The Regression discontinuity design: continuous and binary outcomes

Sara Geneletti

London School of Economics
Department of Statistics

Joint work with Gianluca Baio, Aidan O'Keeffe & Federico Ricciardi (UCL), Sylvia Richardson (MRC-BSU), Linda Sharples (LSHTM) & other collaborators — funded by UK MRC-MRP grant MR/K014838/1

Bayes Pharma 2019

Outline

1. Basic set-up and key concept
2. Sharp and Fuzzy designs
3. Continuous outcomes: LDL cholesterol levels
   - Data
   - Estimators
   - Prior constraints
   - Results
4. Binary outcomes: attaining recommended levels
   - Estimators
   - Prior constraints
5. The future of RDDs

The Regression Discontinuity Design – what is it?

- The Regression Discontinuity Design (RDD) was first introduced in the econometrics literature during the 1960s[5].
- The original idea was to exploit policy thresholds to estimate the causal effect of an educational intervention.
- The RDD has proven to be very useful when treatment is assigned based on a pre-specified rule linked to a continuous variable . For example:
  - Antiretroviral HIV drugs might be prescribed when a patients CD4 count is less than 200 cells/mm^3 [1];
  - Statins might be prescribed when a patient's 10-year risk of a cardiovascular event (10-year CVD risk score) exceeds a certain threshold (*e.g.* in the UK previously 20% and now 10%) [2]

RDD – key idea

- The **key idea** is that the threshold acts like a randomizing device
- For those who are familiar with this: it is an instrumental variable .
- This is possible if we consider the units *close* to the threshold as coming from the **same** population in which the assignment variable has its own natural (random) variability $\Rightarrow$ **(conditional) exchangeability**

Education example

- We want to quantify the effect of going to college on future income
- Comparing the income of individuals who attended college and those who did not will not provide us with the effect of college attendance alone
  - Confounders such as social class, ability, motivation etc will make this difficult
- That is a classic problem of observational studies

Education Example continued

- Often college scholarships are given on the basis of grades obtained in final school examinations, *eg* if the average exam grade is above 75%, the student gets a scholarship
- Suppose one student has an average of 74% and another an average of 76%:
  - Can we really consider them as coming from different populations especially if in other respects (*eg* family income, post code etc) they are similar?
  - Given that there is natural variability in exam performance even for the same individual?

Public health example

- Many medicines are prescribed according to a particular guideline
  - Antiretroviral HIV drugs prescribed when patient's CD4 count is less than 200 cells/mm^3[1];
  - Blood pressure medication is prescribed when patient's BP is 140/90mmHg or above;
  - Statins are prescribed when *eg* 10 year Framingham risk score is over 20%

- Consider a population of HIV patients and suppose patient `A` has a CD4 count of 195 and patient `B` has a count of 205 cells/mm^3
- **Theoretically** , patient `A` gets the drugs while patient `B` does not
- Can we really consider them as coming from different populations?
  - If the two are the same in every other relevant respect (*eg* individual circumstances etc)
  - Given that there is a natural variability in CD4 counts and in the instruments used to measure them?

Application: prescription of statins in primary care UK

Statins

- A class of drugs used to lower cholesterol and prescribed to prevent heart disease
  - Trials show an average reduction of LDL cholesterol of ≈ 2 mmol/l
  - UK NHS guidelines are to prescribe statins to individuals without previous CVD if their 10 year CVD score exceeds 20%(10%)

- **Data**: Clinical practice database containing routine GP prescriptions as well as information on the variables that determine them (THIN)
- 587 general practices in the UK, covering 5.2% of the (2013) UK population — over 10 million individuals living in the UK and fairly representative of the general population
- Individual characteristics (sex, date of birth, date of registration, proxies of socioeconomic status)
- Medical history (GP visits, prescriptions, exams)
- Relevant clinical outcomes (LDL level, CHD events, deaths)

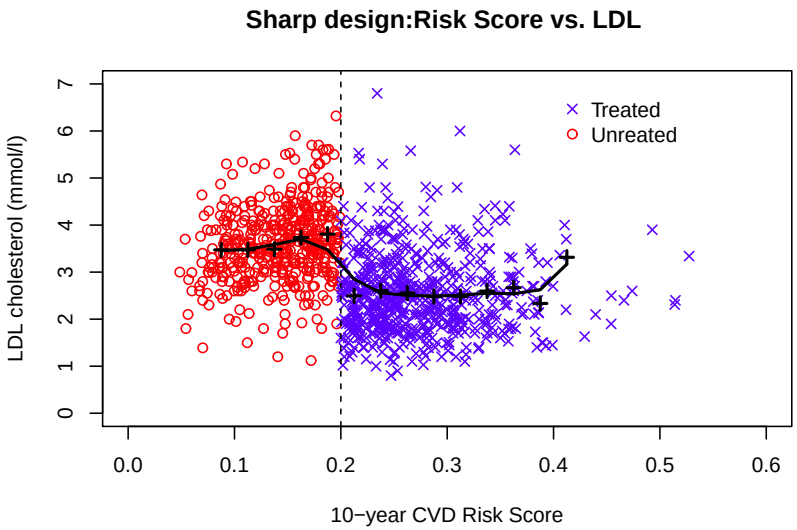
Notation (in our application)

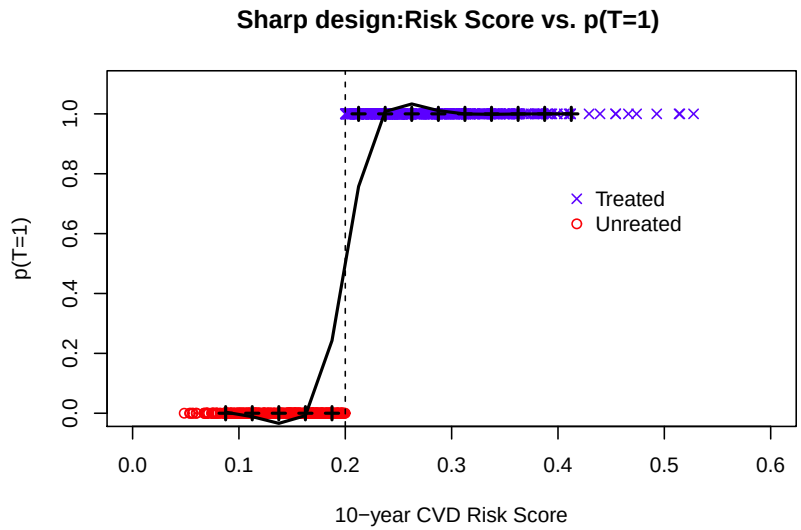
- $X =$ continuous forcing/assignment variable (risk score);
- $Z =$ threshold indicator (is risk score above/below 20%);
- $T =$ treatment **administered** (statin prescription);
- $\boldsymbol{C} \equiv (\boldsymbol{O}, \boldsymbol{U}) =$ observed and unobserved covariates (social class, co-morbidities);
- $Y =$ continuous outcome (LDL cholesterol level).

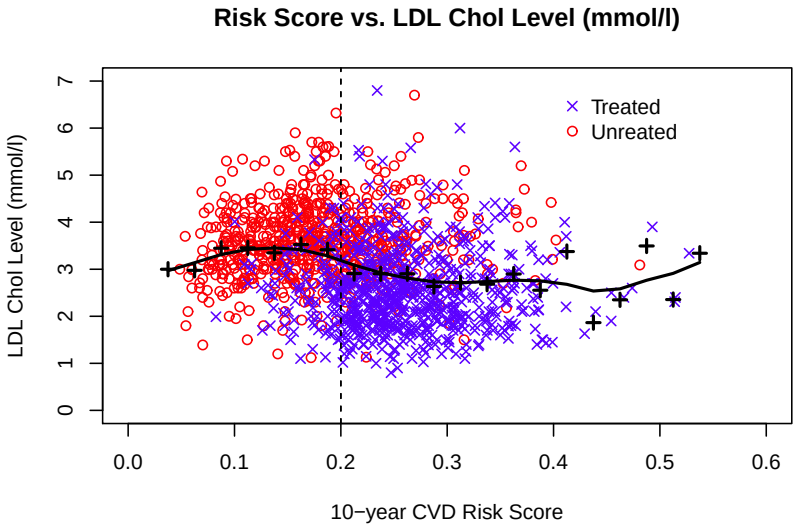
X and Z

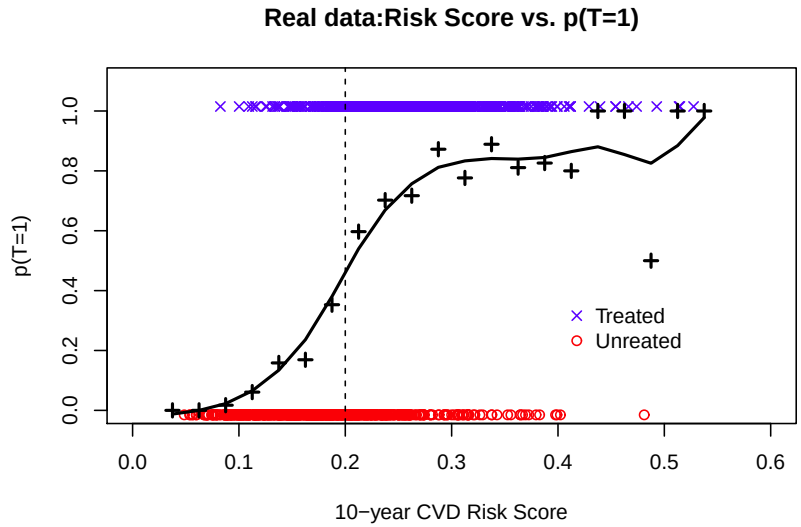
- X is the continuous variable and $X = x_0$ at the threshold
- $Z = 1$ if $X \geq x_0$ and $Z = 0$ if $X < x_0$

Sharp vs Fuzzy RDDs









RDD – Assumptions

1. *Unconfoundedness:* $Y \perp\!\!\!\perp Z \mid (T, \boldsymbol{C}, X)$ guarantees that the units just above and below the threshold are "similar".
2. *Independence of Guidelines:* $Z \perp\!\!\!\perp \boldsymbol{C} \mid X$ the threshold is set by an external body, *e.g.* a government agency.
3. *Monotonicity*: No decision-maker systematically defies the guidelines - i.e. GPs don't only prescibe to those below the threshold(!)
4. *Continuity*: $E(Y|Z, X = x, T, \boldsymbol{C})$ is continuous at in x (at x_0) for $T = 0, 1$
   If the outcome is discontinuous then the effect of threshold indicator will be confounded with the effect of whatever is responsible for the discontinuity

1-3 are instrumental variable assumptions

The Causal Effect

- Denote $x^c = x - x_0$ to be the forcing/assignment variable centred at x_0
- Consider the linear model

$$\mathsf{E}(Y) = \mu_{il} = \beta_{0l} + \beta_{1l} x_{il}^c \qquad l = a\text{bove}, b\text{elow}$$

- **NB**: "close" to the threshold, the covariates $\boldsymbol{C}$ are balanced, so no need to control for them (kind of...) — **but**: how close is close?
- The issue of bandwidth selection (how close) is still unresolved

Sharp RDD

- The formula for the **sharp** causal estimator is

$$\text{ATE} = \text{E}(Y|Z=1) - \text{E}(Y|Z=0) - \text{E}(T|Z=0) = \Delta_\beta = \beta_{0a} - \beta_{0b}$$

Fuzzy RDD

- The formula for the **fuzzy** causal effect estimator is

$$\text{LATE} = \frac{\text{E}(Y|Z=1) - \text{E}(Y|Z=0)}{\text{E}(T|Z=1) - \text{E}(T|Z=0)} = \frac{\Delta_\beta}{\Delta_\pi} = \frac{\beta_{0a} - \beta_{0b}}{\pi_a - \pi_b}$$

- π_l is an estimate of $\Pr(T=1|Z=z)$, *e.g.* the chance of being treated when above or below the threshold.

RDD – examples

- Increasingly popular in Public Health/Epidemiology
- HIV: The CD4 count is often used to deterimine drug assignment[1].
- HPV: The date of birth of a woman (pre/post vaccine availability)[4].
- Prostate cancer: PSA is a chemical produced by the body and used to determine treatment[3].
- Cholesterol: 10-year CVD risk score to determine statin treatment[2].

Problems with the LATE and Bayesian solutions

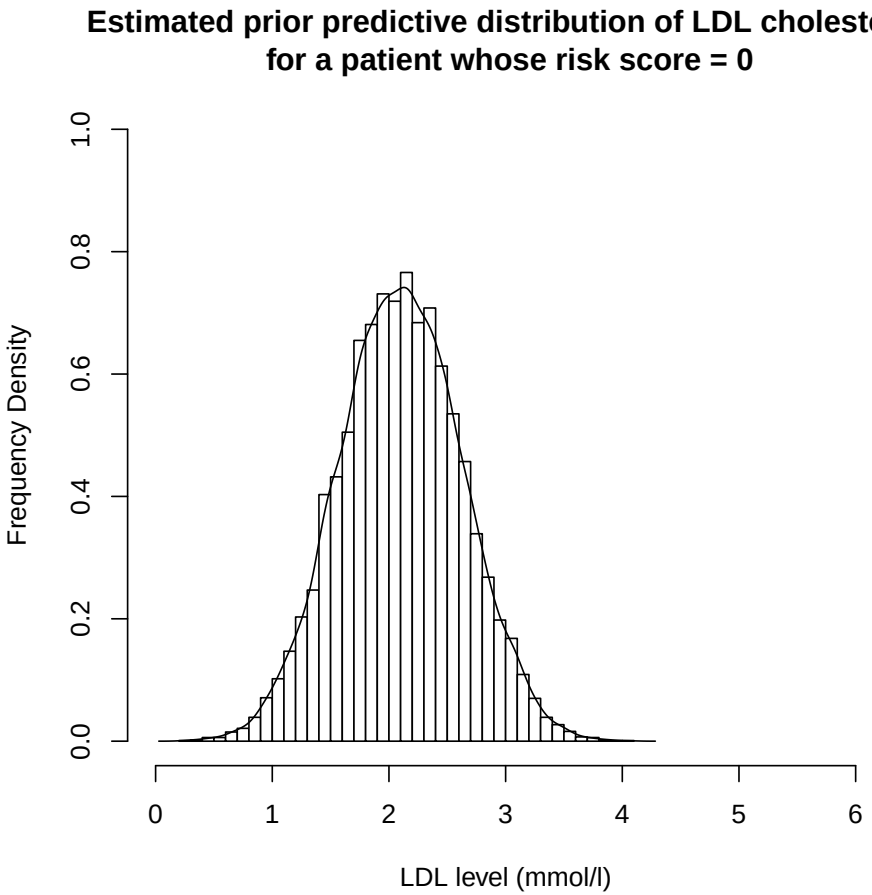
- The denominator of LATE can be very small (*i.e.* $\pi_a \approx \pi_b$)
- Informative priors on the relevant parameters can encode knowledge and assumptions about these two probabilities so that the resulting estimator does not explode to ∞

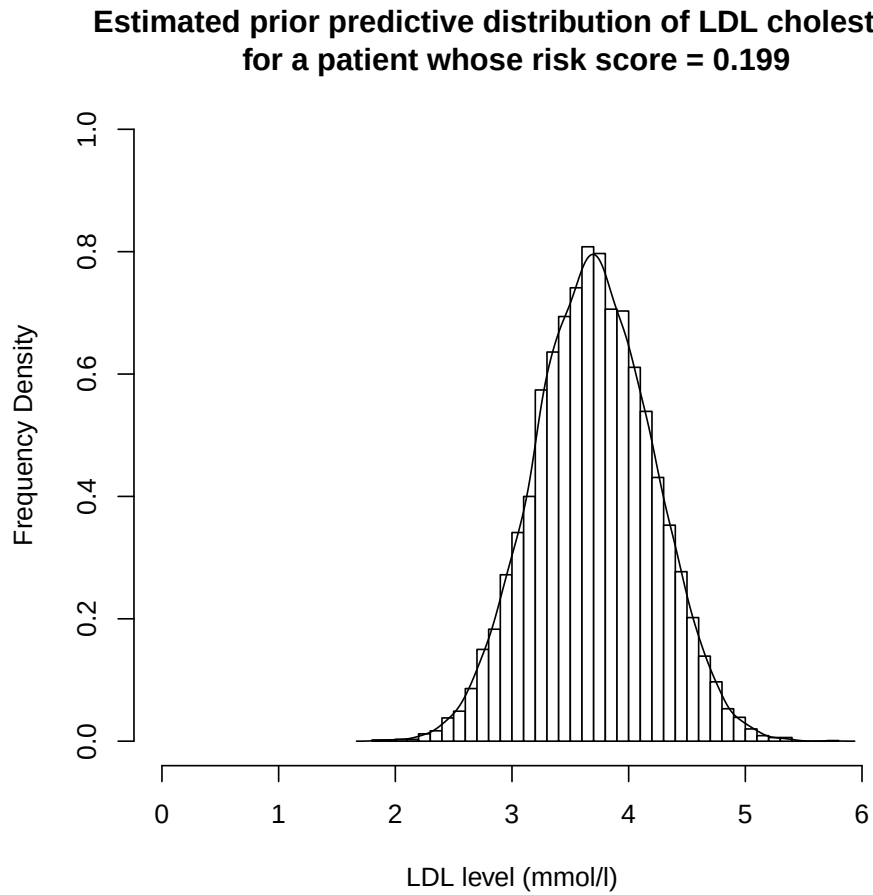
Other advantages of Bayes

- Estimation of variances and intervals does not rely on asymptotics — just a byproduct of MCMC procedures + can naturally include more appropriate models (vs 2SLS)
- Can encode more complex models to account for different levels of compliance in a straightforward manner

Bayesian modelling: $\mu_{il} = \beta_{0l} + \beta_{1l} x_{il}^c$

1. Informative prior on the slopes, based on clinical expert opinions





$\beta_{1l} \sim \text{Normal}(m_{1l}, s_{1l}^2)$, for suitable values of m_{1l} and s_{1l}^2

Bayesian modelling: $\mu_{il} = \beta_{0l} + \beta_{1l} x^c_{il}$

2. Informative priors on the intercepts:
   $\beta_{0b} \sim \text{Normal}(m_0, s_0^2)$ and $\beta_{0a} = \beta_{0b} + \phi$

- **Weakly informative prior**: $\phi \sim \text{Normal}(0, 2)$
  - "Skeptical" prior on the effect of treatment, which is assumed to be null
- **Strongly informative prior**: $\phi \sim \text{Normal}(-2, 1)$
  - "Enthusiastic" prior, strongly based on the available information coming from the RCTs (reduction of 2 mmol/l)[1]
  - Relatively small variance to represent strong belief in the trials

[1] Ward et al (2007) [6]

Bayesian modelling: logit(π_l)

2. Informative priors on the intercepts:
   $\beta_{0b} \sim \text{Normal}(m_0, s_0^2)$ and $\beta_{0a} = \beta_{0b} + \phi$

- **Weakly informative prior**: $\phi \sim \text{Normal}(0, 2)$
  - "Skeptical" prior on the effect of treatment, which is assumed to be null
- **Strongly informative prior**: $\phi \sim \text{Normal}(-2, 1)$
  - "Enthusiastic" prior, strongly based on the available information coming from the RCTs (reduction of 2 mmol/l)[1]
  - Relatively small variance to represent strong belief in the trials

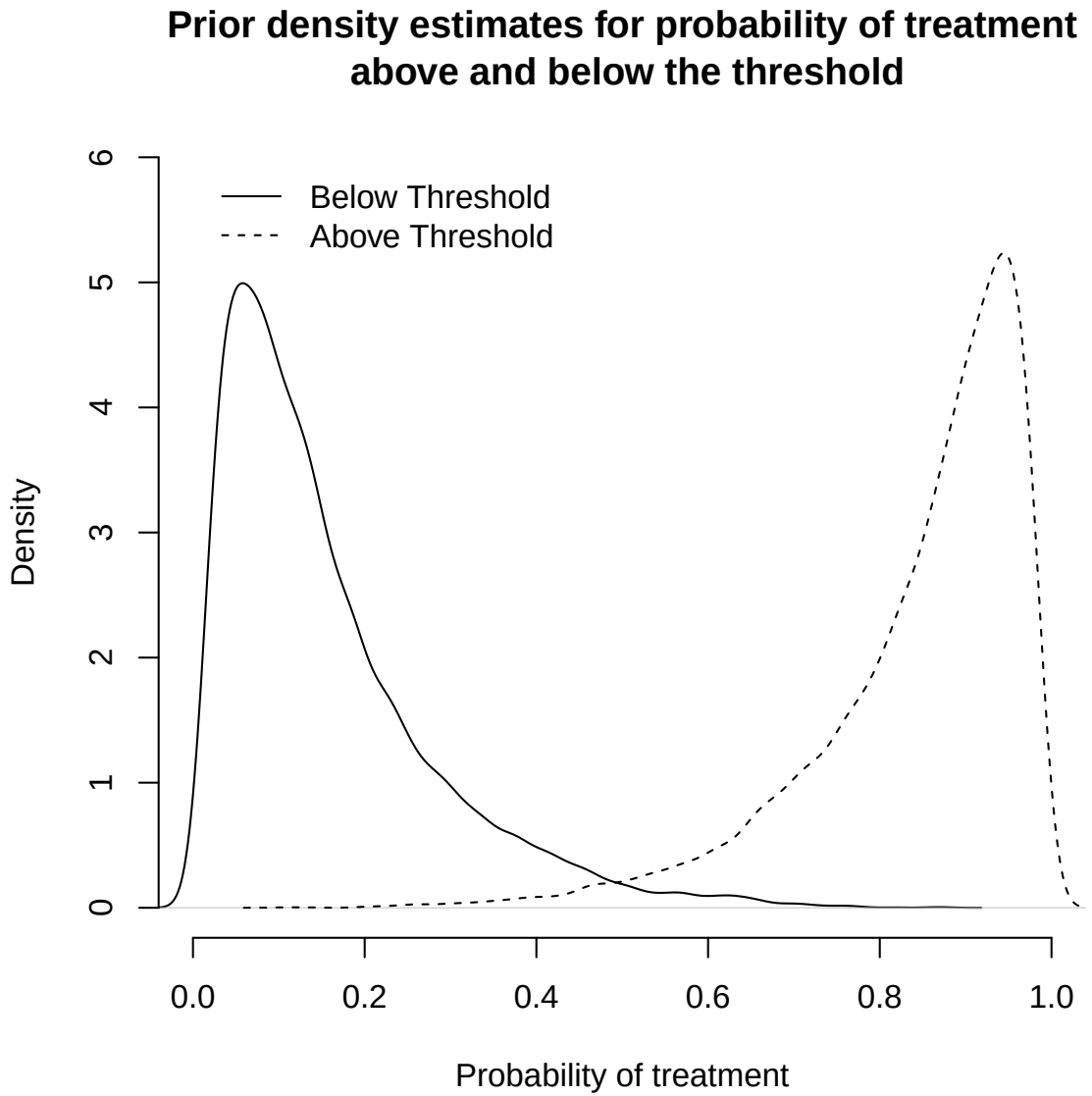
3. **Informative prior on the probability of treatment**:

$$\text{logit}(\pi_a) \sim \text{Normal}(2, 1), \qquad \text{logit}(\pi_b) \sim \text{Normal}(-2, 1)$$

- **NB**: implies that $\Delta_\pi = \pi_a - \pi_b$ is centered far from 0 but can vary
- Helps stabilise the denominator and thus the LATE

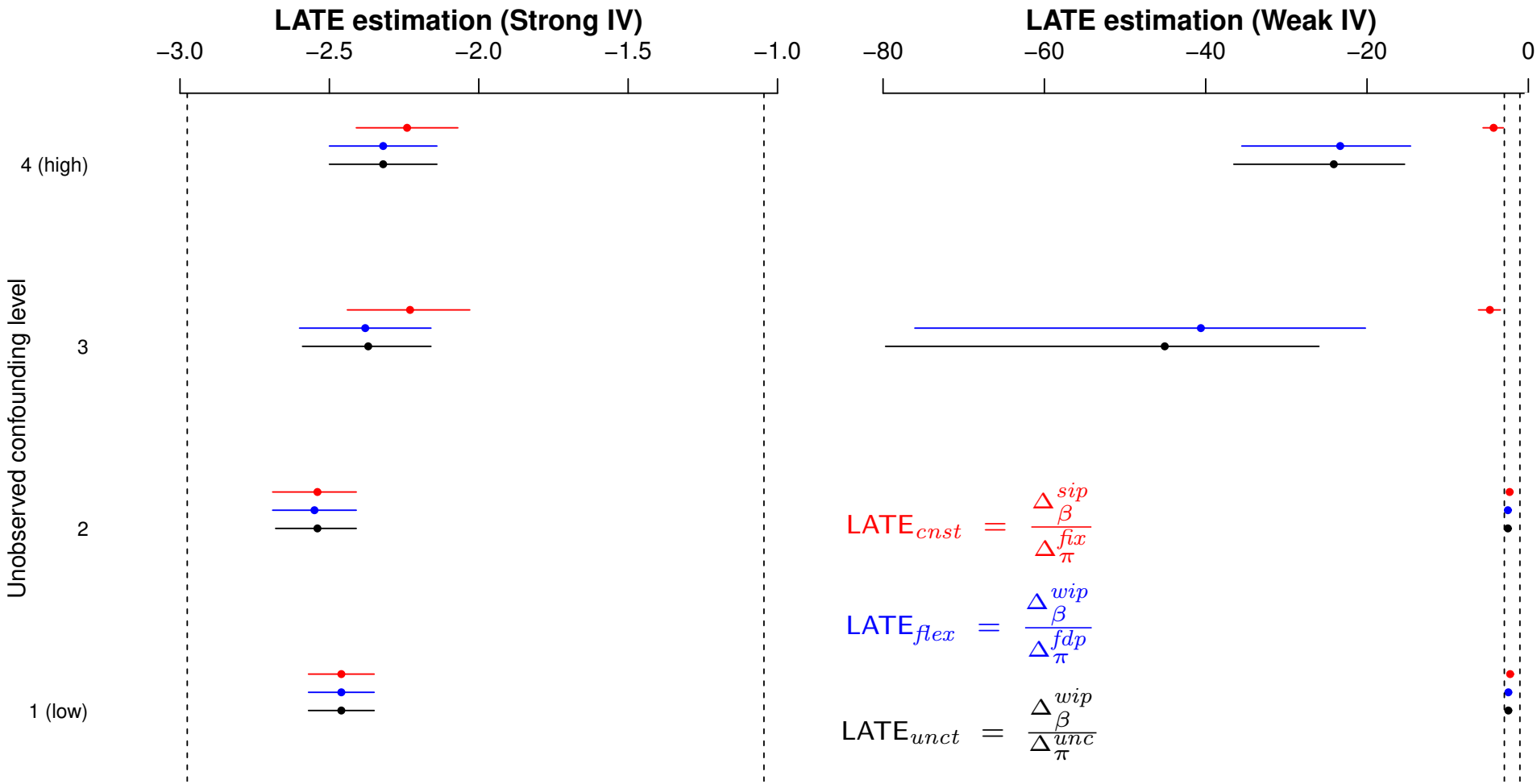
[1] Ward et al (2007) [6]

Bayesian modelling: logit(π_l)



Simulation study results

Bandwidth $= 0.25$ (fairly large!), Treatment effect size $\sim$ Normal$(-2, 0.5^2)$



Binary outcomes

- Most of the RDD literature focusses on continuous outcomes, but often in biostatistics, practitioners are interested in **binary** outcomes

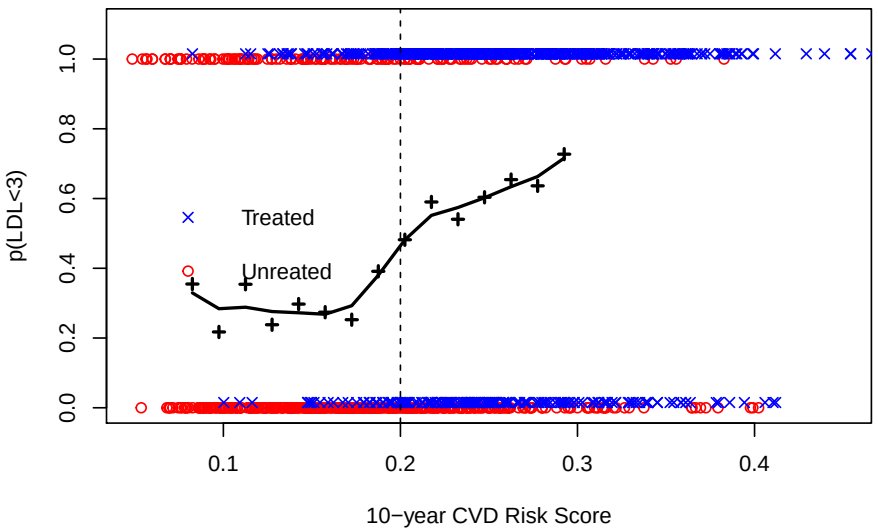
Example

- Did LDL cholesterol levels drop to recommended levels after statin prescription?
- Guidelines in the UK state that LDL cholesterol levels should be below 2 mmol/l for patients who are at high risk (e.g. with multiple co-morbidities)
- And below 3 mmol/l for low risk patients

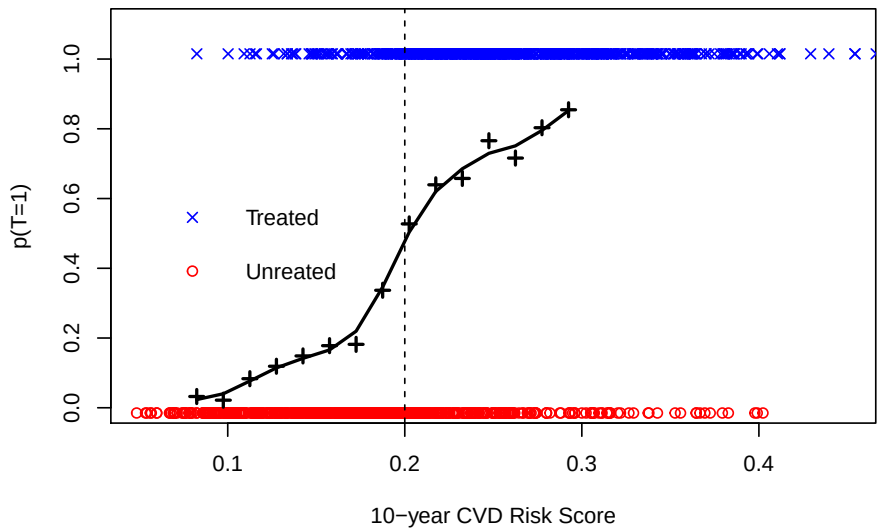
- Using the same data we dichotomised the LDL cholesterol outcome such that $Y = 1$ is LDL cholesterol levels are below 2mmol/l (or 3 mmol/l) and $Y = 0$ otherwise

LDL cholesterol levels

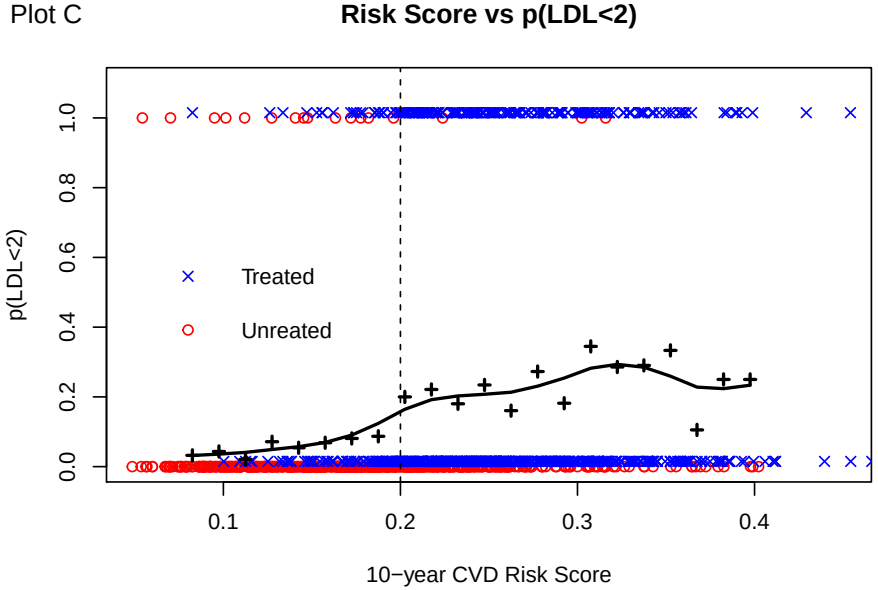
Plot A



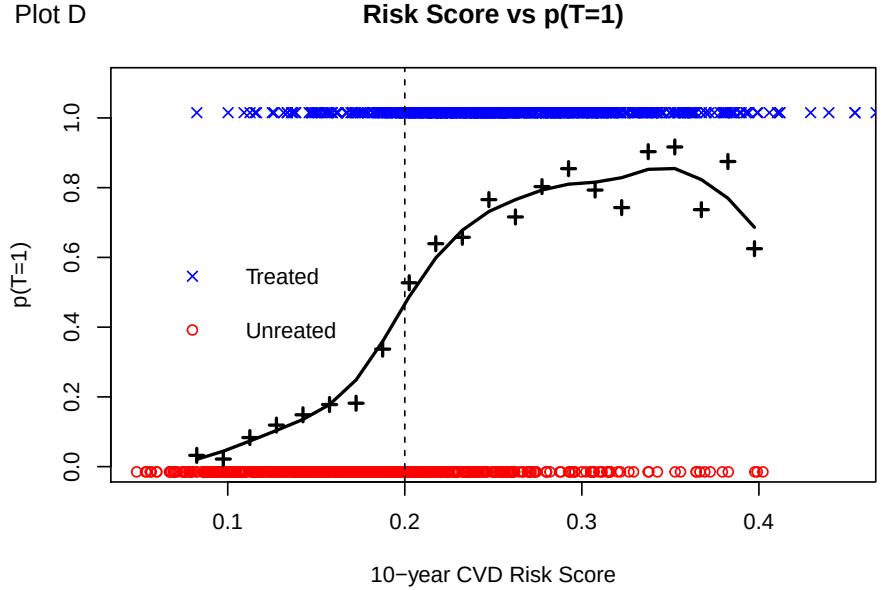
Plot B



Plot C



Plot D



Binary outcomes

- We can draw on the IV-based Multiplicative Structural Mean Models (MSMMs), which consider the causal **Risk Ratio for the Treated** (RRT)

$$\begin{aligned} \text{RRT} &= \frac{\mathsf{E}[\mathsf{E}_a(Y \mid Z) \mid T = 1]}{\mathsf{E}[\mathsf{E}_b(Y \mid Z) \mid T = 1]} \\ &= 1 - \frac{\mathsf{E}(Y \mid Z = 1) - \mathsf{E}(Y \mid Z = 0)}{\mathsf{E}(Y\bar{T} \mid Z = 1) - \mathsf{E}(Y\bar{T} \mid Z = 0)} \end{aligned}$$

  when a set of assumptions holds (log-linear in t + no T-Z multiplicative interaction)[2]
- Known issues of standard estimators (*e.g.* generalised method of moments):
  - May give absurd results (lower 95% interval estimate < 0)
  - The data for the product term ($Y\bar{T}$) are usually sparse $\Rightarrow$ implausibly wide interval estimates
- Can "fix" it by using suitable constraints — similar to those used to stabilise the denominator of the LATE in the continuous case

Constraining the models (1)

- The RRT is expressed as a function of a set of parameters (in the same spirit as the LATE)

$$\text{RRT} = f\left(\exp(\alpha_a) - \exp(\alpha_b)\right)$$

  where:

  - α_a and α_b are the intercepts in the log-linear models for $\text{E}(Y \mid Z = 1)$ and $\text{E}(Y \mid Z = 0)$
  - For convenience, model $y_{il} \sim \text{Poisson}(\mu_{il})$ — consistent with MSMM assumptions

- Typically, we would put priors on α_a and α_b, which would induce a prior on RRT

Constraining the models (1)

- The RRT is expressed as a function of a set of parameters (in the same spirit as the LATE)

$$\text{RRT} = f\left(\exp(\alpha_a) - \exp(\alpha_b)\right)$$

where:

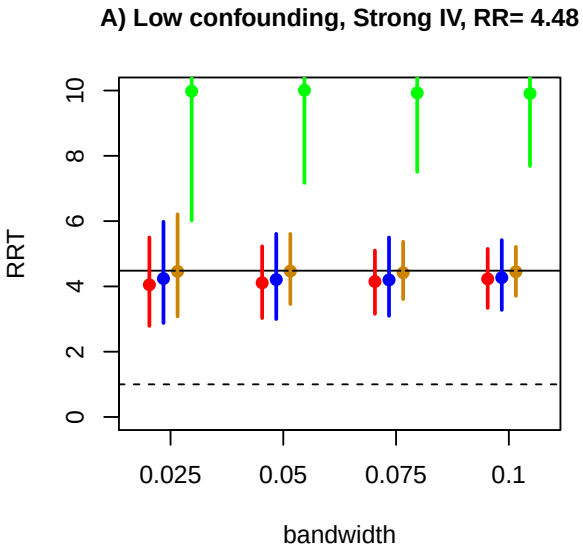
  - α_a and α_b are the intercepts in the log-linear models for $\text{E}(Y \mid Z = 1)$ and $\text{E}(Y \mid Z = 0)$
  - For convenience, model $y_{il} \sim \text{Poisson}(\mu_{il})$ — consistent with MSMM assumptions

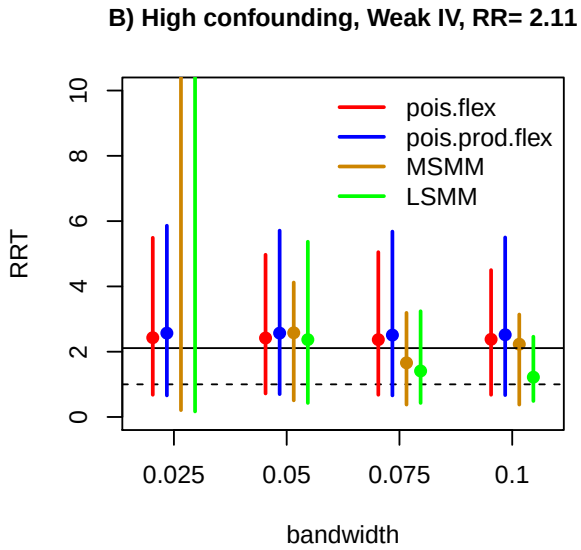
- Typically, we would put priors on α_a and α_b, which would induce a prior on RRT

- **But**: can also put a prior on RRT to ensure that it is > 0 and another on α_a. This would then induce a prior on α_b, *e.g.*

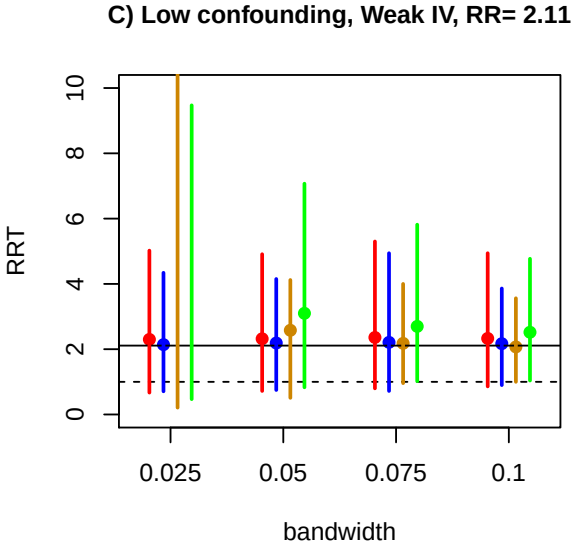
$$\text{RRT} \sim \text{Gamma}(3, 1) \qquad \alpha_a \sim p(\alpha_a) \qquad \text{and} \qquad \alpha_b = g\left(\text{RRT}, \alpha_a\right)$$

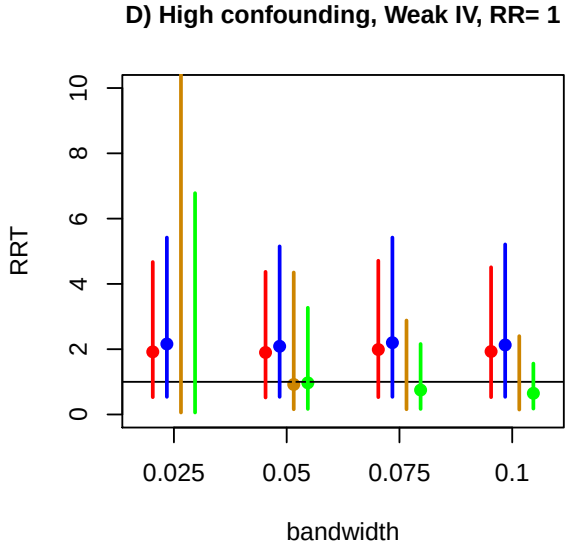
Results





"pois.flex" $\Rightarrow$ based on $\mathsf{E}(Y\bar{T} \mid Z)$
"pois.prod.flex" $\Rightarrow$ based on $\mathsf{E}(Y \mid \bar{T}, Z)\mathsf{E}(\bar{T} \mid Z)$
"MSMM" $\Rightarrow$ Generalised method of moments []
"LSMM" $\Rightarrow$ Logistic Structural Mean Model []





Sensitivity/Range analysis

- For both continuous and binary outcomes we have developed a number of estimators which are based on slightly different assumptions about e.g. how the denominator works or whether data are sparse
- All of them (and indeed the standard RDD/IV estimators) should be used in any real context
- In a best case scenario all of the estimates have substantial overlap (as was the case in our application)
- When they do not then certainly prefer the ones we develop(!) and use the simulations to understand why there are discrepancies
- I am always reluctant to give a single point estimate with a credible interval – plausible ranges are better

Systematic RDDs

There is a lot of potential for making RDDs less opportunistic and more systematic

Picture this...

- Imagine there is a new drug on the market - it's passed trials etc
- The NHS wants to know: where do we set the threshold to optimise benefits (minimise cost?)
- Some rough idea comes from trials but we know they have low external validity
- Run 3/4 RDDs with different thresholds in different sub-populations
- Is this any less ethical than changing the guidelines as evidence from primary care emerges?

Conclusions

- "*Real World Evidence*" (*i.e.* Electronic Health Record data) is increasingly popular in research
  - Causal estimates are still tricky because of issues with self-selection, confounding, etc
- Useful to (critically!) explore specific designs to balance characteristics
  - RDD
  - Interrupted time series
  - ...

Conclusions

- "*Real World Evidence*" (*i.e.* Electronic Health Record data) is increasingly popular in research
  - Causal estimates are still tricky because of issues with self-selection, confounding, etc
- Useful to (critically!) explore specific designs to balance characteristics
  - RDD
  - Interrupted time series
  - ...
- Bayesian modelling particularly helpful
  - Because data are available in registries, administrative databases, there are likely to be RCTs (may be on small samples/time frames) to base priors on
  - Design alone may not be sufficient to obtain balance — may need to impose constraints $\Rightarrow$ explicit and typically relatively easy in a full Bayesian framework

Thank you!

References

Jacob Bor, Shahira Ahmed, Matthew P. Fox, Sydney Rosen, Gesine Meyer-Rath, Ingrid T. Katz, Frank Tanser, Deenan Pillay, and Till Baernighausen.

Effect of eliminating CD4-count thresholds on HIV treatment initiation in South Africa: An empirical modeling study.
PLOS ONE, 12(6), JUN 15 2017.

S. Geneletti, F. Ricciardi, A. G. OKeeffe, and G. Baio.

Bayesian modelling for binary outcomes in the regression discontinuity design.
Submitted to JASA, 2016.

Jonathan Shoag, Joshua Halpern, Brian Eisner, Richard Lee, Sameer Mittal, Christopher E. Barbieri, and Daniel Shoag.

Efficacy of Prostate-Specific Antigen Screening: Use of Regression Discontinuity in the PLCO Cancer Screening Trial.
JAMA ONCOLOGY, 1(7):983–986, OCT 2015.

Leah M. Smith, Linda E. Levesque, Jay S. Kaufman, and Erin C. Strumpf.

Strategies for evaluating the assumptions of the regression discontinuity design: a case study using a human papillomavirus vaccination programme.
INTERNATIONAL JOURNAL OF EPIDEMIOLOGY, 46(3):939–949, JUN 2017.

D. L. Thistlethwaite and D. T. Campbell.

Regression-discontinuity analysis: an alternative to the ex-post facto experiment.
Journal of Educational Psychology, 51:309–317, 1960.

S. Ward, L. Jones, A. Pandor, M. Holmes, R. Ara, A. Ryan, W. Yeo, and N. Payne.

A systematic review and economic evaluation of statins for the prevention of coronary events.
Health Technology Assessment, 11(14), 2007.