODE.

```
# Activation of required libraries
library(rms)
library(mvtnorm)
library(arm)
library(rdd)

# Hypothetical dataset
n_patients          <- 100
R2          <- 0.0
corr        <- sqrt(R2)
treatment_effect <- -10
n_sim     <- 10000
Est_Effect          <- matrix(nrow = n_sim, ncol=6)
Std_error           <- matrix(nrow = n_sim, ncol=6)
DF          <- matrix(nrow = n_sim, ncol=6)

# Variance of prognosis (sigma1), outcome (sigma2)
sigma1   <- 4
sigma2   <- 100
#Covariantie prognosis and outcome
sigma12 <- corr * sqrt(sigma1) * sqrt(sigma2)

# Mean of prognosis and outcome
mu          <- c(10, 90)
# Covariance Matrix
sigma     <- matrix(c(sigma1, sigma12, sigma12, sigma2), nrow = 2, byrow = TRUE)

for(i in 1:n_sim){

dataset   <- rmvnorm(n_patients, mean = mu, sigma = sigma)

dataset   <- as.data.frame(dataset)
names(dataset)    <- c("prognosis", "outcome")

## Randomized Controlled Trial, all patients randomized

# Treatment "Randomize all patients"
```

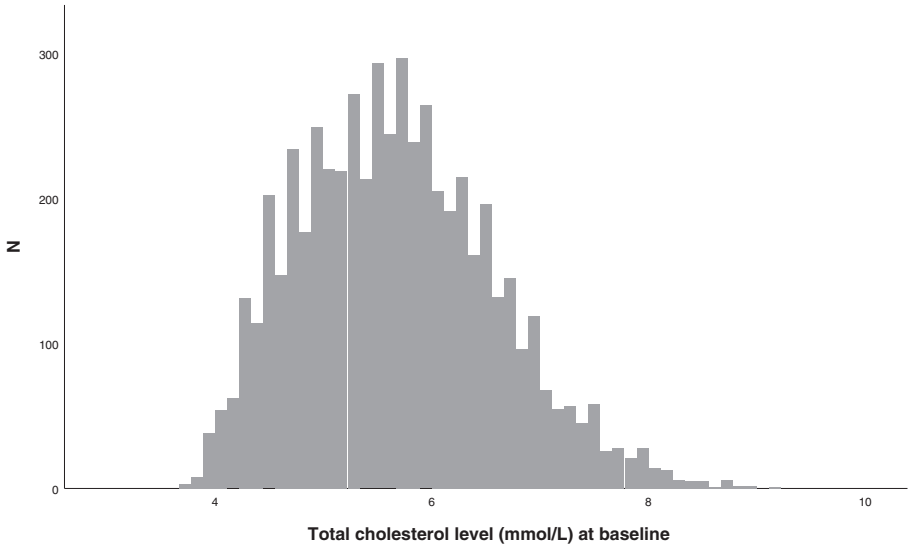
```
dataset$T_RCT    <- as.numeric(runif(n_patients) <= 0.5)

# Outcome "Randomize all patients"
dataset$O_RCT    <- dataset$outcome
dataset$O_RCT[dataset$T_RCT == 1] <- dataset$outcome[dataset$T_RCT == 1] + treat-
ment_effect

## Regression discontinuity, good prognosis control, poor prognosis treatment

#Treatment "Regression discontinuity design"
dataset$T_RDC    <- as.numeric(dataset$prognosis>10)

#Outcome "Regression discontinuity design"
dataset$O_RDC    <- dataset$outcome
dataset$O_RDC[dataset$T_RDC  ==  1]  <-  dataset$outcome[dataset$T_RDC  ==  1]  +
treatment_effect

fit_RCT   <- lm(O_RCT ~ prognosis + T_RCT, data = dataset)
fit_RDC   <- lm(O_RDC ~ prognosis + T_RDC, data = dataset)

fit_RCT_rcs          <- ols(O_RCT ~ rcs(prognosis) + T_RCT, data = dataset, x=T, y=T)
fit_RDC_rcs          <- ols(O_RDC ~ rcs(prognosis) + T_RDC, data = dataset, x=T, y=T)

fit_RCT_llr          <- RDestimate(O_RCT ~ prognosis, cutpoint = 10, data = dataset)
fit_RDC_llr          <- RDestimate(O_RDC ~ prognosis, cutpoint = 10, data = dataset)

Est_Effect[i, ]      <-     c(fit_RCT$coefficients[3],     fit_RDC$coefficients[3],     fit_RCT_
rcs$coefficients[6], fit_RDC_rcs$coefficients[6], fit_RCT_llr$est[1], fit_RDC_llr$est[1])

}

#Mean Effect estimate of treatment
colMeans(Est_Effect)
#Mean squared error of effect estimate treatment
colMeans((Est_Effect - treatment_effect)^2)
```



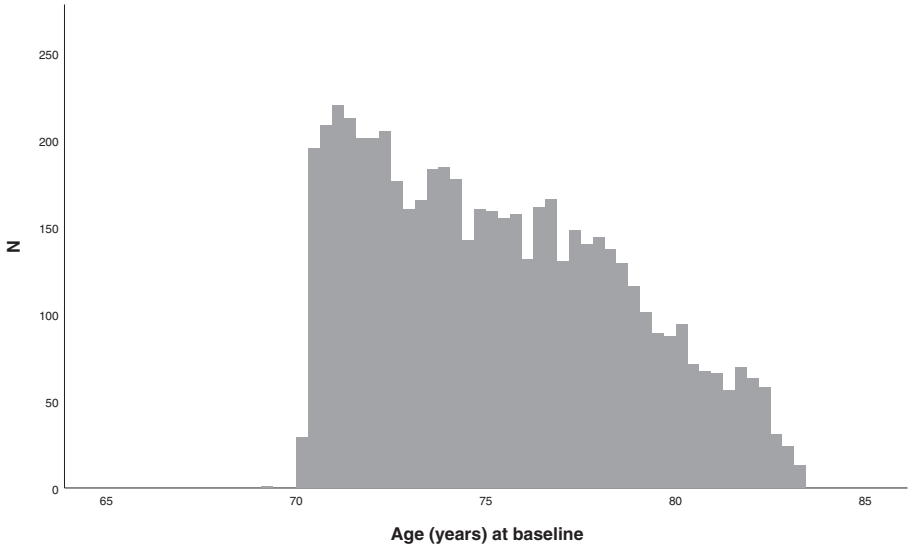
a) Histogram of mean systolic blood pressure (mmHg) at baseline in PreDIVA.



b) Histogram of age (years) at baseline in PreDIVA.



c) Histogram of total cholesterol level (mmol/L) at baseline in PROSPER.



d) Histogram of age (years) at baseline in PROSPER.

Appendix 2. Histogram of the assignment variable in both preDIVA and PROSPER in RCT data.