Statistical issues in the application of the regression discontinuity design for causal inference from clinical administrative databases

Gianluca Baio

University College London
Department of Statistical Science

g.baio@ucl.ac.uk

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

Symposium "Statistical Methods for Recurrent Data workshop"

Universitè de Lille
Monday 7 November 2016

Outline

1. **Regression discontinuity design**
   - **Basic structure & assumptions**
   - **Sharp vs fuzzy**
2. It's good to be Bayesians (1)
   - Link to evidence synthesis
   - Stabilise estimates via prior information
3. It's good to be Bayesians (2)
   - Binary outcomes
   - Wacky frequentist- vs regolarised Bayesian-estimates
4. How close is close enough?
   - Optimal bandwidths
   - No bandwiths?
5. Conclusions

Outline

1. Regression discontinuity design
   - Basic structure & assumptions
   - Sharp vs fuzzy
2. **It's good to be Bayesians (1)**
   - **Link to evidence synthesis**
   - **Stabilise estimates via prior information**
3. It's good to be Bayesians (2)
   - Binary outcomes
   - Wacky frequentist- vs regolarised Bayesian-estimates
4. How close is close enough?
   - Optimal bandwidths
   - No bandwiths?
5. Conclusions

Outline

1. Regression discontinuity design
   - Basic structure & assumptions
   - Sharp vs fuzzy
2. It's good to be Bayesians (1)
   - Link to evidence synthesis
   - Stabilise estimates via prior information
3. **It's good to be Bayesians (2)**
   - **Binary outcomes**
   - **Wacky frequentist- vs regolarised Bayesian-estimates**
4. How close is close enough?
   - Optimal bandwidths
   - No bandwiths?
5. Conclusions

Outline

1. Regression discontinuity design
   - Basic structure & assumptions
   - Sharp vs fuzzy
2. It's good to be Bayesians (1)
   - Link to evidence synthesis
   - Stabilise estimates via prior information
3. It's good to be Bayesians (2)
   - Binary outcomes
   - Wacky frequentist- vs regolarised Bayesian-estimates
4. **How close is close enough?**
   - **Optimal bandwidths**
   - **No bandwiths?**
5. Conclusions

Outline

1. Regression discontinuity design
   - Basic structure & assumptions
   - Sharp vs fuzzy
2. It's good to be Bayesians (1)
   - Link to evidence synthesis
   - Stabilise estimates via prior information
3. It's good to be Bayesians (2)
   - Binary outcomes
   - Wacky frequentist- vs regolarised Bayesian-estimates
4. How close is close enough?
   - Optimal bandwidths
   - No bandwiths?
5. **Conclusions**

The Regression Discontinuity Design

- The Regression Discontinuity Design (RDD) was first introduced in the econometrics literature during the 1960s. The original idea was to exploit policy thresholds to estimate the causal effect of an educational intervention.

Thistle & Campbell (1960) [9]

The Regression Discontinuity Design

- The Regression Discontinuity Design (RDD) was first introduced in the econometrics literature during the 1960s. 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;
  - 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%)

The **key idea** is that the threshold acts like a randomizing device.
This is possible if we consider the units close to the threshold as they come from the same population in which the assignment variable has its own natural variability $\Rightarrow$ **(conditional) exchangeability**

Thistle & Campbell (1960) [9]; NICE (2008) [6]; NICE (2014) [7]

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

- 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

- 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 etc) they are the same?
  - Given that there is natural variability in exam performance even for the same individual?

What is the RD design?

Public health example

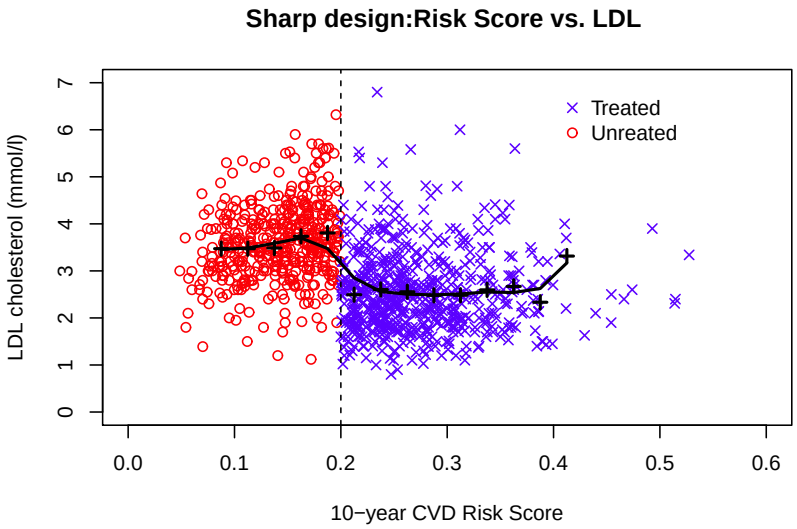
- Many medicines are prescribed according to a particular guideline
  - Antiretroviral HIV drugs prescribed when patient's CD4 count is less than 200 cells/mm^3
  - Blood pressure medication is prescribed when patient's BP is 140/90mmHg or above
  - Statins are prescribed when *eg* 10 year Framingham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

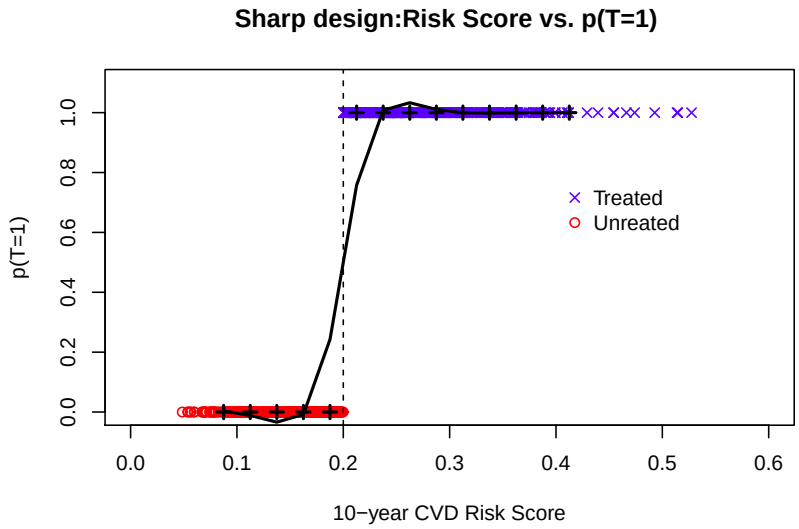
What is the RD design?

Public health example

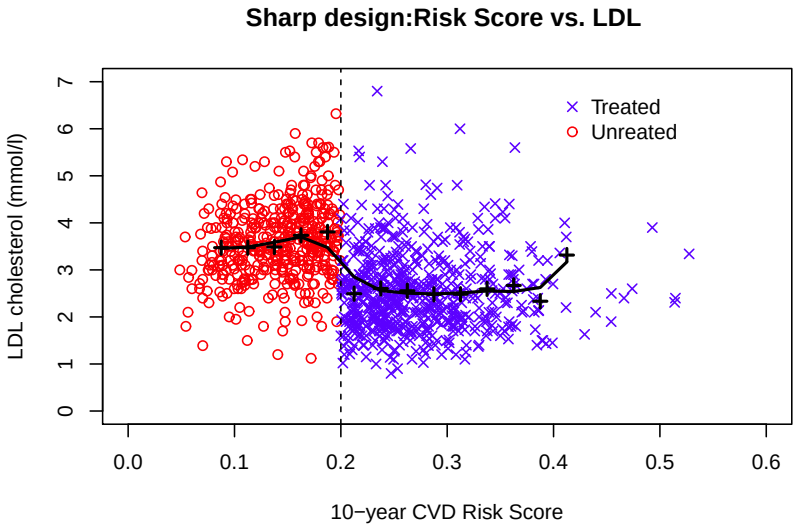
- Many medicines are prescribed according to a particular guideline
  - Antiretroviral HIV drugs prescribed when patient's CD4 count is less than 200 cells/mm^3
  - Blood pressure medication is prescribed when patient's BP is 140/90mmHg or above
  - Statins are prescribed when *eg* 10 year Framingham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

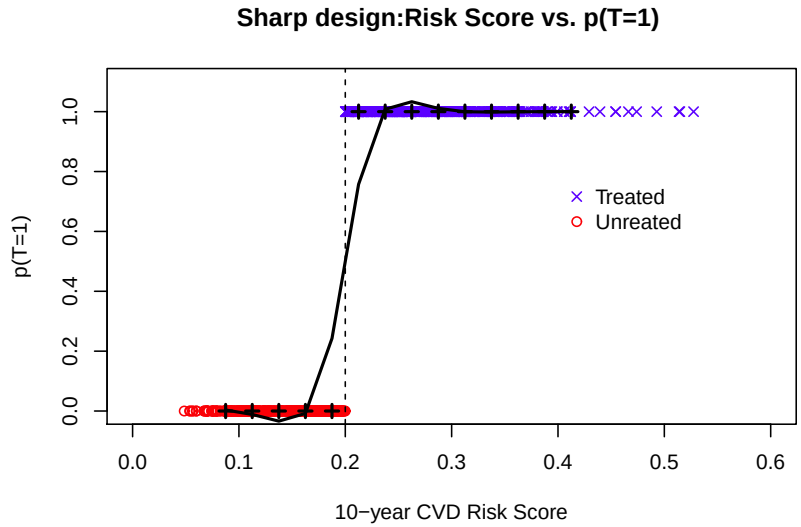
Sharp vs Fuzzy RDDs

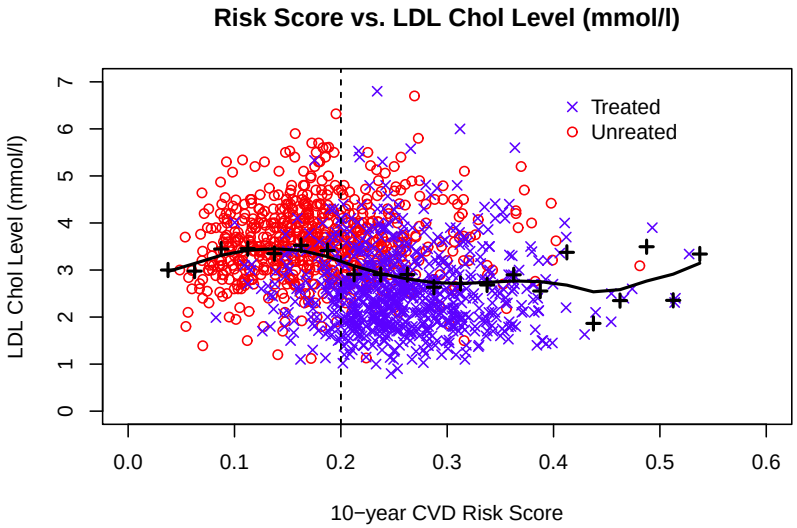


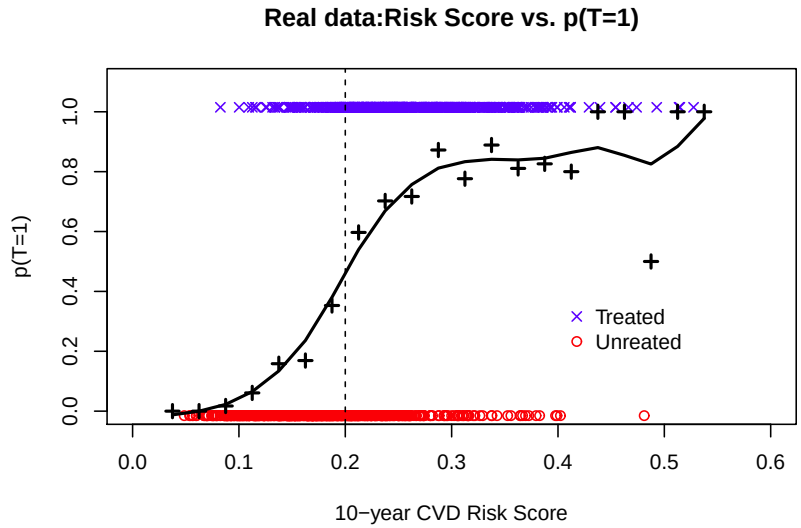


Sharp vs Fuzzy RDDs









Notation & assumptions

- X = continuous forcing/assignment variable;
- Z = threshold indicator;
- T = treatment **administered**;
- $\boldsymbol{C} \equiv (\boldsymbol{O}, \boldsymbol{U})$ = observed and unobserved covariates;
- Y = outcome.

(Main) 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 governmental agency.

3. *Monotonicity:*
   No decision-maker systematically defies the guidelines.

Geneletti et al (2015) [3]

The Causal Effect

- Denote $x^c = x - x_0$ to be the centered forcing variable
- Consider the linear model

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

- **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? (more on this later)

The Causal Effect: Sharp RDD

- Denote $x^c = x - x_0$ to be the centered forcing variable
- 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? (more on this later)

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

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

The Causal Effect: Fuzzy RDD

- Denote $x^c = x - x_0$ to be the centered forcing variable
- Consider the linear model

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

- **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? (more on this later)

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

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

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

- **NB**: The RDD can be linked to instrumental variables (IVs)

Case study: prescription of statins

- **Statins** are 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%)

Case study: prescription of statins

- **Statins** are 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**: Simulation + real clinical practice database containing routine GP prescriptions as well as information on the variables that determine them (THIN: www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database)
  - 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)

The importance of being a Bayesian (1)

- Stabilising the estimators
  - 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 ∞

Geneletti et al (2015) [3]

The importance of being a Bayesian (1)

- Stabilising the estimators
  - 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 ∞
- Computational advantages
  - Estimation of variances and intervals does not rely on asymptotics — just a byproduct of MCMC procedures + can naturally include more appropriate models (vs 2SLS)

Geneletti et al (2015) [3]; O'Keeffe and Baio (2016) [8]

The importance of being a Bayesian (1)

- Stabilising the estimators
  - 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 ∞
- Computational advantages
  - Estimation of variances and intervals does not rely on asymptotics — just a byproduct of MCMC procedures + can naturally include more appropriate models (vs 2SLS)
- Expand the model to include extra information & deal with the two levels of compliance (GP vs patients)
  - For example, logistic regression models to explain the treatment assignment in terms of practice-level covariates
  - Mixture models to include individual level covariates to account for proxies of compliance with treatment

Geneletti et al (2015) [3]; O'Keeffe and Baio (2016) [8]

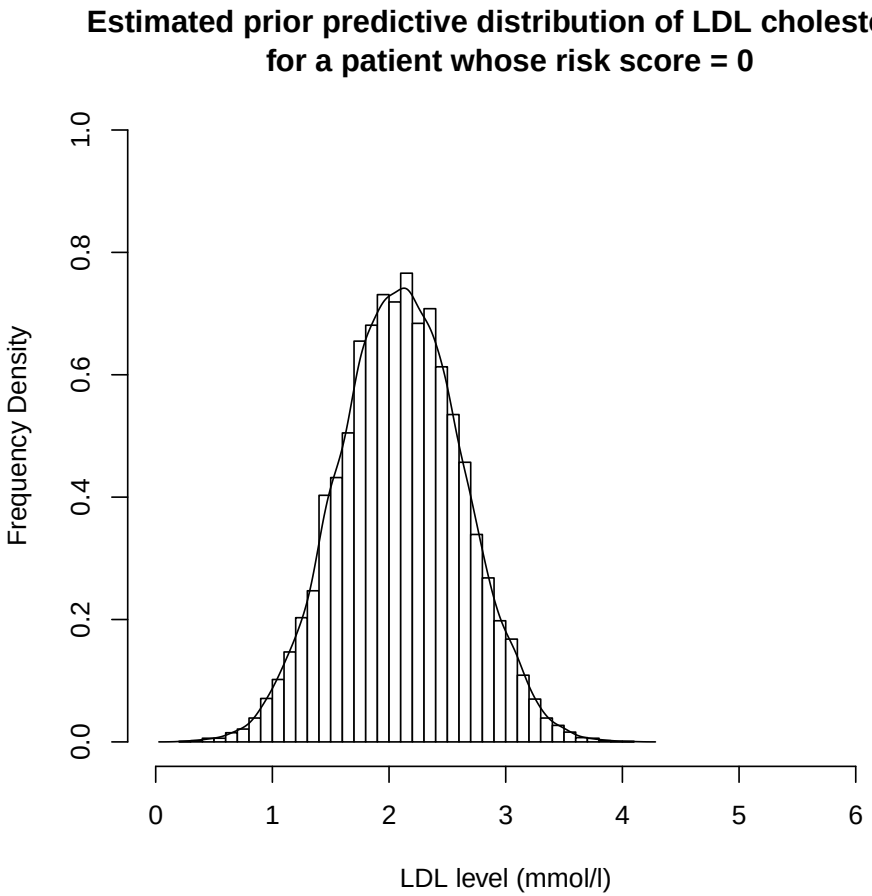
The importance of being a Bayesian (1)

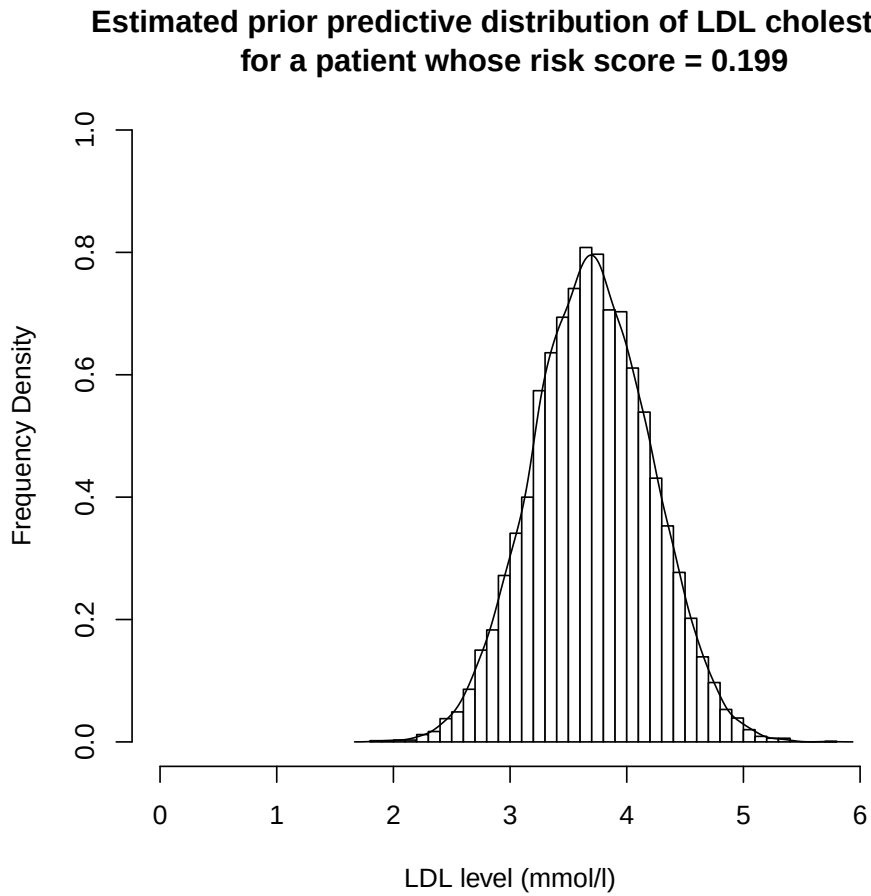
- Stabilising the estimators
  - The denominator of LATE can be very small (*i.e.* $\pi_a \approx \pi_b$)
  - Informative priors on the relevant parameters can encode knowledge and assumptions about these two probabilities so that the resulting estimator does not explode to ∞
- Computational advantages
  - Estimation of variances and intervals does not rely on asymptotics — just a byproduct of MCMC procedures + can naturally include more appropriate models (vs 2SLS)
- Expand the model to include extra information & deal with the two levels of compliance (GP vs patients)
  - For example, logistic regression models to explain the treatment assignment in terms of practice-level covariates
  - Mixture models to include individual level covariates to account for proxies of compliance with treatment
- Cooler! 

Geneletti et al (2015) [3]; O'Keeffe and Baio (2016) [8]

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)
  - Relatively small variance to represent strong belief in the trials

Ward et al (2007) [10]

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)
     - Relatively small variance to represent strong belief in the trials

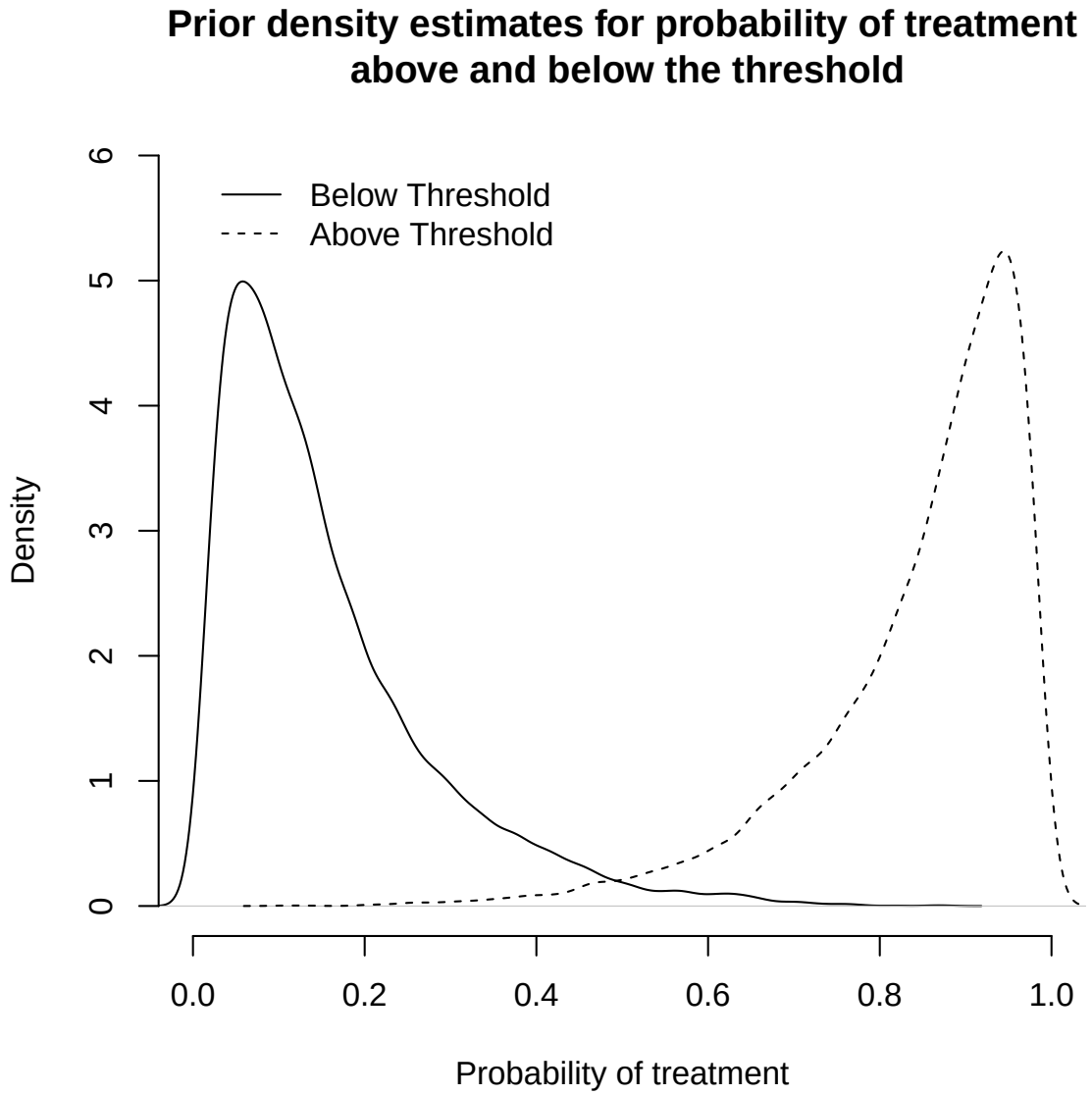
3. **Informative prior on the probability of treatment**:

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

   - **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

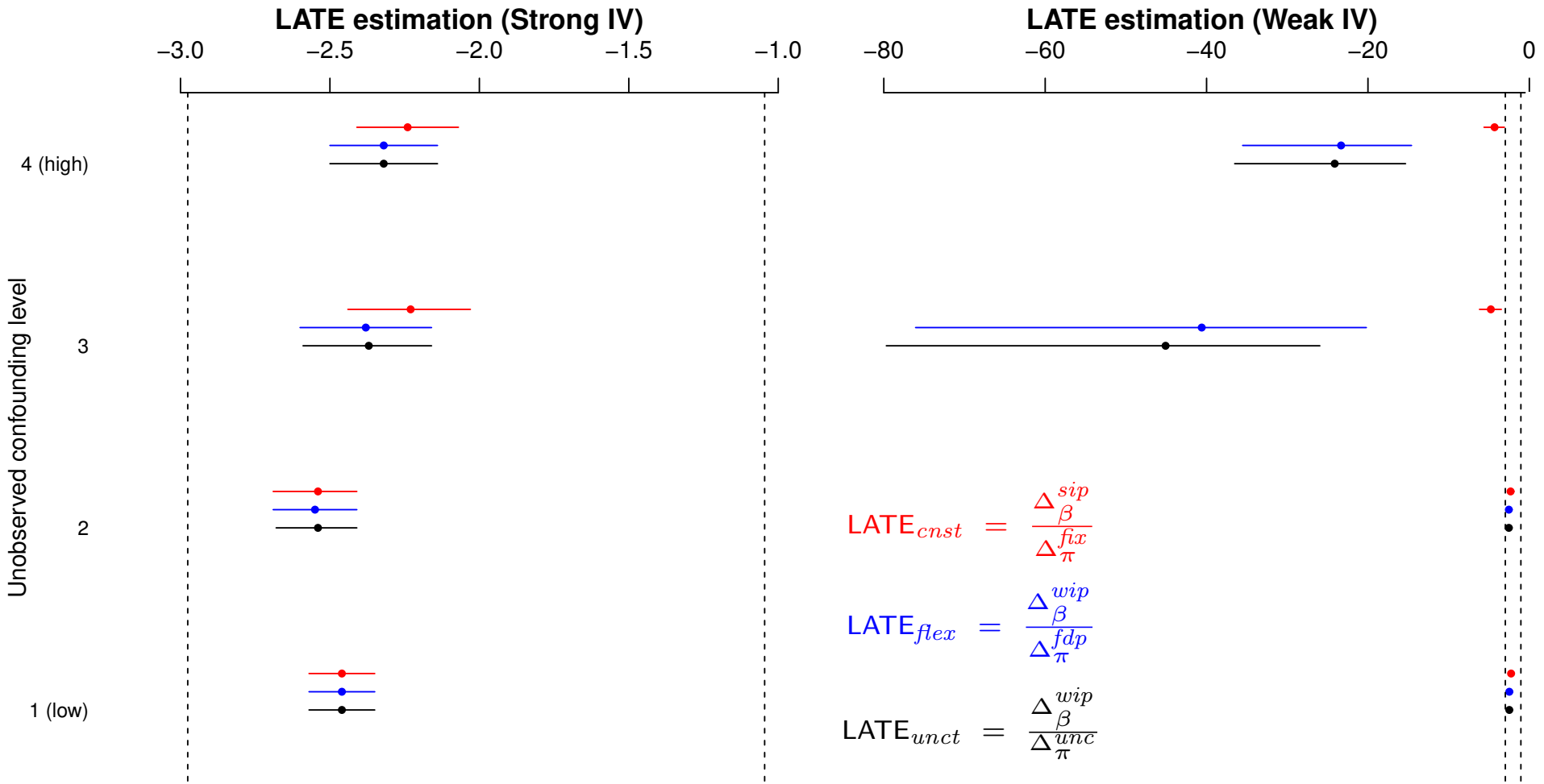
Ward et al (2007) [10]

Bayesian modelling: $\text{logit}(\pi_l)$



Results

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



The importance of being a Bayesian (2)

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

Geneletti et al (2016) [4]

The importance of being a Bayesian (2)

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

- Known issues of standard estimators (*e.g.* generalised method of moments):
  - May give absurde 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 — fairly easy in a Bayesian setting

Geneletti et al (2016) [4]

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

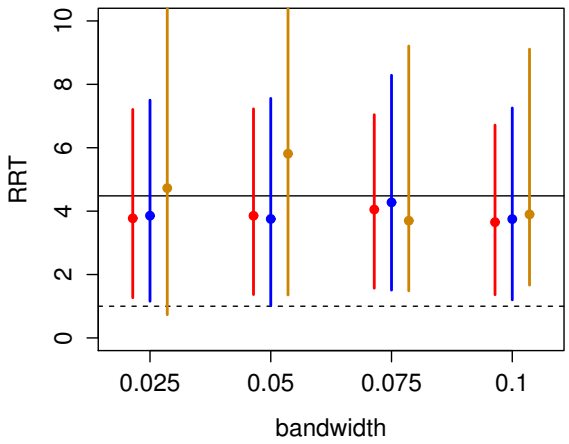
- **But**: can also put a prior on RRT to ensure that it is > 0 and, say, α_a and then induce a prior on α_b, *e.g.*

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

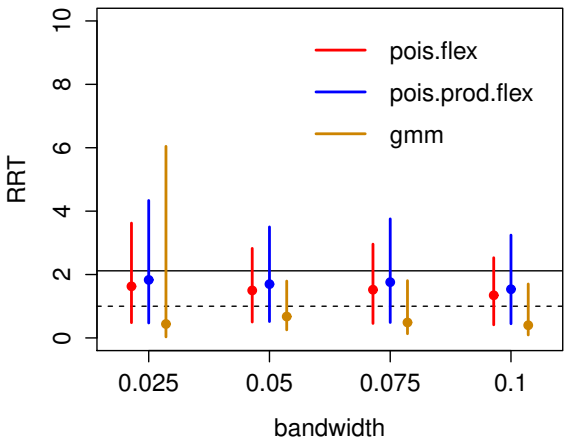
Constraining the models (2)

NB: "Industry standard" methods (based on generalised mehod of moments) fail to give reasonable results in many scenarios

A) High confounding, Weak IV, RR= 4.48

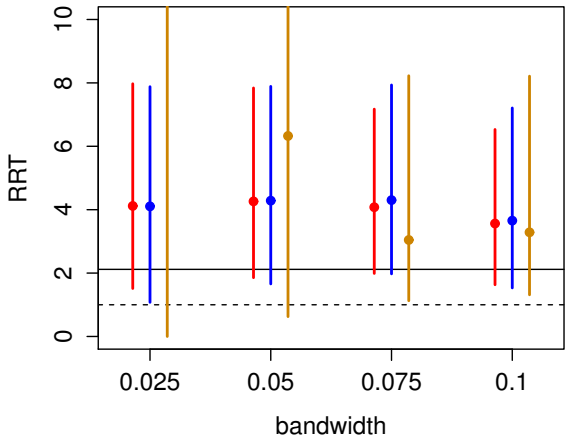


B) High confounding, Weak IV, RR= 2.12

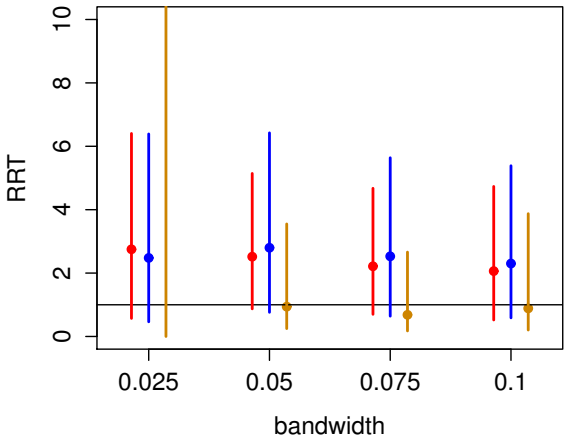


$\Rightarrow$ based on $\mathsf{E}(Y\bar{T} \mid Z)$
$\Rightarrow$ based on $\mathsf{E}(Y \mid \bar{T}, Z)\mathsf{E}(\bar{T} \mid Z)$

C) Low confounding, Weak IV, RR= 2.12



D) High confounding, Weak IV, RR= 1



Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

1. Producing results using data within a number of different bandwidths, that may be selected also with the guidance of an expert in the field of study.

Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

1. Producing results using data within a number of different bandwidths, that may be selected also with the guidance of an expert in the field of study.
2. Selecting an "optimal" bandwidth aimed at minimizing an error term related to the estimation of the effect in a non-parametric fashion.

Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

1. Producing results using data within a number of different bandwidths, that may be selected also with the guidance of an expert in the field of study.

2. Selecting an "optimal" bandwidth aimed at minimizing an error term related to the estimation of the effect in a non-parametric fashion.

   CV A **Cross Validation** based approach

Des Jardins & McCall (2008) [2]

Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

1. Producing results using data within a number of different bandwidths, that may be selected also with the guidance of an expert in the field of study.

2. Selecting an "optimal" bandwidth aimed at minimizing an error term related to the estimation of the effect in a non-parametric fashion.
   - CV A **Cross Validation** based approach
   - IK A **Mean Square Error** minimization based method, designed to give unbiased *point* estimator for the effect;

Des Jardins & McCall (2008) [2]; Imbens & Kalyanaraman (2011) [5]

Bandwidth selection

Bandwidth selection for RDD is addressed in the literature in two main ways:

1. Producing results using data within a number of different bandwidths, that may be selected also with the guidance of an expert in the field of study.

2. Selecting an "optimal" bandwidth aimed at minimizing an error term related to the estimation of the effect in a non-parametric fashion.
   - CV A **Cross Validation** based approach
   - IK A **Mean Square Error** minimization based method, designed to give unbiased *point* estimator for the effect;
   - CCT A **bias-correction** and robust inference method recently, focusing on getting an unbiased *interval* estimator for the effect.

Des Jardins & McCall (2008) [2]; Imbens & Kalyanaraman (2011) [5]; Calonico et al (2015) [1]

What's wrong with that?

Fail to account properly for the **real** issue — **exchangeability**

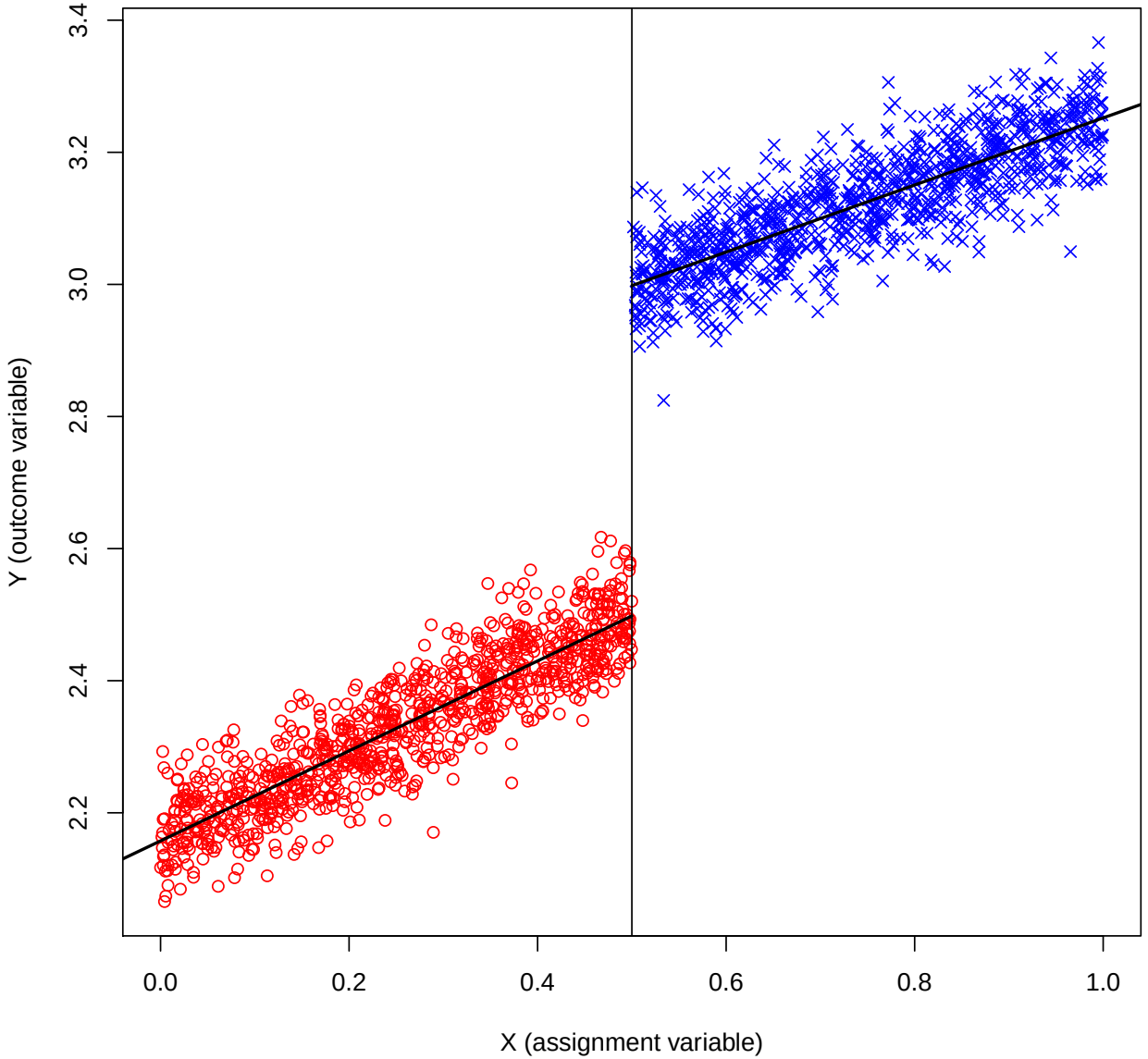
```
set.seed(90)                                  # sets the random number generator seed
tr <- 0.5                                     # sets the value for the threshold
X <- runif(2000,0,1)                          # generates assignment variable
Z <- as.numeric(X>tr)                         # generates treatment indicator
X.c <- X-tr                                   # centers assignment variable
e <- rnorm(2000, 0, .05)                      # generates random error (white noise)
Y <- Z*(3+0.5*X.c) + (1-Z)*(2.5+0.7*X.c) + e  # generates the outcome variable
```

NB: In this case

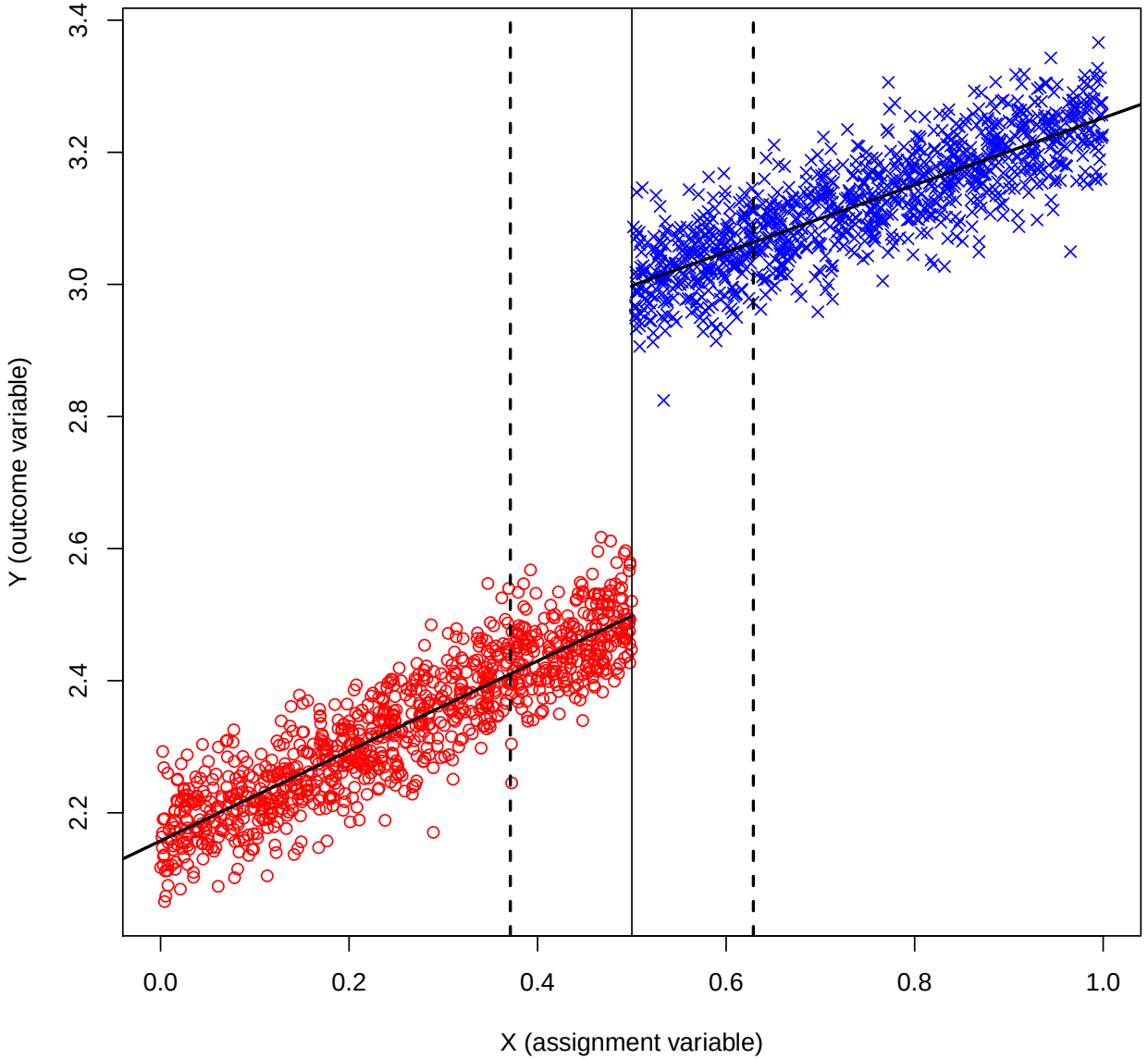
- The design is sharp
- No unobserved confounders
- The outcome is fully determined by the forcing variable

$\Rightarrow$ observations are (conditionally) exchangeable below and above the threshold!
$\Rightarrow$ shouldn't we be able to use **all** (most?) the data to estimate the causal effect (and gain precision)?

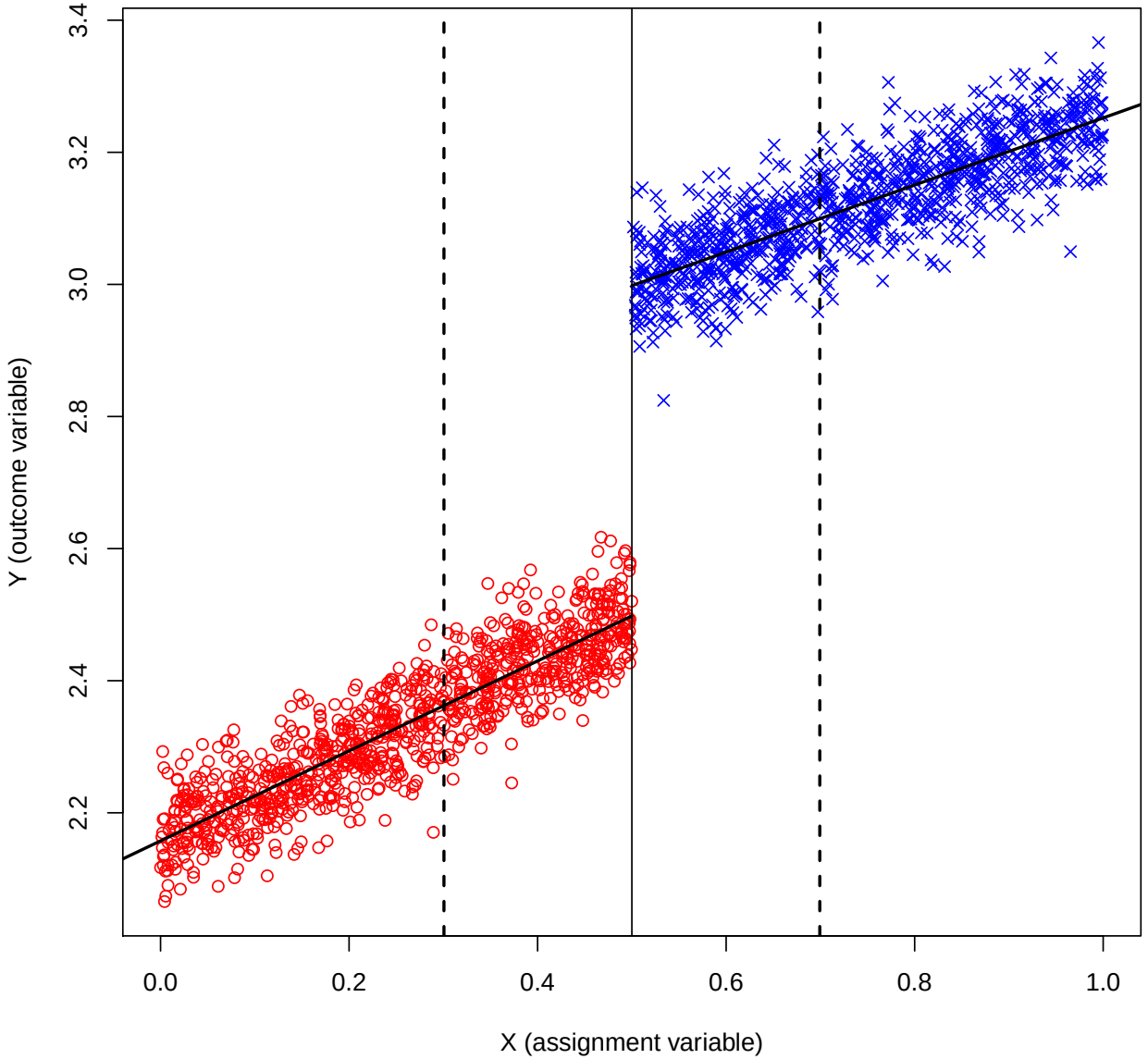
Use them all?


Y (outcome variable)
2.2
2.4
2.6
2.8
3.0
3.2
3.4
0.0
0.2
0.4
0.6
0.8
1.0
X (assignment variable)

Use them all? Default optimisation (min MSE)



Use them all? Alternative optimisation (sum for regression estimates)



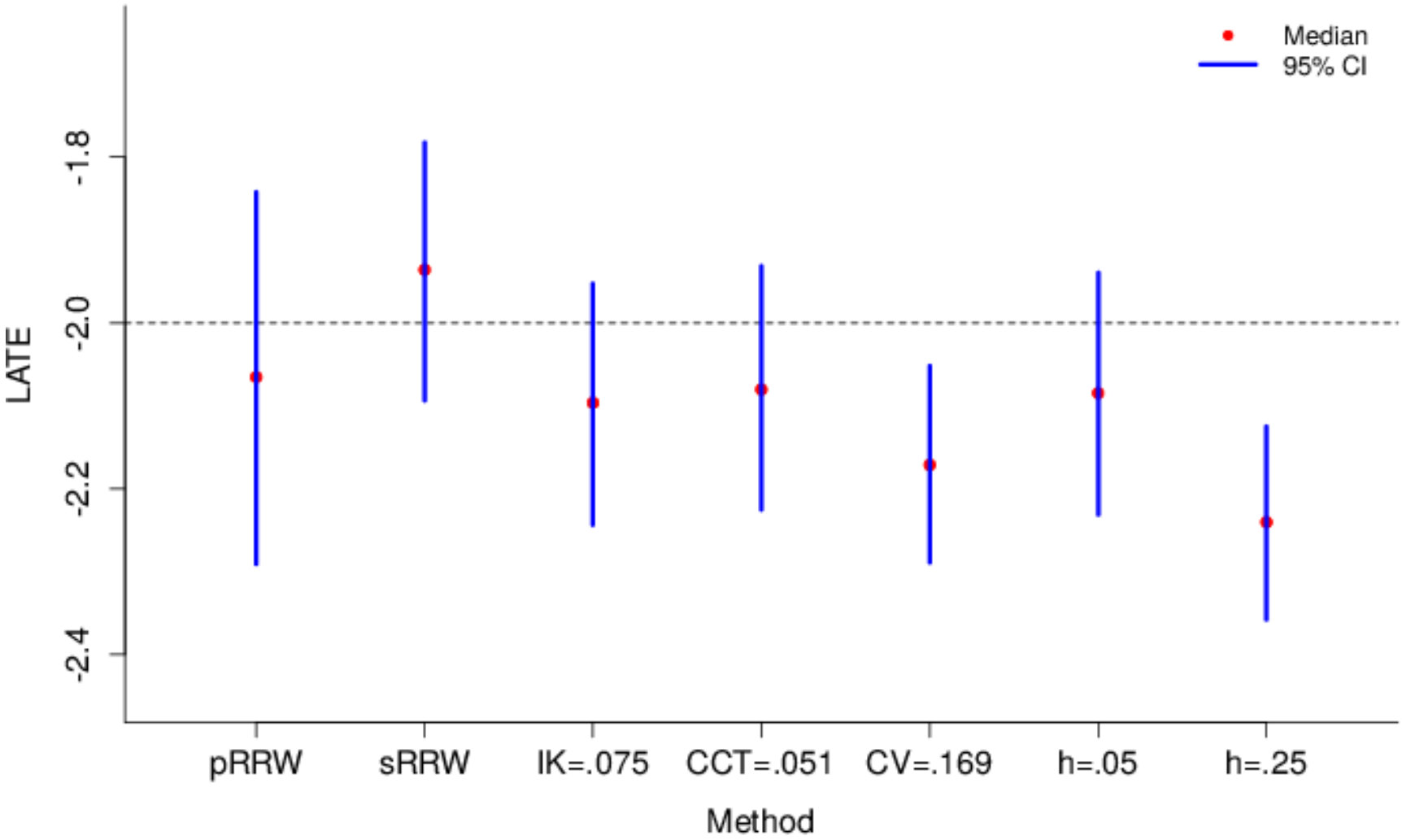
Alternative methods?

- Possibly use flexible regressions (*e.g.* splines) — but standard setting may not be flexible enough...
  - We may **know** that individuals at either extreme really are too different and do not want them to basically matter at all...

Alternative methods?

- Possibly use flexible regressions (*e.g.* splines) — but standard setting may not be flexible enough...
  - We may **know** that individuals at either extreme really are too different and do not want them to basically matter at all...
- Other avenues?
  - **Reverse random walk priors** — anchor priors to one extreme and "filter" irrelevant data [work in progress]
  - Spatially structured models — formally account for spatial distance from the threshold [work in progress?!]
  - Clustering — individuals with similar characteristics are clustered together and can be used as "exchageable" [work in progress]
- More importantly, selection should happen according to balancing in the confounders above & below the threshold!

Some tentative results



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

References

[1] S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295 – 2326, 2015.

[2] S. DesJardins and B. McCall. The impact of the gates millennium scholars program on the retention, college finance and work-related choices, and future educational aspirations of low-income minority students. 2008.

[3] S. Geneletti, AG O'Keeffe, L. D. Sharples, S. Richardson, and G. Baio. Bayesian regression discontinuity designs: incorporating clinical knowledge in the causal analysis of primary care data. *Statistics in Medicine*, 2015.

[4] S. Geneletti, F. Ricciardi, AG O'Keeffe, and G. Baio. Bayesian modelling for binary outcomes in the regression discontinuity design. *Submitted to JASA*, 2016.

[5] G. Imbens and K. Kalyanaraman. Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 2011.

[6] NICE. *Quick reference guide: Statins for the prevention of cardiovascular events*, 2008.

[7] NICE. `https://www.nice.org.uk/guidance/cg181`, (accessed 7th November 2016).

[8] AG O'Keeffe and G Baio. Approaches to the estimation of the local average treatment effect in a regression discontinuity design. *Scandinavian Journal of Statistics*, 2016.

[9] 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.

[10] 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.

Thank you!