Regression discontinuity was a valid design for dichotomous outcomes in three randomized trials

Nikki van Leeuwen (MSc)[1], Hester F. Lingsma (PhD)[1], Simon P. Mooijaart (MD, PhD)[2,3], Daan Nieboer (MSc)[1], Stella Trompet[2], Ewout W. Steyerberg (PhD)[1, 4]

[1] Center for Medical Decision Making, Department of Public Health, Erasmus Medical Center, PO Box 2040, 3000 CA, Rotterdam, The Netherlands

[2] Department of Gerontology and Geriatrics, Leiden University Medical Centre, Leiden, The Netherlands

[3] Institute for Evidence-Based Medicine in Old Age, Leiden, The Netherlands

[4] Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, Leiden, The Netherlands

Date: January 31, 2018

Correspondence: Nikki van Leeuwen

PO Box 2040

3000 CA Rotterdam

The Netherlands

Phone: +31 (0)6 17126187

Fax: +31 (0)10 7038456

E-mail: n.vanleeuwen.1@erasmusmc.nl

Reprints: No reprints will be supplied

Word Count: Abstract: 203 words

Paper: 3517 words

Abstract

Regression discontinuity (RD) is a quasi-experimental design that may provide valid estimates of treatment effects in case of continuous outcomes. We aimed to evaluate validity and precision in the RD design for dichotomous outcomes.

We performed validation studies in three large RCTs (CRASH, GUSTO and PROSPER). To mimic the RD design, we selected patients above and below a cut-off (e.g. age 75 years) randomized to treatment and control respectively. Adjusted logistic regression models using restricted cubic splines (RCS) and polynomials, and local logistic regression models estimated the odds ratio (OR) for treatment, with 95% confidence intervals to indicate precision.

In CRASH, treatment increased mortality with OR 1.22 [95% CI 1.06; 1.40] in the RCT. The RD estimates were 1.42 [0.94; 2.16] and 1.13 [0.90; 1.40] with RCS adjustment and local regression respectively. In GUSTO, treatment reduced mortality (OR 0.83 [0.72; 0.95]), with more extreme estimates in the RD analysis (OR 0.57 [0.35; 0.92] and 0.67 [0.51; 0.86] respectively). In PROSPER, similar RCT and RD estimates were found, again with less precision in RD designs.

We conclude that the RD design provides similar but substantially less precise treatment effect estimates compared to an RCT, with local regression being the preferred method of analysis.

Keywords: Regression discontinuity design, quasi-experimental trials, trial design, causal inference, logistic regression, restricted cubic splines, polynomials, local logistic regression

Abbreviations

RD = Regression Discontinuity design

RCT = Randomized Controlled Trial

PROSPER = PROspective Study of Pravastatin in elderly individuals at risk of vascular disease

CRASH = Corticosteroid Randomisation After Significant Head injury

GUSTO = The Global Utilization of Streptokinase and Tissue plasminogen activator for Occluded coronary arteries

CI = Confidence Interval

What is new?

- RD design provides similar but substantially less precise treatment effect estimates compared to an RCT in dichotomous outcomes
- local regression is the preferred method of analysis when using an RD design with dichotomous outcomes
- global treatment effect estimates from RD designs should only be presented secondary to local average treatment effect estimates and never as the primary parameter of interest
- a strength of this study is the use of data from three large RCTs to be able to compare the RD results with the RCT estimates and therefore we were able to carefully assess interaction between the assignment variable and treatment
- our results suggest when there is no interaction between the assignment variable and treatment – and thus a global treatment effect can be estimated – the results from the RCS adjusted analyses and local logistic regression are more similar to each other than when there is interaction

Introduction

Randomized clinical trials (RCTs) provide the most reliable evidence of effectiveness of medical interventions.[1] Nevertheless, recruitment of sufficient numbers of patients is a challenge in RCTs; it is estimated that less than 50% of the RCTs meet their recruitment targets.[2,3] Patients' treatment preferences and clinicians equipoise are often cited as barriers to recruitment in RCTs.[2,4,5,6,7]. Patients participating in trials may poorly represent the population of interest.[8,9] Especially, under-representation of older participants and women is well known in RCTs.[8,10]

The quasi-experimental "regression discontinuity" (RD) design is an alternative epidemiological design to assess effectiveness of treatment. It has been suggested that RD is the observational design that most resembles an RCT.[11,12] In the RD design, treatment is not assigned randomly, but is allocated to a subset of patients, based on a baseline assignment variable, often related to the outcome. The control group consists of a complementary subset of patients, not receiving treatment. E.g. all patients with an age over 75 years receive treatment and patients with an age below 75 years do not receive treatment and are considered as the control group. Such treatment assignment method may closely resemble clinical practice and may thus facilitate patient inclusion. In the analysis of the treatment effect, a regression model is used to compare treatment to the control group, while adjusting for the treatment assignment variable, in this example age.

The RD design is attractive because some of the challenges of the randomization process are avoided. However, the estimates from this quasi-experimental design may be substantially less efficient compared to an RCT.[13] The validity of RD estimates on

continuous outcomes are well studied[13,14,15], but the validity of the RD design with binary outcomes is less known. Only a few examples have been described before[16,17], while many health outcomes are dichotomous. Moreover, the efficiency of modeling approaches is unclear, i.e. the precision of estimated treatment effects. The aim of this study was to assess validity and precision of the RD design in studies with dichotomous outcome compared to an RCT. We hereto analyzed data from three large RCTs.

Methods

Patients

Three trials were used to validate the RD design in empirical data. To assess the internal validity of the RD design we compared RD estimates with the estimates resulting from the RCT data. For the RD design we used a continuous baseline variable as assignment variable and the dichotomous endpoints of the RCTs.

The "Corticosteroid Randomisation After Significant Head injury" (CRASH) trial studied the effect of corticosteroids on death and disability after head injury.[18] CRASH enrolled 10,008 patients between 1999 and 2005. The primary outcome in CRASH was 14-day mortality. We included 9,554 patients with complete outcome data of whom 2,323 died before 14 days (24%). The median age was 33 years (IQR: 23 – 47 years).

Second, we analyzed 30,510 patients from "The Global Utilization of Streptokinase and Tissue plasminogen activator for Occluded coronary arteries" trial (GUSTO). Patients were entered between 1990 and 1993. 10,348 patients were assigned to treatment (accelerated tissue plasminogen activator, t-PA) and 20,162 patients were used as control patients receiving streptokinase.[19] The primary outcome was 30-day mortality. The median age was 61 (IQR: 52 – 69) and mortality occurred in 2,128 (7%). For both CRASH and GUSTO, age was used as the treatment allocation variable.

Third, we analyzed data from "PROspective Study of Pravastatin in elderly individuals at risk of vascular disease" (PROSPER).[20] This study enrolled 5,804 patients between December 1997 and May 1999, who were assigned to pravastatin (n = 2,891) or placebo (n = 2,913) to reduce the risk of coronary disease in elderly individuals. The

outcome was a composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke at 3.2 years on average after randomization. 881 (15%) of the patients experienced the composite endpoint. The median total cholesterol level was 5.6 mmol/L (IQR: 5.0 – 6.3 mmol/L) at baseline (Table 1). For PROSPER, we considered baseline total cholesterol as the treatment allocation variable.

Statistical analysis

To analyze the data as an RD design, we selected those patients with a baseline value above the median of the assignment variable, who were assigned to treatment in the original RCT as the intervention group, and those with a baseline value below the median and not assigned to treatment in the RCT as control group. Histograms of the baseline assignment variables for each study were plotted, as well as binned scatterplots for outcome means for treated and controls at each baseline assignment value. The analysis was based on the intention-to-treat principle. This led to inclusion of approximately half of the RCT patients. The treatment effect was expressed as odds ratios (OR) with 95% confidence intervals (95% CI), with adjustment for the baseline variable in a logistic regression model. To compare the RD estimates to the RCT estimates in comparable sample sizes, random samples of 50% from the complete RCT data were drawn (5000 times). To compare the designs in terms of efficiency we calculated the ratio of variances between both designs based on estimated standard errors (SEs) of the estimated treatment effects: $(SE_{design\ 2} / SE_{design\ 1})^2$.
Previous work has shown that the validity of the RD design is highly dependent on the quality of the adjustment in the analysis phase, and on assumptions of a local or global

effect of the treatment.[13] All analyses (RCT and RD) were adjusted for the baseline variable that was used to attribute treatment; age in both CRASH and GUSTO and baseline cholesterol in the PROSPER trial. We assessed non-linearity of the effect of the baseline variable with non-linear restricted cubic splines (RCS) functions. An RCS function is a smooth function that consists of pieced-together cubic splines that are restricted to be linear in the tails. We used three knots for adequate flexibility.[21] Consequently we used the RCS of the baseline variable in the adjustment model for optimal adjustment. To consider a different approach to estimate RD estimates, we also used polynomials of the baseline variables in the adjustment model. R^2 statistics were calculated to indicate the explained variance of the adjustment model.

The approach described above assumes a global treatment effect. It has been argued that this assumption is hard to make and can never be proven.[11] We therefore also analyzed the RD design with local logistic regression models. In local logistic regression, only patients around the cut-off were used in the analysis to estimate the treatment effect. For the local estimations, the *gam* package in R was used, in which a default span of 0.5 is set. Gaussian kernel was used for the local logistic regression analysis. Using this kernel, the observations outside the span have lower influence on the estimation, but all the data are used in the analysis. To assess differential treatment effects, we studied interaction between the baseline variable and the treatment in the RCT data. For all three trials we assessed treatment effect heterogeneity in the complete RCT data, using interaction terms between treatment and the assignment variable. Moreover, to study the stability of the estimates for all three validation studies, we added RD analyses on an additional cut-off.

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

Results

In CRASH the treatment was harmful. The adjusted OR was 1.22 [95% CI: 1.06; 1.40] for the effect of treatment on mortality in the 50% subset of the RCT. For the hypothetical RD design, the estimated OR was 1.42 [0.94; 2.16], with RCS adjustment for age. When analyzed with polynomial adjustment the OR for treatment was 1.09 [0.81; 1.46]. The alternative method to analyze this hypothetical RD design, local logistic regression, resulted in an estimated OR of 1.13 [0.90; 1.40] (Table 2).

In GUSTO the estimated OR for mortality was 0.83 [0.72; 0.95] in a subset of 50% of the patients. The estimated OR, in the RD scenario was 0.57 [0.35; 0.92] adjusted with RCS for age. The OR for treatment from RD estimated with polynomial adjustment for age was 0.82 [0.63; 1.07]. The analysis with local logistic regression resulted in an estimated OR of 0.67 [0.51; 0.86] (Table 3).

In the PROSPER trial, the adjusted OR for the composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke was 0.85 [95% CI; 0.69; 1.04] when assessed in the subset of 50% of the RCT. The estimated OR was 0.80 [0.46; 1.38] in the hypothetical RD design adjusted for baseline cholesterol with RCS. The OR for treatment from RD estimated with polynomial adjustment for age was 0.81 [0.56; 1.16]. The RD design analyzed with local logistic regression showed an OR for treatment of 0.79 [0.56; 1.13] (Table 4).

In none of the RCTs we found statistically significant interaction between treatment and the assignment variable. However, this interaction test has limited statistical power. In all three trials there appeared to be a differential treatment effect over the range of the assignment variable, (Figure 1d, 2d and 3d). This is confirmed in the additional RD analysis with treatment assignment based on a different cut-off (Table 2, 3 and 4). In these validation studies we see slightly different RD estimates between the two different assignment approaches in all three studies.

In terms of efficiency, the RD with adjustment was 7.2 to 12.1 times less efficient than the adjusted RCT, compared to 3.1 to 4.5 less efficient estimates from RD with polynomial adjustment. The RD design analyzed with local logistic regression was 2.5 to 3.5 times less efficient than the adjusted RCT (Table 5).

Discussion

This validation study, with data from three large RCTs, showed that the treatment effect estimates from the hypothetical RD were similar to the treatment effect estimates from the RCTs, either with RCS and polynomial adjustment or local logistic regression. In all three studies the confidence interval of all RD estimates overlapped with the point estimate of the RCT. However, RD estimates were substantially less precise.

Causality in regression discontinuity design

The advantage of a quasi-experimental, prospective, RD design over an observational study is the controlled assignment of treatment. This property is shared with an RCT. As Labrecque et al. stated, in both an RCT as in an RD design, we have good knowledge of the assignment mechanism.[11] In RCTs, treatment is randomly allocated and in RD treatment is assigned to patients using a baseline assignment variable. The treated and untreated patients in an RCT are unconditionally exchangeable. Therefore, RCTs are accepted to make causal inference. In an RD design the treated and the control patients are not exchangeable across the whole baseline range since they have a systematically different baseline characteristic. In RD the treated and untreated are only exchangeable close to the cut-off of the baseline assignment variable.[11,12] Therefore, causal inference can only be made around the cut-off in an RD design, where patients can be considered to be exchangeable. The causal treatment effect estimated in RD is a local treatment effect estimate. This means that comparing estimates from RCT and RD may not be completely straightforward, even with comparable RCT and RD data.[11] Therefor it may not be entirely fair to interpret the concordance between local RD estimates and global RCT estimates as a measure of validity of RD estimates. The overall RCT estimate is

253 the average treatment effect in the whole RCT population, although we can condition on
254 the assignment variable for more efficient analysis.[22-25] An RD estimate is a local
255 treatment effect among patients at the cut-off and may vary dependent on the cut-off
256 chosen.[13] At the end of the day, it is the RCT estimate that is the average of local
257 estimates across the distribution of the assignment variable.

258

259 *Global vs. local treatment effects*

260 Only when treatment does not interact with the baseline assignment variable the
261 estimate from an RD design can be interpreted as a global treatment effect estimate.[11]
262 In order to estimate a global treatment effect estimate in RD, one would have to feel
263 confident modeling the relationship between the assignment variable and the outcome
264 even where it is not observed in the data.[11,26,27] In other words, the model for the
265 assignment variable–outcome relationship in both the treated and untreated groups
266 would have to be extrapolated to the side of the cutoff where they were not observed.[11]

267 When using RCS or polynomial adjustment, the treatment effect in CRASH was slightly
268 different compared to the RCT. Graphical inspection showed qualitative interaction
269 between treatment and the adjustment variable age (Figure 1d). At the cut-off (age 33
270 years) the treatment effect – the difference between the plotted line for the control
271 patients and the plotted line for the treated patients – was larger than the global RCT
272 effect which is shown in Figure 1a. This explains the difference between the RD
273 estimate and the RCT. The presence of a heterogeneous treatment effect over the
274 range of age was confirmed in the RD analysis with treatment based on a different cut-
275 off, resulting in less similar treatment estimates compared to the RCT estimates.

Qualitative interaction was also observed in GUSTO (Figure 2d), and could have led to more extreme RD estimates (0.57 and 0.67) compared to the OR estimated in the RCT (0.83). At the cut-off of 62 years in Figure 2d a larger treatment effect is shown compared to the global treatment effect in Figure 2a. However, in RD with polynomial adjustment for age, the treatment effect is similar (0.82) to the RCT estimate. A smaller treatment effect was estimated when the cut-off for treatment assignment was set at 70 years. This is also confirmed in Figure 2d; after the age of 62 the treatment effect decreases.

In PROSPER, also qualitative interaction was found and shown in Figure 3d. However, RD with treatment assignment set at cholesterol 5.6 mmol/L, RD estimates (0.80, 0.81 and 0.79) and the RCT estimate (0.85) were quite similar. When the treatment assignment rule was set at cholesterol 6.2 mmol/L for RD, the RD estimates were slightly different from the RCT estimate. These results confirm that the RD estimate is not equal to the global RCT treatment effect estimate when the treatment effect is heterogeneous across the baseline assignment variable.

In a prospective RD design, it is hard to know whether the baseline assignment variable interacts with treatment. It can be formally tested but since the treatment groups each have data on only one side of the cut-off, the result only represents possible interaction at a small range of the assignment variable, around the cut-off. Thus, the assumptions required to estimate the global treatment effect are not verifiable within the RD design. This is why it has been suggested that global treatment effect estimates from RD designs should only be presented secondary to local average treatment effect estimates and never as the primary parameter of interest.[11,12]

In this study we also assessed and compared RCS and polynomials for adjustment in RD. The advantage of an RCS function over polynomial adjustment is the restriction of the function to be linear in the tails. This is important when using this for optimal adjustment in for example RCTs, to estimate global effects over the whole range of the population studied. However, in RD we are primarily interested in local estimates and thus optimal adjustment around the cut-off for treatment assignment. So the advantage of RCS spline functions over polynomial adjustment in for example RCTs, may be less applicable to optimal adjustment in RD.

Our results suggest when there is no interaction between the assignment variable and treatment – and thus a global treatment effect can be estimated – the results from the RCS and polynomial adjusted analyses and local logistic regression are more similar to each other than when there is interaction. If there was some interaction between the assignment variable and treatment, the results from local logistic regression and the RCS and polynomial adjusted analyses were less similar. So, the comparison of both RD estimates could be a way to have more information on the assignment variable – treatment relationship.

Efficiency of RD design

The RD estimates with adjustment appeared to be substantially less efficient than the RCT estimates. An RD design analyzed with adjusted logistic regression using RCS adjustment implies that 7.2 to 12.1 times more patients need to be included in the study compared to an RCT design. RD with polynomial adjustment would need 3.1 to 4.5 more patients compared to an RCT. If one would analyze the RD design with local

logistic regression, this study would need about 2.4 to 3.6 times more patients than an RCT. So, the local regression approach was more efficient compared to the adjusted logistic regression. Also in terms of efficiency, local logistic regression would be preferred to analyze an RD design.

In absolute numbers an RD design needs more patients to obtain similar efficiency, compared to an RCT to estimate global treatment effect estimates. Although RD is described as less efficient than RCT in identifying the global average causal effect, it may be nearly as good in identifying the local causal effect, which may be of interest depending on the context. From a power perspective, it would be a fair comparison if the RCTs were powered to estimate treatment effects in the assignment variable subgroups around the discontinuity and compare these with the local RD treatment effect estimates. However, in our study we focus on the comparison between global RCT estimates and estimates from an RD design, and the efficiency of an RD design to estimate global treatment effect estimates.

Also, an RD design could facilitate patient recruitment, especially when the cut-off for treatment assignment closely resembles clinical practice. In these specific cases an RD design may be cheaper and less-time intensive than an RCT. Besides, RD designs could be conducted in different settings than RCTs; one can assume that RD design have less stringent inclusion criteria. This would be especially the case in a retrospective RD design when data from (clinical) registries are used. Therefore, some argue that data used in RD designs could lead to more external validity.[28,29]

Linear versus logistic models in RD

In this study we specifically assess the performance of RD vs RCT in the context of dichotomous outcomes and logistic regression, which is not the standard in RD designs, but is common in health research. RD is underused with logistic regression models; only a few examples are described before.[16,17] RD can be easily extended to generalized linear models like logistic regression.[30] When using dichotomous endpoints in RD it is straightforward to obtain more interpretable parameters like risk differences and risk ratios even in the logistic regression context, because the predicted probabilities at the threshold can be obtained directly from the model. The only barrier using logistic models in RD would be the absence of a data driven optimal bandwidth selector for the logistic model, like Imbens-Kalyanamaran[31] optimal bandwidth calculation is available for local linear regression models. For the local estimations in this study the *gam* package in R was used, in which a default span a 0.5 proportion of the observations over the assignment range is included. This can be adjusted specifying "span" in the gam function, for example span=0.2. When one is interested in a local treatment effect estimate, extending the span will in theory decrease validity but also increase reliability.

Strengths and limitations

We used RCT data to evaluate a hypothetical RD design, in which we artificially set the cut-off to "assign" treatment. This resulted in perfect adherence to the defined cut-off. This is unlikely to be the case in real life where which patients are prospectively assigned to treatment. A strength of this study is the use of data from three large RCTs to be able to compare the RD results with the RCT estimates. Moreover, because of the

RCT data we were able to carefully assess interaction between the assignment variable and treatment.

Conclusion and recommendations

Our findings for dichotomous outcomes are in line with previous work on RD for continuous outcomes.[13] The RD design may provide similar treatment effect estimates compared to RCT estimates for dichotomous outcome measures, but has some strong disadvantages that should be carefully considered when choosing an RD design to assess the effectiveness of a medical intervention. First, to be able to estimate the same global treatment effect in an RD design as in an RCT, the assumption of a global treatment effect over the range of the assignment variable is required. In prospectively collected RD data this assumption of a global treatment effect cannot be proven. Global treatment effect estimates from RD designs should therefore only be reported secondary to local treatment effect estimates. Second, the RD design is substantially less efficient than an RCT, requiring sample sizes at least three times higher than for the conventional RCT to obtain the same precision for the treatment effect estimate. In this study we found local logistic regression would be most efficient to analyze an RD design. Future research on the RD design should focus on more efficient application of the RD design, considering different approaches to estimate treatment effects from an RD design and examining their properties.

Acknowledgements

The authors would like to thank A.J. (Ton) de Craen for his insightful comments regarding the design of this paper, before he passed away in January 2016.

We would like to thank P. (Pablo) Perel and I. (Ian) Roberts for providing the CRASH data for this study.

None of the authors reports any conflict of interest.

The Institute for Evidence-Based Medicine in Old Age (IEMO) is funded by the Dutch Ministry of Health and Welfare and supported by ZonMw (project number 62700.3002).

References

[1] McDonald AM, Knight RC, Campbell MK, Entwistle VA, Grant AM, Cook JA, et al. What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. Trials. 2006;7:9.

[2] Mills N, Blazeby JM, Hamdy FC, Neal DE, Campbell B, Wilson C, et al. Training recruiters to randomized trials to facilitate recruitment and informed consent by exploring patients' treatment preferences. Trials. 2014;15:323.

[3] Treweek S, Mitchell E, Pitkethly M, Cook J, Kjeldstrom M, Taskila T, et al. Strategies to improve recruitment to randomised controlled trials. Cochrane Database Syst Rev. 2010(1):MR000013.

[4] Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. J Clin Epidemiol. 1999 Dec;52(12):1143-56.

[5] Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, et al. Barriers to participation in clinical trials of cancer: a meta-analysis and systematic review of patient-reported factors. Lancet Oncol. 2006 Feb;7(2):141-8.

[6] King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, et al. Impact of participant and physician intervention preferences on randomized trials: a systematic review. JAMA. 2005 Mar 2;293(9):1089-99.

[7] Fayter D, McDaid C, Eastwood A. A systematic review highlights threats to validity in studies of barriers to cancer trial participation. J Clin Epidemiol. 2007 Oct;60(10):990-1001.

[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] 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.

[10] 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.

[11] Labrecque JA, Kaufman JS. Commentary: Can a quasi-experimental design be a better idea than an experimental one? Epidemiology. 2016 Jul;27(4):500-2.

[12] Lee DS, Lemieux T. Regression discontinuity designs in economics. J Econ Lit. 2010;48:281–355.

[13] Van Leeuwen N, Lingsma HF, De Craen AJ, Nieboer D, Mooijaart SP, Richard E, Steyerberg EW. Regression discontinuity design: simulation and application in two cardiovascular trials with continuous outcomes. Epidemiology. 2016 Jul;27(4):503-11.

[14] Maas IL, Nolte S, Walter OB, Berger T, Hautzinger M, Hohagen F, Lutz W, Meyer B, Schröder J, Späth C, Klein JP, Moritz S, Rose M. The regression discontinuity design showed to be a valid alternative to a randomized controlled trial for estimating treatment effects. J Clin Epidemiol. 2017;82:94-102.

[15] Hansen H, Klejnstrup NR, Andersen OW. A Comparison of Model-Based and Design-Based Impact Evaluations of Interventions in Developing Countries. Am J Eval. 2013;34(3):320-338.

[16] Berk RA, De Leeuw J. An Evaluation of California's Inmate Classification System Using a Generalized Regression Discontinuity Design. JASA. 1999;94(448):1045-1052.

[17] Imbens GW, Lemieux T. Regression discontinuity designs: A guide to practice. J Econom. 2008;142(2):615-635.

[18] Edwards P, Farrell B, Lomas G, Mashru R, Ritchie N, Roberts I, et al. The MRC CRASH Trial: study design, baseline data, and outcome in 1000 randomised patients in the pilot phase. Emerg Med J. 2002 Nov;19(6):510-4.

[19] An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO investigators. N Engl J Med. 1993 Sep 2;329(10):673-82.

[20] 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.

[21] 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).

[22] Hernandez AV, Steyerberg EW, Butcher I., Mushkudiani N, Taylor GS, Murray GD, Marmarou A, Choi SC, Lu J, Habbema JD, Maas AI. Adjustment for strong predictors of outcome in traumatic brain injury trials: 25% reduction in sample size requirements in the IMPACT study. J.Neurotrauma. 2006;23(9):1295-303.

[23] Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J.Clin Epidemiol. 2004;57(5):454-60.

[24] Steyerberg EW, Bossuyt PM, Lee KL. Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? Am Heart J. 2000;139(5):745-751.

[25] Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression models. Int Stat Rev.1991;58(2):227-240.

[26] Bor J, Moscoe E, Bärnighausen T. Three approaches to causal inference in regression discontinuity designs. Epidemiology. 2015;26:e28–30; discussion e30.

[27] Rubin DB. Assignment to Treatment Group on the Basis of a Covariate. Journal of Educational and Behavioral statistics. 1977;6:377–401.

[28] Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol. 2015;68(2):122-33.

[29] Venkataramani AS, Bor J, Jena AB. Regression discontinuity designs in healthcare research. BMJ. 2016;352:i1216

[30] Bor J, Moscoe E, Mutevedzi P, Newell ML, Bärnighausen T. Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology. 2014;25(5):729-37.

[31] Imbens G, Kalyanaraman K, Optimal Bandwidth Choice for the Regression Discontinuity Estimator. https://scholar.harvard.edu/files/imbens/files/rd_09feb3.pdf.

Table 1. Patient characteristics of CRASH (n = 9554), GUSTO (n = 30,510) and PROSPER (n = 5804).

Characteristic	CRASH	GUSTO	PROSPER
N in treatment arm (%)	4800 (50)	10348 (34)	2891 (50)
N in control arm (%)	4454 (50)	20162 (66)	2913 (50)
Median (IQR) of baseline variable for treatment assignment*	33 (23 - 47) *years*	61 (52 – 69) *years*	5.6 (5.0 – 6.3) *mmol/L*
N outcome (%)**	2323 (24)	2128 (7)	881 (15)

** Baseline measurement is age in years in CRASH and GUSTO and total cholesterol in mmol/L in PROSPER.*

*** Outcome is 14-day all-cause mortality in CRASH, 30-day all-cause mortality in GUSTO and a composite endpoint of coronary death, non-fatal myocardial infarction and fatal or non-fatal stroke at 3.2 years on average in PROSPER.*

Table 2. RCT and RD analyses in the CRASH trial (n = 9554).

Analysis	N total	R2 (%)	OR (95% CI) for 14-day mortality	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment*	4777	7	1.22 (1.06; 1.40)	0.071
RD – assignment: age ≤ 33 Tx-, age > 33 Tx+				
RCS adjustment*	4844	10	1.42 (0.94; 2.16)	0.212
Polynomial adjustment*	4844	10	1.09 (0.81; 1.46)	0.151
Local logistic regression	4844	NA	1.13 (0.90; 1.40)	0.112
RD – assignment: age ≤ 40 Tx-, age > 40 Tx+				
RCS adjustment*	4806	10	1.04 (0.68; 1.60)	0.218
Polynomial adjustment*	4806	10	0.94 (0.72; 1.23)	0.138
Local logistic regression	4806	NA	1.02 (0.80; 1.32)	0.129

**Linear, RCS or polynomial adjustment means that baseline age was used as a linear, RCS or polynomial term in the regression analysis to control for age.*

Table 3. RCT and RD analyses in the GUSTO trial (n = 30,510).

Analysis	N total	R2 (%)	OR (95% CI) for 30-day mortality	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment*	15255	12	0.83 (0.72; 0.95)	0.071
RD – assignment: age ≤ 62 Tx-, age > 62 Tx+				
RCS adjustment*	15423	11	0.57 (0.35; 0.92)	0.246
Polynomial adjustment*	15423	11	0.82 (0.63; 1.07)	0.133
Local logistic regression	15423	NA	0.67 (0.51; 0.86)	0.132
RD – assignment: age ≤ 70 Tx-, age > 70 Tx+				
RCS adjustment*	17846	10	0.94 (0.72; 1.22)	0.133
Polynomial adjustment*	17846	10	0.95 (0.75; 1.21)	0.121
Local logistic regression	17846	NA	0.90 (0.74; 1.10)	0.102

**Linear, RCS or polynomial adjustment means that baseline age was used as a linear, RCS or polynomial term in the regression analysis to control for age.*

Table 4. RCT and RD analyses in the PROSPER trial (n = 5804).

Analysis	N total	R2 (%)	OR (95% CI) for composite endpoint	Standard error (SE) of treatment effect estimate
RCT				
Linear adjustment	2902	0.4	0.85 (0.69; 1.04)	0.104
RD – assignment: cholesterol ≤ 5.6 Tx-, cholesterol > 5.6 Tx+				
RCS adjustment	2919	0.7	0.80 (0.46; 1.38)	0.279
Polynomial adjustment	2919	0.7	0.81 (0.56; 1.16)	0.185
Local logistic regression	2919	NA	0.79 (0.56; 1.13)	0.181
RD – assignment: cholesterol ≤ 6.2 Tx-, cholesterol > 6.2 Tx+				
RCS adjustment	2969	0.7	1.30 (0.71; 2.40)	0.311
Polynomial adjustment	2969	0.6	1.03 (0.69; 1.53)	0.205
Local logistic regression	2969	NA	1.07 (0.75; 1.56)	0.187

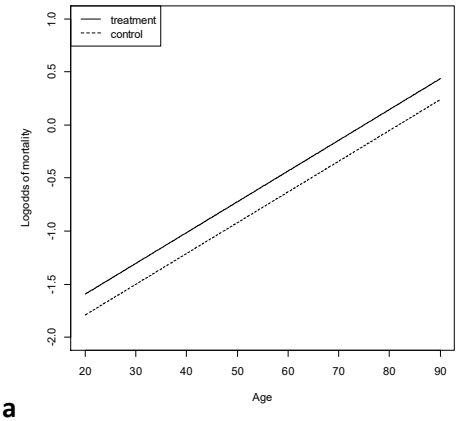
**Linear, RCS or polynomial adjustment means that baseline cholesterol level was used as a linear, RCS or polynomial term in the regression analysis to control for cholesterol level.*

Table 5. Relative efficiency in terms of required sample size for different designs in CRASH, GUSTO and PROSPER*.

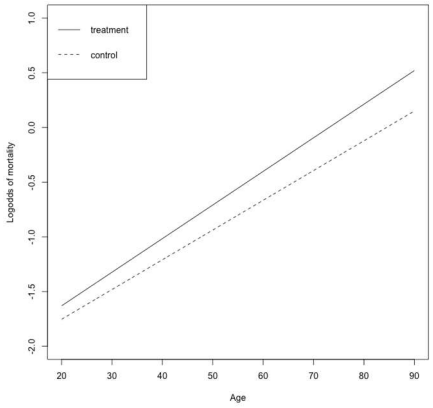
	CRASH	GUSTO	PROSPER
RCT adjusted vs RD RCS adjustment	8.9	12.1	7.2
RCT adjusted vs RD polynomial adjustment	4.5	3.5	3.1
RCT adjusted vs RD local logistic regression	2.5	3.5	3.0

*Formula: $(SE_{design\ 2} / SE_{design\ 1})^2$

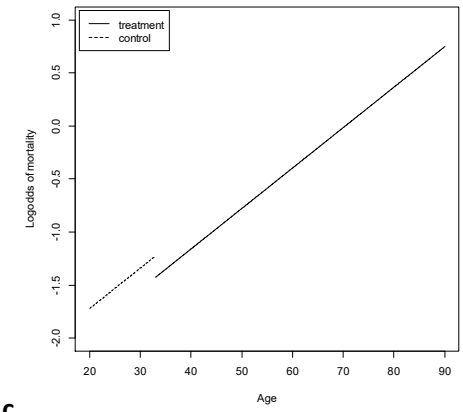
Figure 1. CRASH



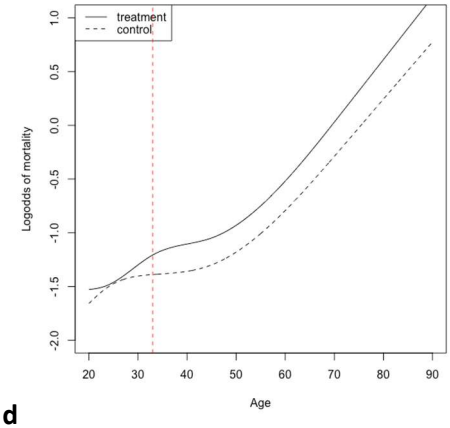
a



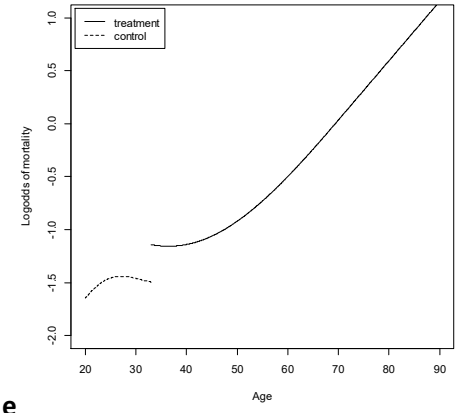
b



c



d



e

a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

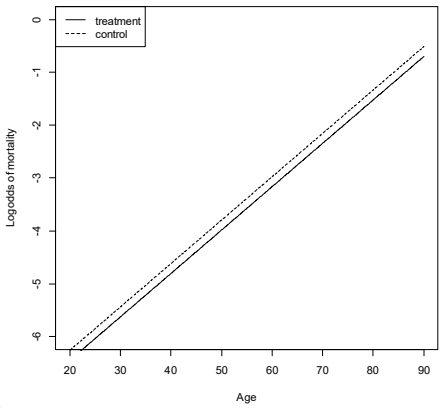
b Linear interaction function of the treatment effect over the baseline variable in RCT data.

c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

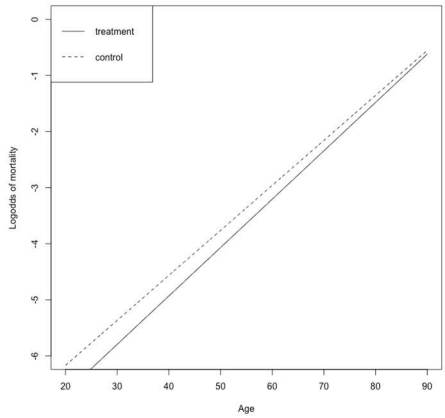
d RCS interaction function of the treatment effect over the baseline variable in RCT data.

e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

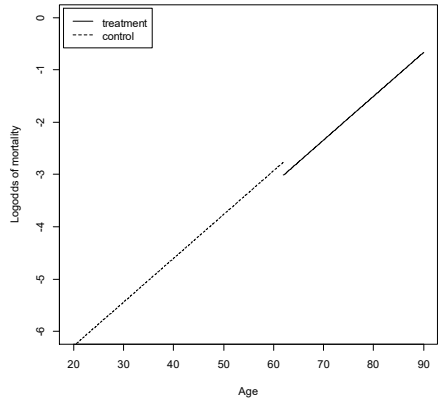
Figure 2. GUSTO



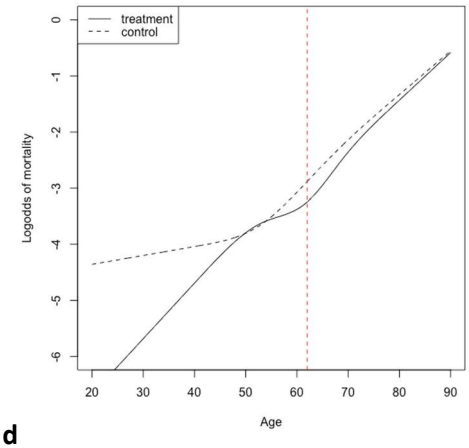
a



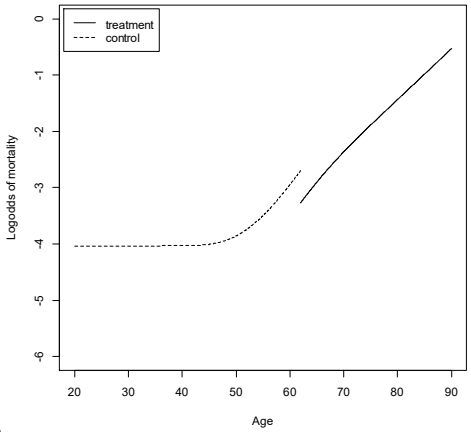
b



c



d



e

a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

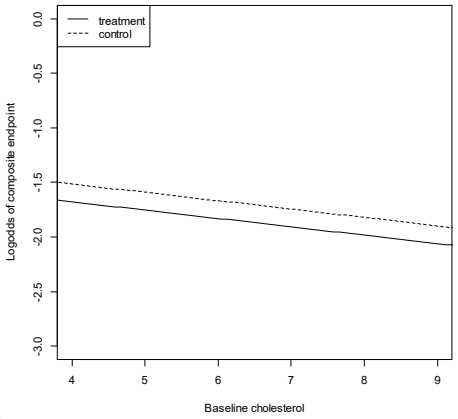
b Linear interaction function of the treatment effect over the baseline variable in RCT data.

c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

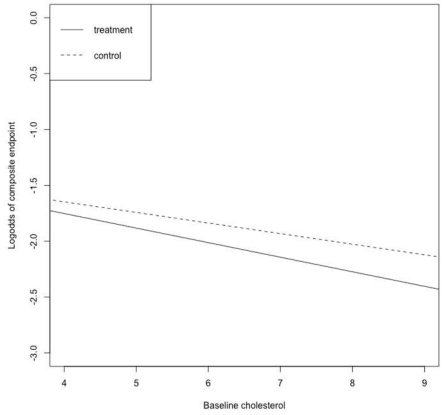
d RCS interaction function of the treatment effect over the baseline variable in RCT data.

e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

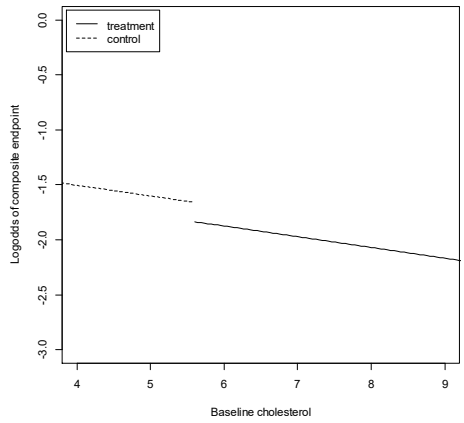
Figure 3. PROSPER



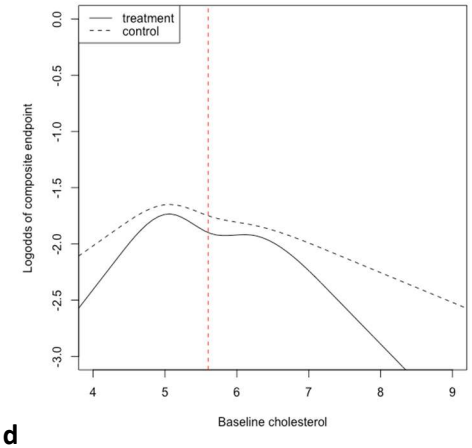
a



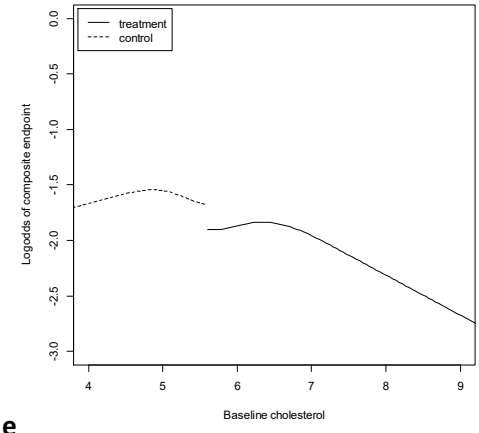
b



c



d



e

a Linear function of the baseline variable over the outcome variable in RCT data. The space between both lines indicates the main treatment effect in the RCT.

b Linear interaction function of the treatment effect over the baseline variable in RCT data.

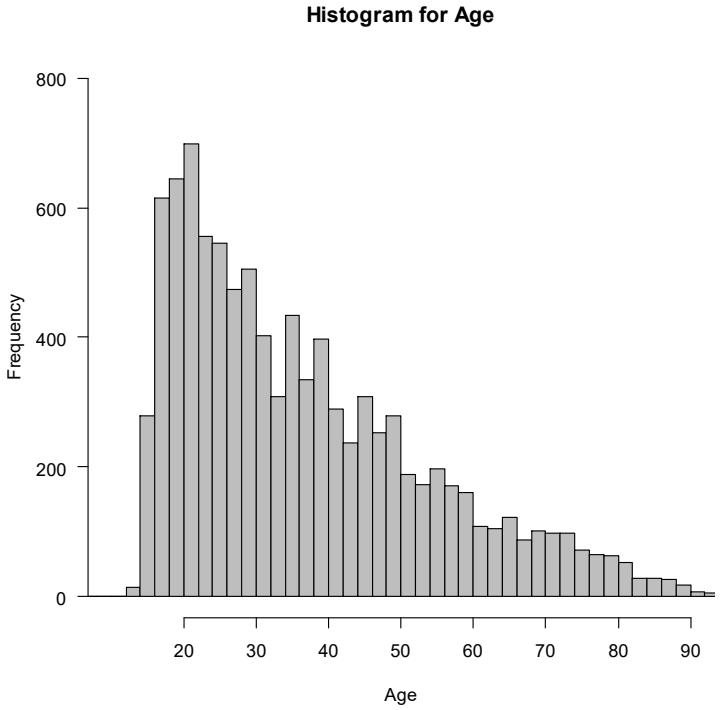
c Linear function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

d RCS interaction function of the treatment effect over the baseline variable in RCT data.

e RCS function of the baseline variable over the outcome variable in RD design. The space between both lines at the cut-off value indicates the treatment effect in the RD design.

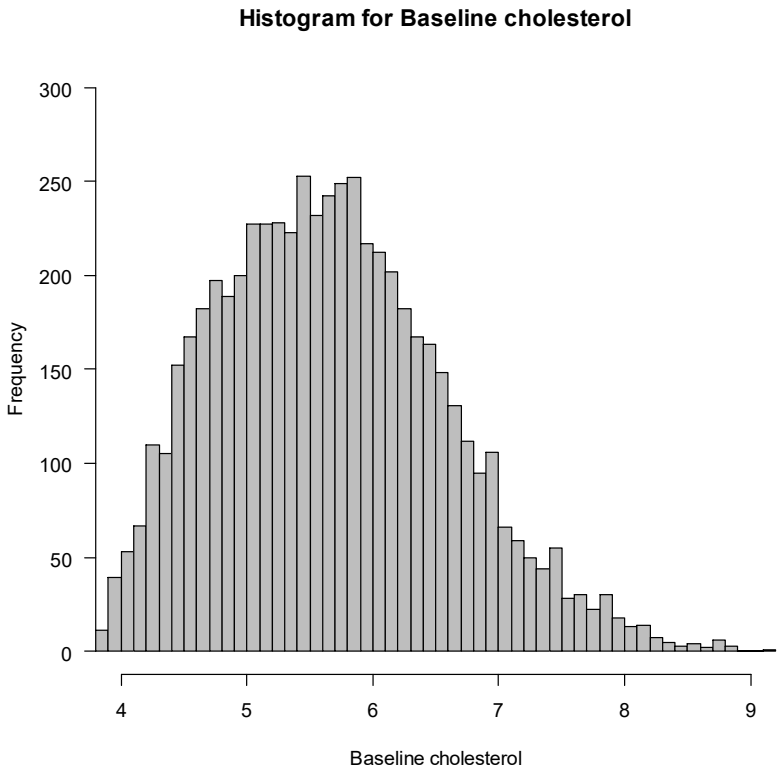
Supplement 1.

a. Distribution of baseline age in years in CRASH.



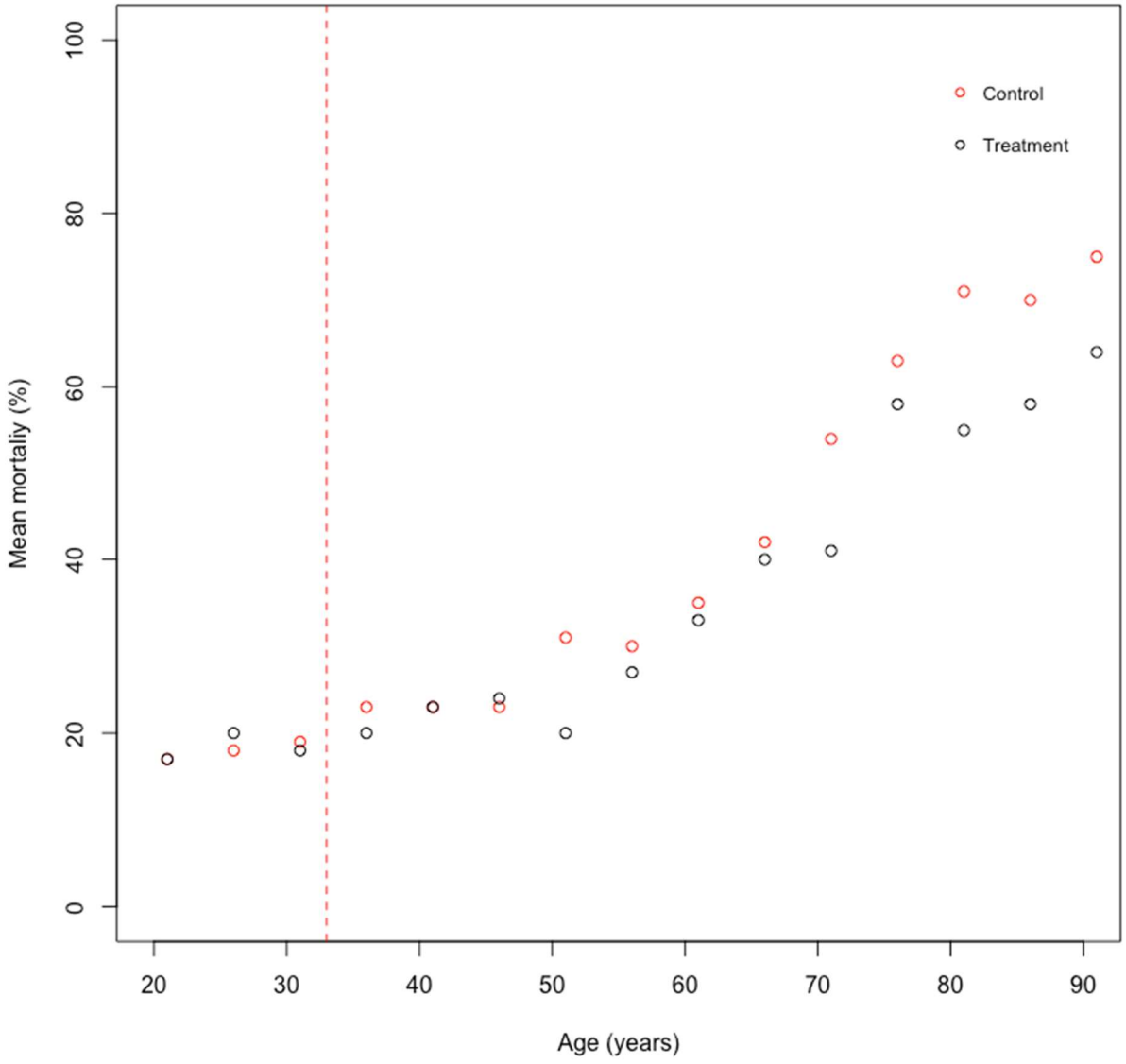
b. Distribution of baseline age in years in GUSTO.

c. Distribution baseline cholesterol level in mmol/L in PROSPER.

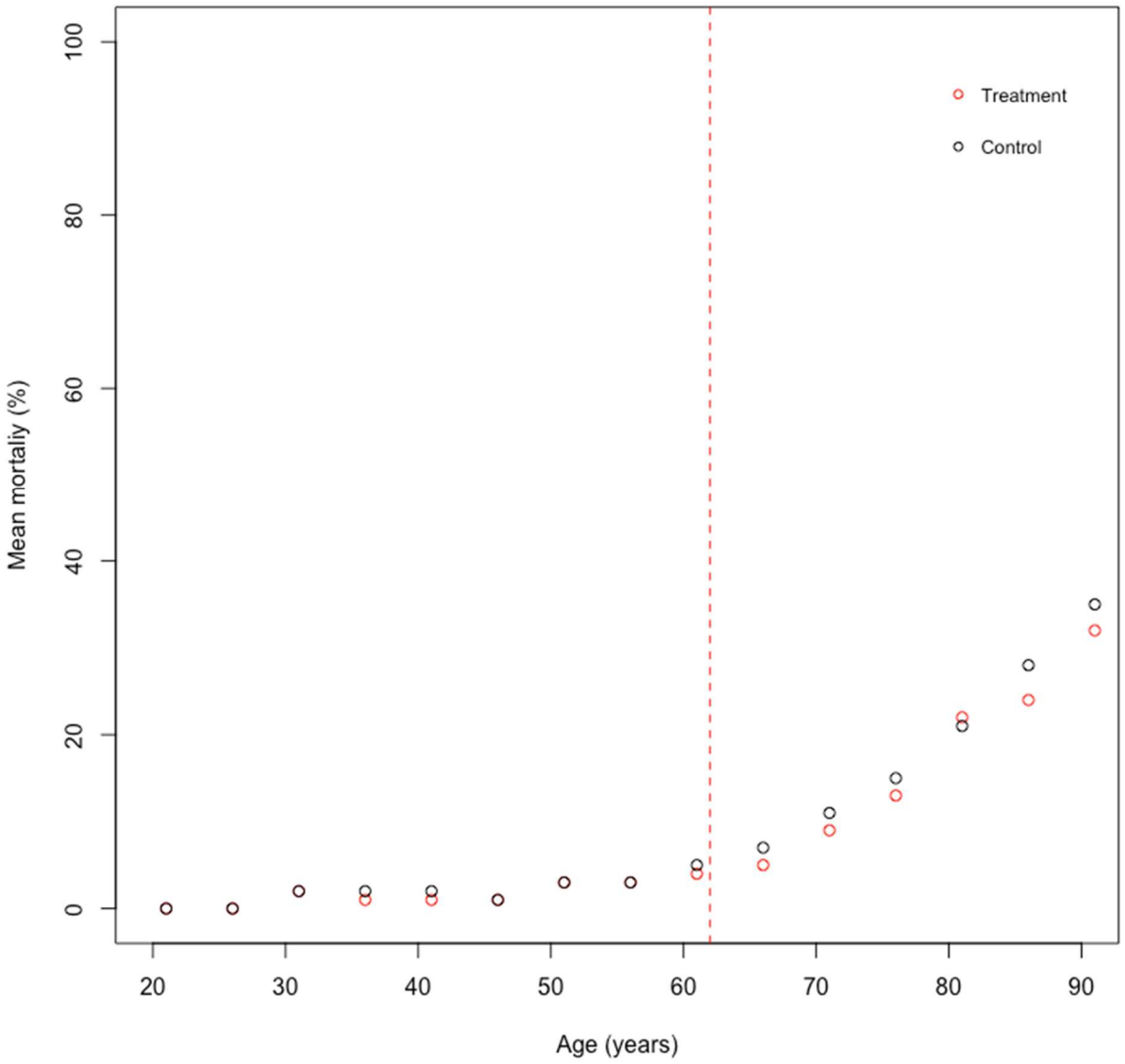


Supplement 2.

a. Binned scatterplot for mortality average, over the baseline age range in CRASH.



b. Binned scatterplot for mortality average, over the baseline age range in the GUSTO.



c. Binned scatterplot for composite endpoint average, over the baseline cholesterol level in mmol/L range in PROSPER.