The Regression Discontinuity Design In Epidemiology: An Application To Statins

Gianluca Baio[1] Sara Geneletti[2] Aidan O'Keeffe[1]

[1] University College London

[2] London School of Economics and Political Science

gianluca@stats.ucl.ac.uk

LSHTM Centre for Statistical Methodology Seminar

Friday 29 November 2013

Ackowledgments

- This talk describes preliminary work and future plans for a project funded by the UK Medical Research Council "*The regression discontinuity design: a novel approach to evaluating the effects of drugs and treatments in primary care*"
- The project has started earlier this year (in September) and the research team spans across UCL, LSE and MRC Biostatistics Unit, Cambridge

- More info and description available at the webpage `www.statistica.it/gianluca/RDD`

Outline of presentation

1. **What is the Regression Discontinuity Design?**
   - Brief intro
   - Sharp vs Fuzzy RD
   - Assumptions
2. RD design applied to statin prescriptions in THIN data
   - Bayesian modelling
   - Simulations
   - Some results
3. Further work & conclusions

Outline of presentation

1. What is the Regression Discontinuity Design?
   - Brief intro
   - Sharp vs Fuzzy RD
   - Assumptions
2. **RD design applied to statin prescriptions in THIN data**
   - Bayesian modelling
   - Simulations
   - Some results
3. Further work & conclusions

Outline of presentation

1. What is the Regression Discontinuity Design?
   - Brief intro
   - Sharp vs Fuzzy RD
   - Assumptions
2. RD design applied to statin prescriptions in THIN data
   - Bayesian modelling
   - Simulations
   - Some results
3. **Further work & conclusions**

What is the RD design?

- The regression discontinuity (RD) design was first introduced in the educational econometrics literature in the 1960's (Thistlethwaite and Campbell, 1960)
- The RD design has not been extensively treated in the epidemiology literature

What is the RD design?

- The regression discontinuity (RD) design was first introduced in the educational econometrics literature in the 1960's (Thistlethwaite and Campbell, 1960)
- The RD design has not been extensively treated in the epidemiology literature

- Recently other econometricians have become interested in formal causal aspects (Imbens and Lemieux, 2008; van der Klaauw, 2008)
- The original idea was to exploit policy thresholds to estimate the causal effect of an educational intervention

What is the RD design?

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

What is the RD design?

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 Design

- The idea of the RD design is that the threshold behaves like a **randomising device**
- If we imagine that the thresholds are adhered to **very strictly** (which is usually termed **sharp** design), then we can think of the RD design as removing the confounding due to unobserved factors, *eg*
  - Academic history, talent, motivation
  - Unobserved health/personal characteristics

RD design and confounding

Sharp Design

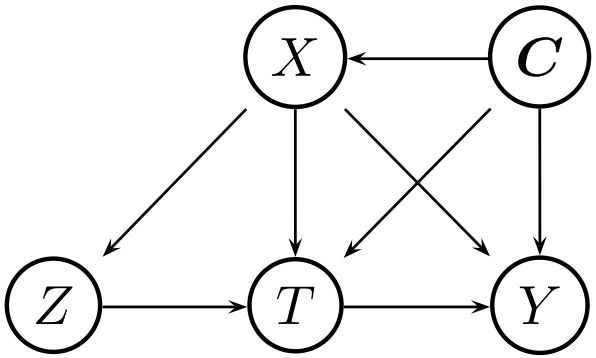
- The idea of the RD design is that the threshold behaves like a **randomising device**
- If we imagine that the thresholds are adhered to **very strictly** (which is usually termed **sharp** design), then we can think of the RD design as removing the confounding due to unobserved factors, *eg*
  - Academic history, talent, motivation
  - Unobserved health/personal characteristics

Fuzzy Design

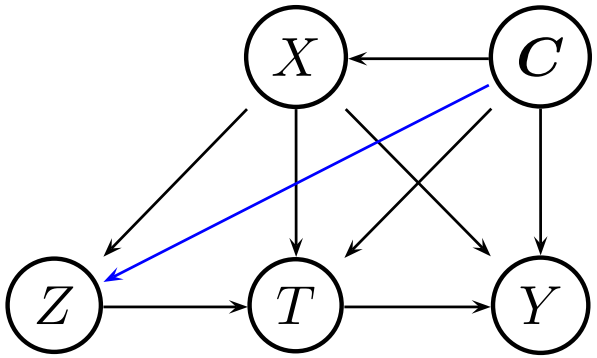
- Particularly in public health contexts the sharp threshold is unlikely to be adhered to (a situation termed **fuzzy** design)
- For example, often GPs override guidelines — generally because, contrary to their recommendations, they feel that patients will benefit from medication
  - Links with economic theory of asymmetric information

RD design and confounding

Sharp RD



Fuzzy RD



- $X =$ guideline variable (*eg* cardiovascular risk score)
- $Z =$ threshold indicator (*ie* $Z = 1$ if $X > x_0$ and 0 otherwise)
- $T =$ treatment administered (prescribed)
- $\boldsymbol{C} = (\boldsymbol{O} \cup \boldsymbol{U}) =$ set of **confounders**
  - $\boldsymbol{O}$ fully observed (*eg* sex, age)
  - $\boldsymbol{U}$ fully or partially unobserved (*eg* smoking status)
- $Y =$ continuous outcome (*eg* LDL cholesterol level)

The RD design — assumptions

A1. *Association of treatment with the threshold indicator*: $Z \not\perp\!\!\!\perp T$

- Can be directly tested from the observed data

The RD design — assumptions

A1. *Association of treatment with the threshold indicator*: $Z \not\perp\!\!\!\perp T$

- Can be directly tested from the observed data

A2. *Independence of guidelines*: $Z \perp\!\!\!\perp \boldsymbol{C}$

- Generally plausible, as the threshold is set by the powers-that-be, *eg* governmental agencies
- **NB**: This assumption does not necessarily hold in its "strong" form, as prescription can be done according to some extra criteria. In this case, a "weaker" form can be used to imply $Z \perp\!\!\!\perp \boldsymbol{U} \mid \boldsymbol{O}$

The RD design — assumptions

A1. *Association of treatment with the threshold indicator*: $Z \not\perp\!\!\!\perp T$

- Can be directly tested from the observed data

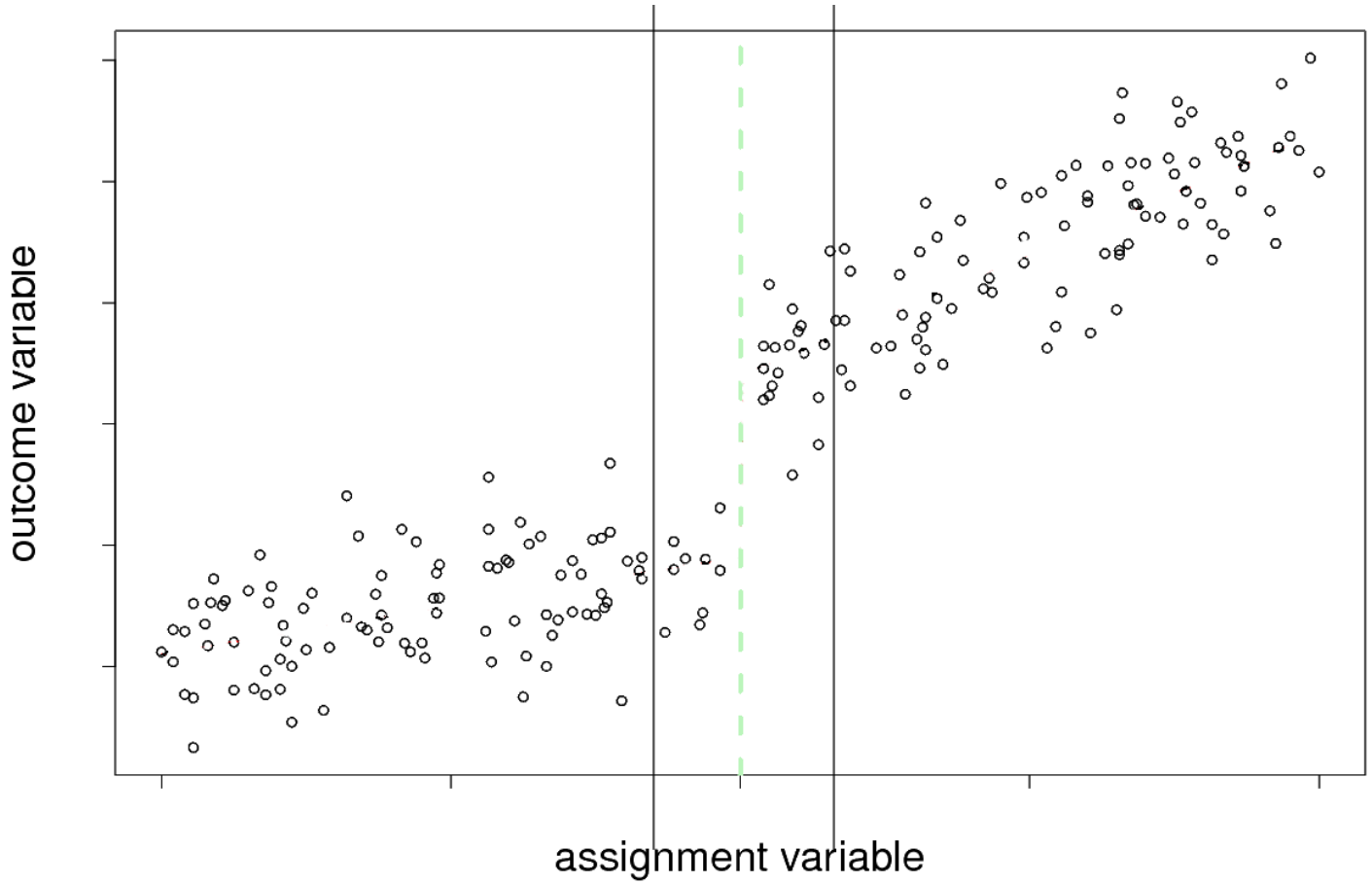
A2. *Independence of guidelines*: $Z \perp\!\!\!\perp \boldsymbol{C}$

- Generally plausible, as the threshold is set by the powers-that-be, *eg* governmental agencies
- **NB**: This assumption does not necessarily hold in its "strong" form, as prescription can be done according to some extra criteria. In this case, a "weaker" form can be used to imply $Z \perp\!\!\!\perp \boldsymbol{U} \mid \boldsymbol{O}$

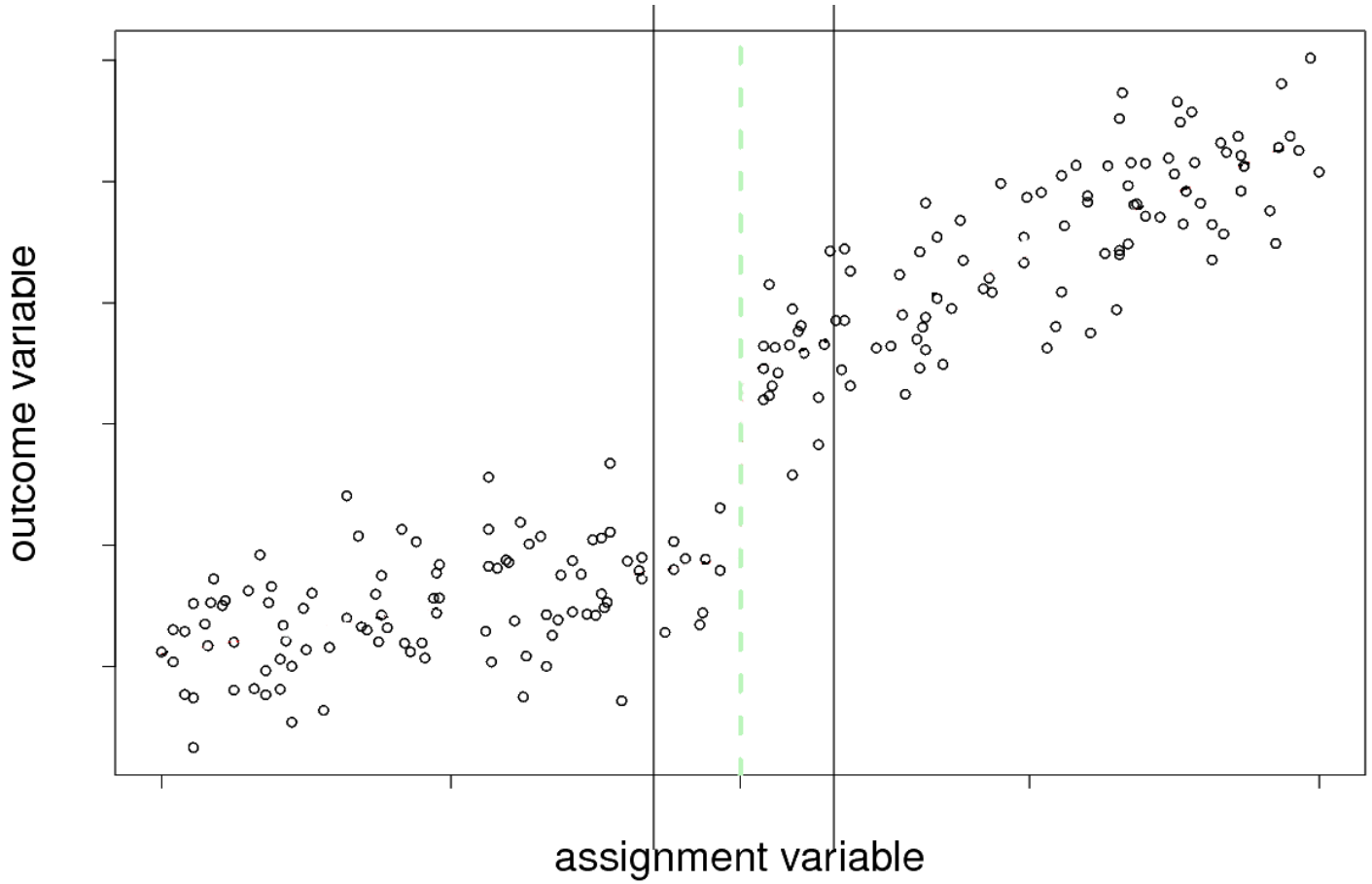
A3. *Unconfoundedness*: $Y \perp\!\!\!\perp Z \mid (T, X, \boldsymbol{C})$

- Implies that the individuals just above and below the threshold are "similar" (exchangeable)
- This assumption is violated if individuals can change their outcome to fall on either side (*eg* benefit fraud: individuals might say their income is below a threshold in order to fall into a category that receives benefits)
- **NB**: this assumption can be expressed equivalently by considering the threshold as a **randomising device**, thus a comparison of above and below gives us a **causal effect estimate of the treatment**, at the threshold

The RD design — assumptions (cont'd)



The RD design — assumptions (cont'd)



- **But**: how far above and below the threshold?

The RD design — assumptions (cont'd)

A4. *Continuity*: $\mathrm{E}(Y \mid Z, X = x, \boldsymbol{C})$ is continuous in x (at x_0) for $T = 0, 1$

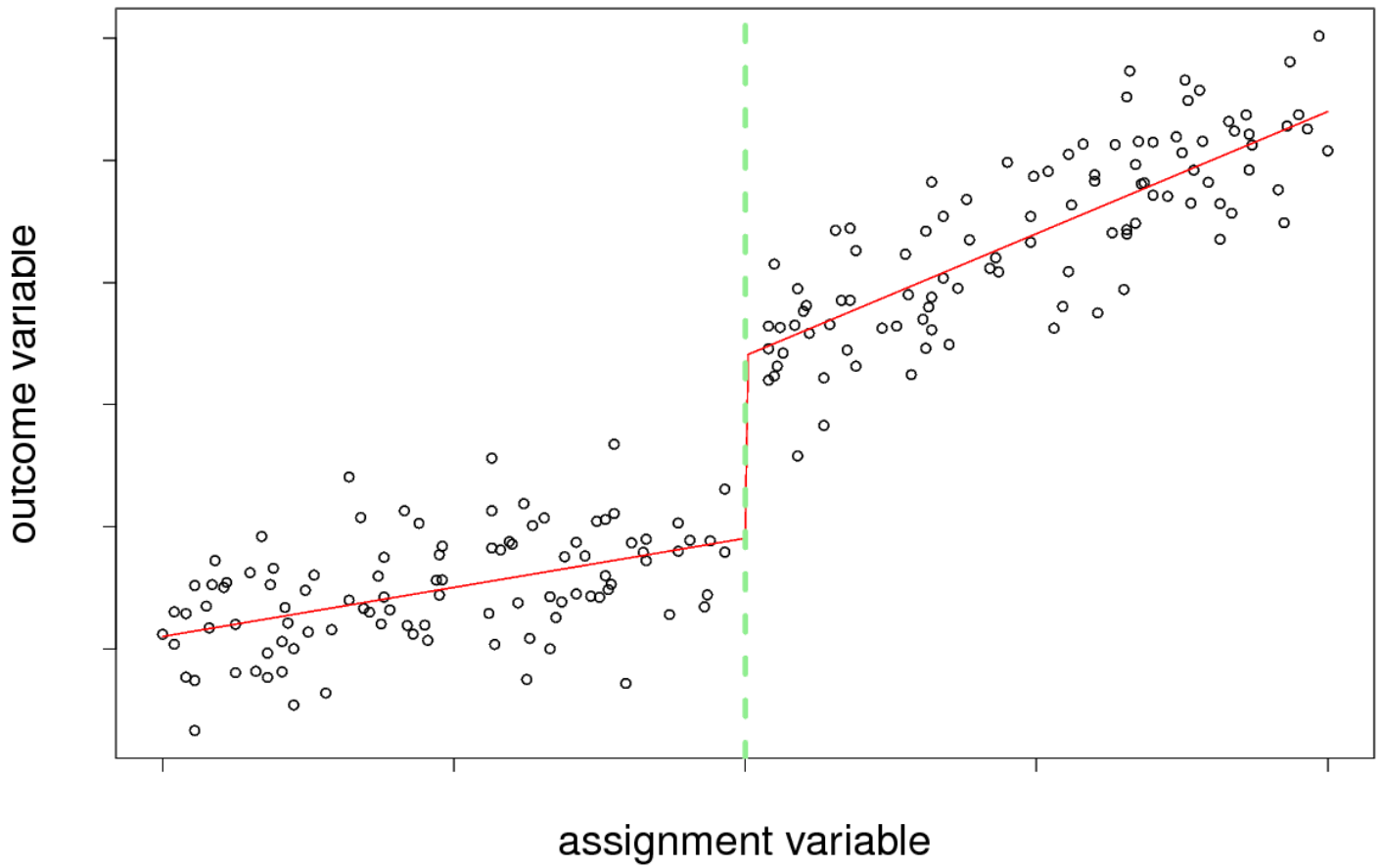
- We can fit two separate regressions, one above and one below the threshold; or can assume a common slope and fit one regression (this assumes effect is the same everywhere)

The RD design — assumptions (cont'd)

A4. *Continuity*: $\mathrm{E}(Y \mid Z, X = x, \boldsymbol{C})$ is continuous in x (at x_0) for $T = 0, 1$

- We can fit two separate regressions, one above and one below the threshold; or can assume a common slope and fit one regression (this assumes effect is the same everywhere)

Marginal conditional expectation $\mathrm{E}(Y \mid X)$



The continuous case: Sharp threshold

- Let Y be the outcome, X^c the **centered** assignment variable and T the treatment indicator
- If the regressions are given by

$$\mathsf{E}(Y_l) = \beta_{0l} + \beta_{1l} X_l^c$$

and:

- $x_0^c = 0$ is the value of X_l^c at the threshold;
- $l = b \Rightarrow X_l^c < x_0^c$ (below)
- $l = a \Rightarrow X_l^c \geq x_0^c$ (above)

then an estimate of the causal effect of the treatment is

$$\begin{aligned} \mathsf{ATE} &= \mathsf{E}(Y|T=1) - E(Y|T=0) \\ &= \beta_{0a} - \beta_{0b} := \Delta_\beta \end{aligned}$$

The continuous case: Fuzzy threshold

- In this case, we also need to consider the threshold indicator, Z
- The formula for the fuzzy estimator is

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

  where π_l is an estimate of $\Pr(T=1|Z=z)$, *eg* the chance of being treated when above or below the threshold
- π_l can be based on the raw frequencies (not recommended!) or modelled more formally
- LATE is based on the compliance literature (Dawid, 2003)

The continuous case: Fuzzy threshold

- In this case, we also need to consider the threshold indicator, Z
- The formula for the fuzzy estimator is

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

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

- π_l can be based on the raw frequencies (not recommended!) or modelled more formally
- LATE is based on the compliance literature (Dawid, 2003)

RD design and "compliance"

- In applications of the RD design in the GP prescription context there are two layers of **compliance**
  1. Adherence of GPs to prescription guidelines (*ie* only give the antiretroviral drug to patients with CD4 count below 200 cells/mm^3)
  2. Compliance of patients to prescription (*ie* always take the antiretroviral drug twice a day every day)

RD design and "compliance"

- In applications of the RD design in the GP prescription context there are two layers of **compliance**
  1. Adherence of GPs to prescription guidelines (*ie* only give the antiretroviral drug to patients with CD4 count below 200 cells/mm^3)
  2. Compliance of patients to prescription (*ie* always take the antiretroviral drug twice a day every day)

- **NB**: the RD design is related to compliance of the first type
- The RD's relationship with compliance means that, in its standard form, it is also related to intention-to-treat experiments
- Nevertheless, both types can be taken into account by specifying the analysis methods
  - Links to Bayesian models to specify informative priors and/or selection models

The RD design — assumptions (cont'd)

A5. *Monotonicity* (fuzzy design only)

- No decision-maker systematically defies the guidelines
- For example, if we consider a pair of binary **strategies** (S_a, S_b) for $a$bove and $b$elow the threshold, this is equivalent to assuming that

$$\Pr(S_a = 0, S_b = 1) = 0$$

- This has links with the counterfactual framework (*eg* "no defiers"), but can be re-framed in a fully decision-theoretic setting

A5. *Monotonicity* (fuzzy design only)

- No decision-maker systematically defies the guidelines
- For example, if we consider a pair of binary **strategies** (S_a, S_b) for $a$bove and $b$elow the threshold, this is equivalent to assuming that

$$\Pr(S_a = 0, S_b = 1) = 0$$

- This has links with the counterfactual framework (*eg* "no defiers"), but can be re-framed in a fully decision-theoretic setting

NB: In the case of our example, this is probably **less** problematic

- Unlike patients, GPs *should* be more aware of the current guidelines and decide in a more rational way
- It is possible that they decide to overrule the guideline (for good reasons!) but on average they should reasonably follow them
- Patients compliance is a more difficult issue!

The RD design — assumptions (cont'd)

A5. *Monotonicity* (fuzzy design only)

- No decision-maker systematically defies the guidelines
- For example, if we consider a pair of binary **strategies** (S_a, S_b) for $a$bove and $b$elow the threshold, this is equivalent to assuming that

$$\Pr(S_a = 0, S_b = 1) = 0$$

- This has links with the counterfactual framework (*eg* "no defiers"), but can be re-framed in a fully decision-theoretic setting

NB: In the case of our example, this is probably **less** problematic

- Unlike patients, GPs *should* be more aware of the current guidelines and decide in a more rational way
- It is possible that they decide to overrule the guideline (for good reasons!) but on average they should reasonably follow them
- Patients compliance is a more difficult issue!

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 (Ward et al., 2007)
- UK NHS guidelines are to prescribe statins to individuals without previous CVD if their 10 year CVD score exceeds 20% (NICE, 2008)

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 (Ward et al., 2007)
- UK NHS guidelines are to prescribe statins to individuals without previous CVD if their 10 year CVD score exceeds 20% (NICE, 2008)

- **Objective**: use the RD design to evaluate
  - Whether the effect of statins is the same as in the trials
  - Whether the threshold is currently set at the point in which most patients have the highest benefit. This has also clear health economic implications

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 (Ward et al., 2007)
- UK NHS guidelines are to prescribe statins to individuals without previous CVD if their 10 year CVD score exceeds 20% (NICE, 2008)

- **Objective**: use the RD design to evaluate
  - Whether the effect of statins is the same as in the trials
  - Whether the threshold is currently set at the point in which most patients have the highest benefit. This has also clear health economic implications

- **Data**: Simulation study, based on a real clinical practice database containing routine GP prescriptions as well as information on the variables that determine them
  - 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)

Going Bayesian

There are several reasons for framing the problem within the Bayesian approach

Going Bayesian

There are several reasons for framing the problem within the Bayesian approach

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

Going Bayesian

There are several reasons for framing the problem within the Bayesian approach

- Stabilising the estimators
  - The denominator of LATE can be very small (*ie* $\pi_a \approx \pi_b$)
  - 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

Going Bayesian

There are several reasons for framing the problem within the Bayesian approach

- Stabilising the estimators
  - The denominator of LATE can be very small (*ie* $\pi_a \approx \pi_b$)
  - 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
- Expand the model to include extra information & deal with the two levels of compliance
  - 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

Going Bayesian

There are several reasons for framing the problem within the Bayesian approach

- Stabilising the estimators
  - The denominator of LATE can be very small (*ie* $\pi_a \approx \pi_b$)
  - 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
- Expand the model to include extra information & deal with the two levels of compliance
  - 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! 

Models for the ATE

- We model LDL cholesterol (y) as a function of the threshold indicator

$$y_{il} \sim \text{Normal}(\mu_{il}, \sigma^2)$$
$$\mu_{il} = \beta_{0l} + \beta_{1l} x_{il}^c \quad \Rightarrow \quad \text{ATE} = \Delta_\beta = \beta_{0a} - \beta_{0b}$$

Models for the ATE

- We model LDL cholesterol (y) as a function of the threshold indicator

$$y_{il} \sim \text{Normal}(\mu_{il}, \sigma^2)$$
$$\mu_{il} = \beta_{0l} + \beta_{1l} x^c_{il} \quad \Rightarrow \quad \text{ATE} = \Delta_\beta = \beta_{0a} - \beta_{0b}$$

- We choose a weakly informative prior for the individual sd: $\sigma \sim \text{Uniform}(0, 5)$
  - Especially with (at least moderately) large datasets, likely to not affect posteriors by much
  - Genuine information or other form can be used

Models for the ATE

- We model LDL cholesterol (y) as a function of the threshold indicator

$$y_{il} \sim \text{Normal}(\mu_{il}, \sigma^2) \qquad \Rightarrow \qquad \text{ATE} = \Delta_\beta = \beta_{0a} - \beta_{0b}$$
$$\mu_{il} = \beta_{0l} + \beta_{1l} x^c_{il}$$

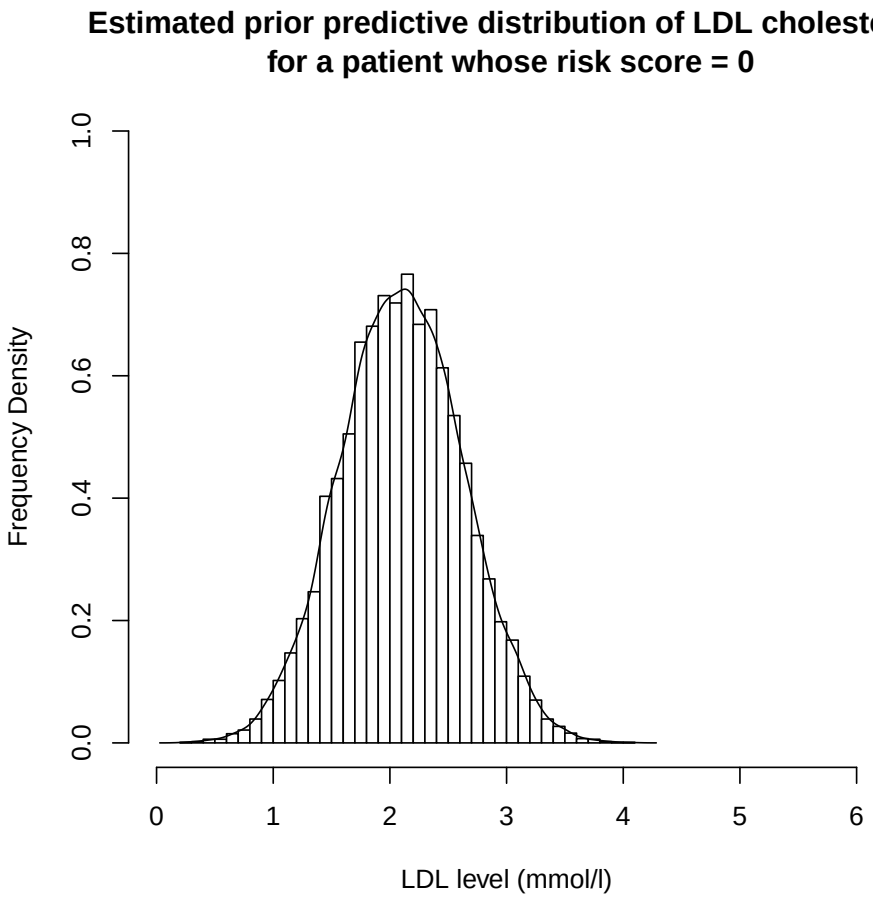
- We choose a weakly informative prior for the individual sd: $\sigma \sim \text{Uniform}(0, 5)$
  - Especially with (at least moderately) large datasets, likely to not affect posteriors by much
  - Genuine information or other form can be used

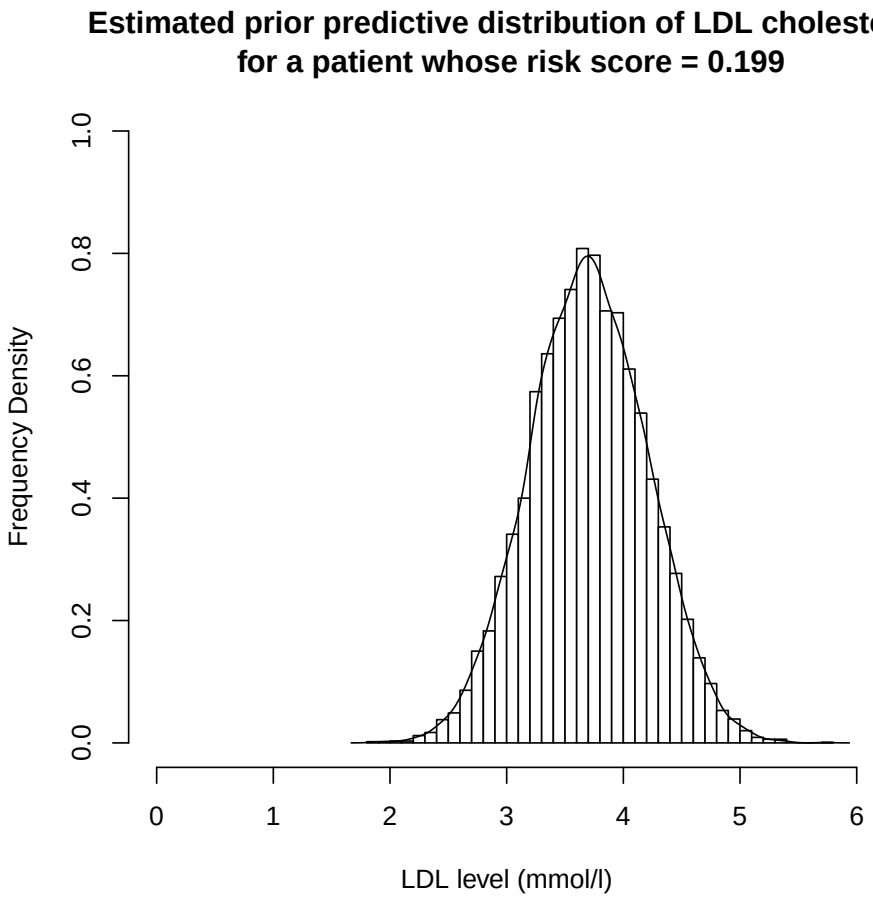
- Then we model the coefficients for the regression below the threshold as

$$\beta_{0b} \sim \text{Normal}(m_0, s_0^2) \qquad \text{and} \qquad \beta_{1b} \sim \text{Normal}(m_{1b}, s_{1b}^2)$$

  - The parameters (m_0, s_0^2) and (m_{1b}, s_{1b}^2) are chosen to induce reasonable values for the estimated LDL in correspondence with centered risk scores in the range $[-0.2; 0]$

Models for the ATE (cont'd)





NB: The selected values are $m_0 = 3.7$, $m_{1b} = 8$ and $s_0 = 0.5$

Models for the ATE (cont'd)

- Finally we model the coefficients for the regression below the threshold as

$$\beta_{0a} = \beta_{0b} + \phi \qquad \text{and} \qquad \beta_{1a} \sim \text{Normal}(m_{1a}, s^2_{1a})$$

  - The parameters of β_{1a} encode the assumption for individuals with very high risk score, the effect is marginally lower than for those closer to the threshold
  - ϕ is the "jump" due to the causal effect of the treatment

Models for the ATE (cont'd)

- Finally we model the coefficients for the regression below the threshold as

$$\beta_{0a} = \beta_{0b} + \phi \qquad \text{and} \qquad \beta_{1a} \sim \text{Normal}(m_{1a}, s^2_{1a})$$

  - The parameters of β_{1a} encode the assumption for individuals with very high risk score, the effect is marginally lower than for those closer to the threshold
  - ϕ is the "jump" due to the causal effect of the treatment

- **Weakly informative prior**: $\phi \sim \text{Normal}(0, 2)$
  - "Sketpical" prior on the effect of treatment, which is assumed to be null

Models for the ATE (cont'd)

- Finally we model the coefficients for the regression below the threshold as

$$\beta_{0a} = \beta_{0b} + \phi \qquad \text{and} \qquad \beta_{1a} \sim \text{Normal}(m_{1a}, s_{1a}^2)$$

  - The parameters of β_{1a} encode the assumption for individuals with very high risk score, the effect is marginally lower than for those closer to the threshold
  - ϕ is the "jump" due to the causal effect of the treatment

- **Weakly informative prior**: $\phi \sim \text{Normal}(0, 2)$
  - "Sketpical" 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

Models for the ATE (cont'd)

- Finally we model the coefficients for the regression below the threshold as

$$\beta_{0a} = \beta_{0b} + \phi \qquad \text{and} \qquad \beta_{1a} \sim \text{Normal}(m_{1a}, s^2_{1a})$$

  - The parameters of β_{1a} encode the assumption for individuals with very high risk score, the effect is marginally lower than for those closer to the threshold
  - ϕ is the "jump" due to the causal effect of the treatment

- **Weakly informative prior**: $\phi \sim \text{Normal}(0, 2)$
  - "Sketpical" 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

- We term the resulting ATE estimators obtained under the two different formulations as Δ_β^{wip} and Δ_β^{sip}

Models for the denominator of the LATE

- Model the (sum of the) treatment indicator, for $l = a, b$

$$\sum_{i=1}^{n_l} t_{il} \sim \text{Binomial}(\pi_l, n_l) \quad \Rightarrow \quad \Delta_\pi = \pi_a - \pi_b \text{ and LATE} = \frac{\Delta_\beta}{\Delta_\pi}$$

Models for the denominator of the LATE

- Model the (sum of the) treatment indicator, for $l = a, b$

$$\sum_{i=1}^{n_l} t_{il} \sim \text{Binomial}(\pi_l, n_l) \quad \Rightarrow \quad \Delta_\pi = \pi_a - \pi_b \text{ and LATE} = \frac{\Delta_\beta}{\Delta_\pi}$$

1. Unconstrained prior $\pi_l \sim \text{Beta}(1, 1)$ — the resulting estimator is Δ_π^{unc}

Models for the denominator of the LATE

- Model the (sum of the) treatment indicator, for $l = a, b$

$$\sum_{i=1}^{n_l} t_{il} \sim \text{Binomial}(\pi_l, n_l) \quad \Rightarrow \quad \Delta_\pi = \pi_a - \pi_b \text{ and LATE} = \frac{\Delta_\beta}{\Delta_\pi}$$

1. Unconstrained prior $\pi_l \sim \text{Beta}(1, 1)$ — the resulting estimator is Δ_π^{unc}
2. Fixed difference prior (Δ_π^{fix})

$$\begin{aligned}
\pi_b &\sim \text{Beta}(\alpha_b, \beta_b) && \text{with } \beta_b = (n_b + 1) \Rightarrow \text{all untreated below} \\
\pi_a &\sim \text{Beta}(\alpha_a, \beta_a) && \text{with } \beta_a = 1 \Rightarrow \text{all treated above} \\
\alpha_b &\sim \text{Uniform}(1, U) && \text{for a large value } U \text{ (\textit{eg } } 10\,000) \\
\alpha_a &= \nu + \alpha_b && \text{with } \nu \sim \text{Uniform}(200, 10\,000)
\end{aligned}$$

NB: implies *at least* 200 more subjects are treated above than below

Models for the denominator of the LATE

- Model the (sum of the) treatment indicator, for $l = a, b$

$$\sum_{i=1}^{n_l} t_{il} \sim \text{Binomial}(\pi_l, n_l) \quad \Rightarrow \quad \Delta_\pi = \pi_a - \pi_b \text{ and LATE} = \frac{\Delta_\beta}{\Delta_\pi}$$

1. Unconstrained prior $\pi_l \sim \text{Beta}(1, 1)$ — the resulting estimator is Δ_π^{unc}
2. Fixed difference prior (Δ_π^{fix})

$$\begin{aligned}
&\pi_b \sim \text{Beta}(\alpha_b, \beta_b) && \text{with } \beta_b = (n_b + 1) \Rightarrow \text{all untreated below} \\
&\pi_a \sim \text{Beta}(\alpha_a, \beta_a) && \text{with } \beta_a = 1 \Rightarrow \text{all treated above} \\
&\alpha_b \sim \text{Uniform}(1, U) && \text{for a large value } U \text{ (\textit{eg } } 10\,000) \\
&\alpha_a = \nu + \alpha_b && \text{with } \nu \sim \text{Uniform}(200, 10\,000)
\end{aligned}$$

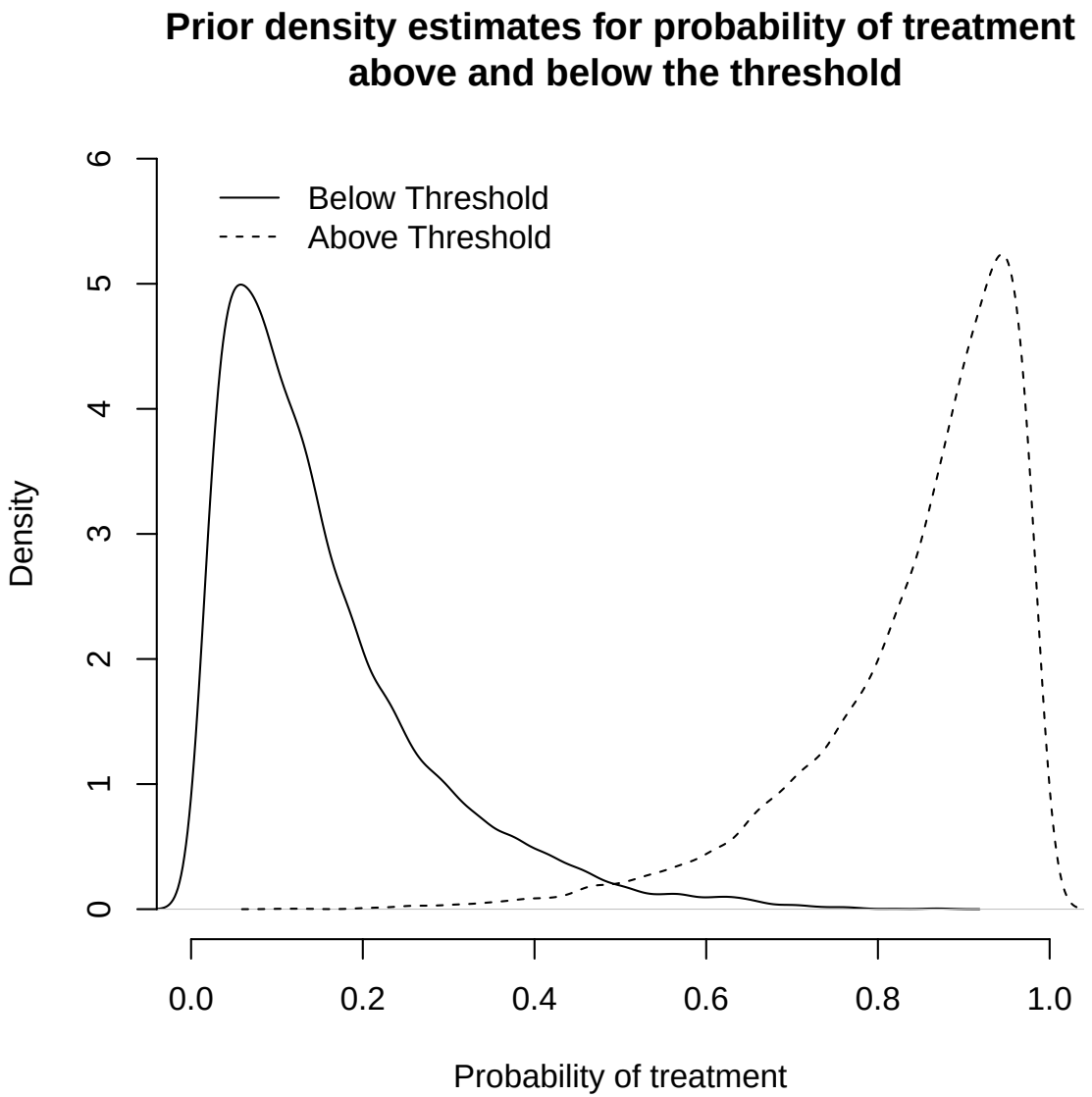
NB: implies *at least* 200 more subjects are treated above than below

3. Flexible difference prior (Δ_π^{fdp})

$$\theta_a \sim \text{Normal}(2, 1), \qquad \theta_b \sim \text{Normal}(-2, 1), \qquad \pi_l = \frac{\exp(\theta_l)}{1 + \exp(\theta_l)}$$

NB: implies that the denominator is centered around far from 0 but can vary

Models for the denominator of the LATE (cont'd)



Simulation study

We simulated data based on real clinical practice data (from THIN) and considered three levels of uncertainty

1. **Unobserved confounding**
   - We use HDL cholesterol as an unobserved confounder and modify the strength of the relationship with the outcome (LDL cholesterol)
   - We consider 4 levels: $1 =$ low, $\ldots$, $4 =$ high

Simulation study

We simulated data based on real clinical practice data (from THIN) and considered three levels of uncertainty

1. **Unobserved confounding**
   - We use HDL cholesterol as an unobserved confounder and modify the strength of the relationship with the outcome (LDL cholesterol)
   - We consider 4 levels: $1 =$ low, $\ldots$, $4 =$ high

2. **Threshold as instrumental variable**
   - We consider the threshold x_0 as either a strong or weak instrument

Simulation study

We simulated data based on real clinical practice data (from THIN) and considered three levels of uncertainty

1. **Unobserved confounding**
   - We use HDL cholesterol as an unobserved confounder and modify the strength of the relationship with the outcome (LDL cholesterol)
   - We consider 4 levels: $1 =$ low, $\ldots$, $4 =$ high

2. **Threshold as instrumental variable**
   - We consider the threshold x_0 as either a strong or weak instrument

3. **Bandwith around the threshold**
   - The bandwith determines the sample size used for the local regressions
   - Smaller bandwidths imply smaller sample size, although exchangeability on either side of the threshold is more likely to hold
   - Larger bandwidths increase the sample size, but include observations that potentially violate exchangeability

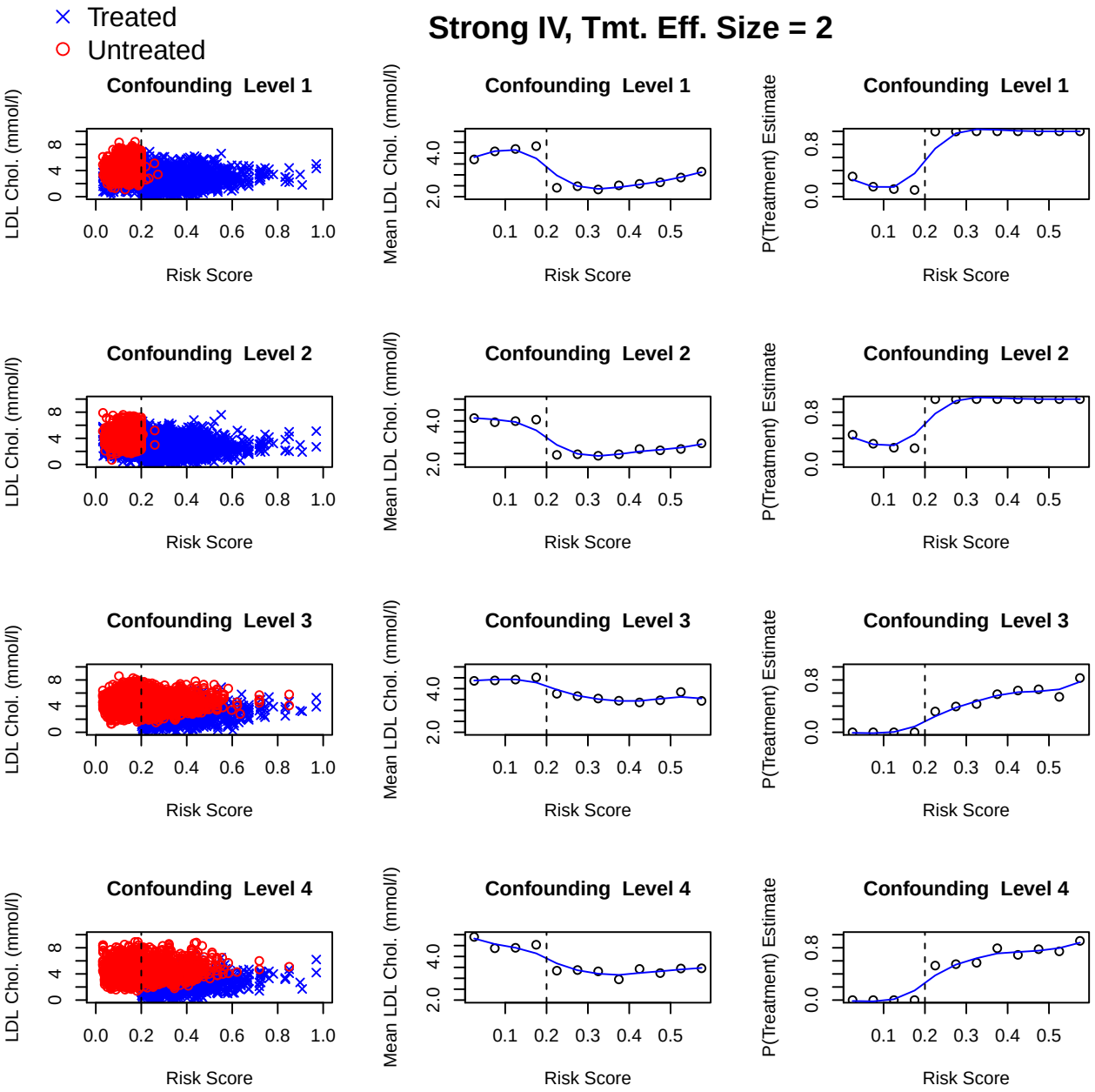
Simulation study

We simulated data based on real clinical practice data (from THIN) and considered three levels of uncertainty

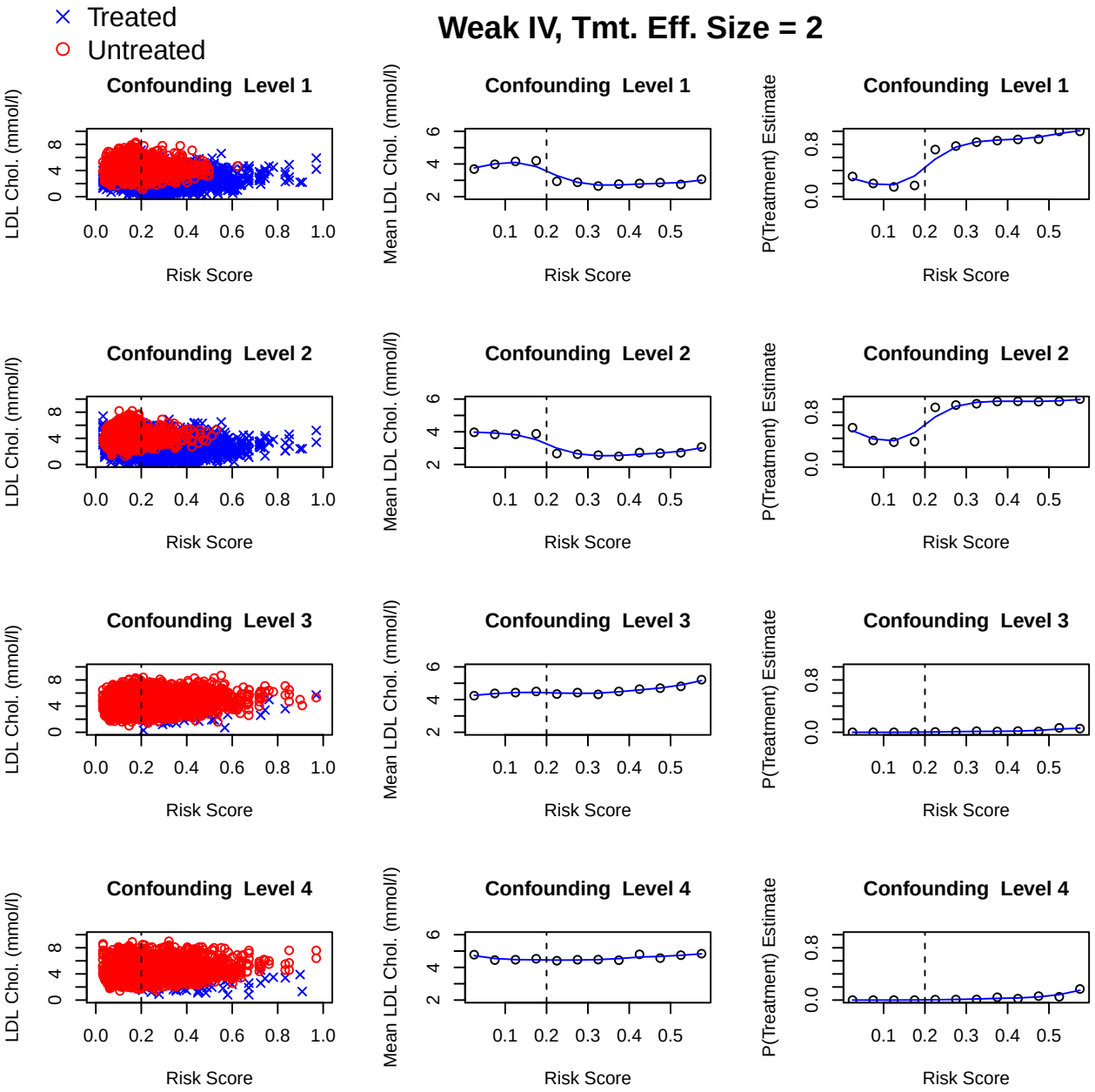
1. **Unobserved confounding**
   - We use HDL cholesterol as an unobserved confounder and modify the strength of the relationship with the outcome (LDL cholesterol)
   - We consider 4 levels: $1 =$ low, ..., $4 =$ high
2. **Threshold as instrumental variable**
   - We consider the threshold x_0 as either a strong or weak instrument
3. **Bandwith around the threshold**
   - The bandwith determines the sample size used for the local regressions
   - Smaller bandwidths imply smaller sample size, although exchangeability on either side of the threshold is more likely to hold
   - Larger bandwidths increase the sample size, but include observations that potentially violate exchangeability

Also, we consider several versions of the ATE and LATE, upon varying the prior distributional assumptions selected

"Descriptive" analysis — useful plots

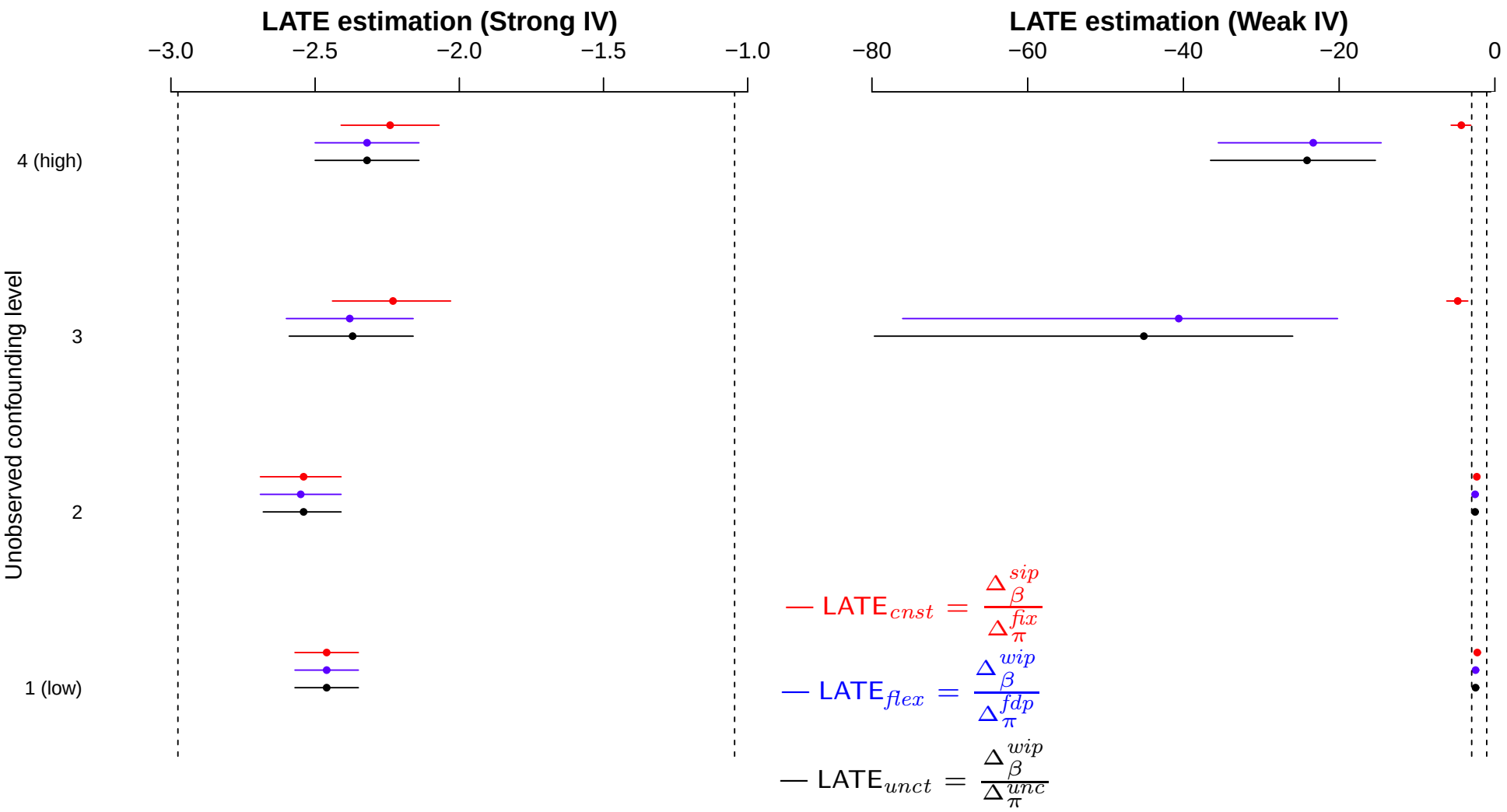


"Descriptive" analysis — useful plots



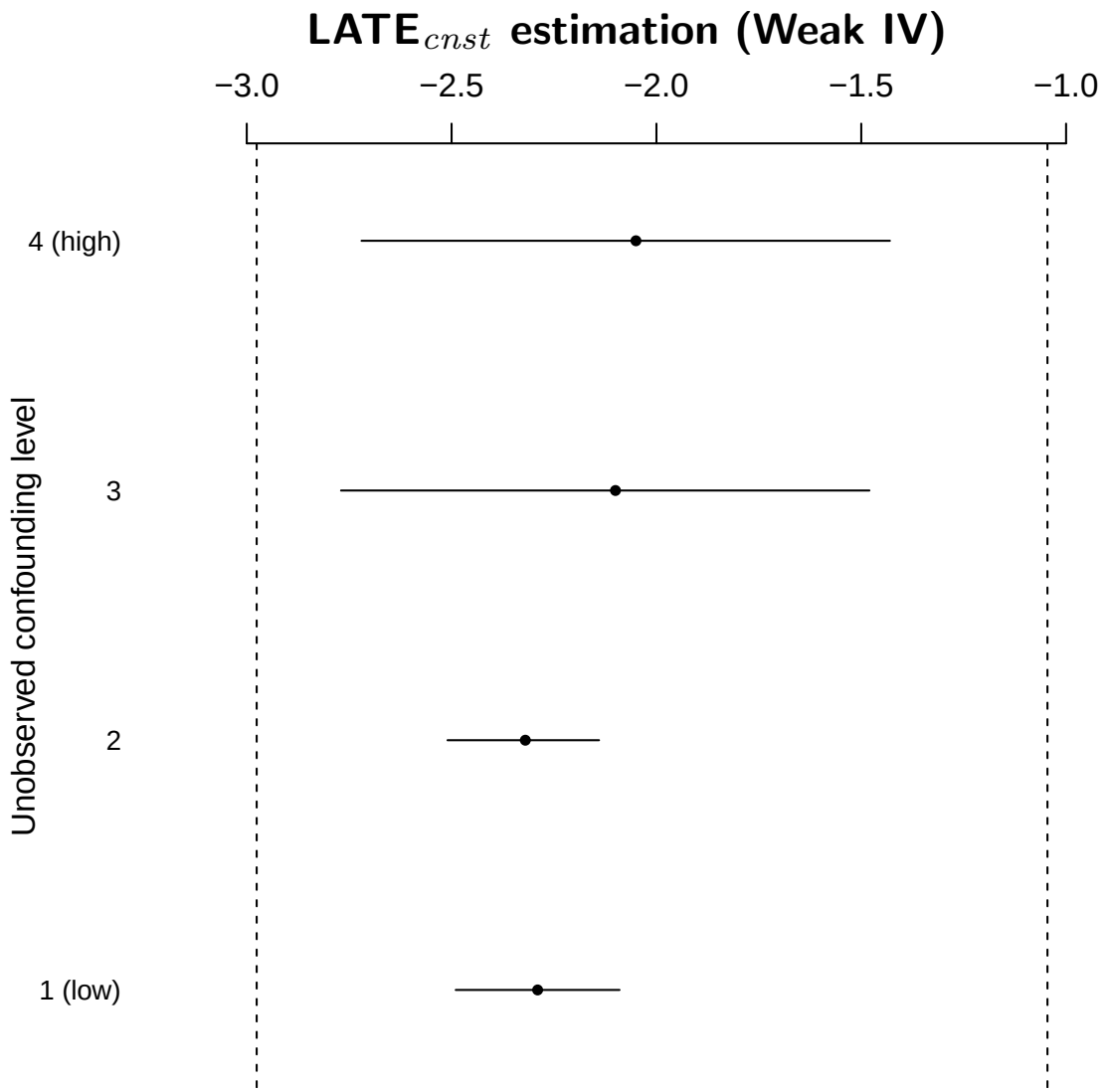
Results

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



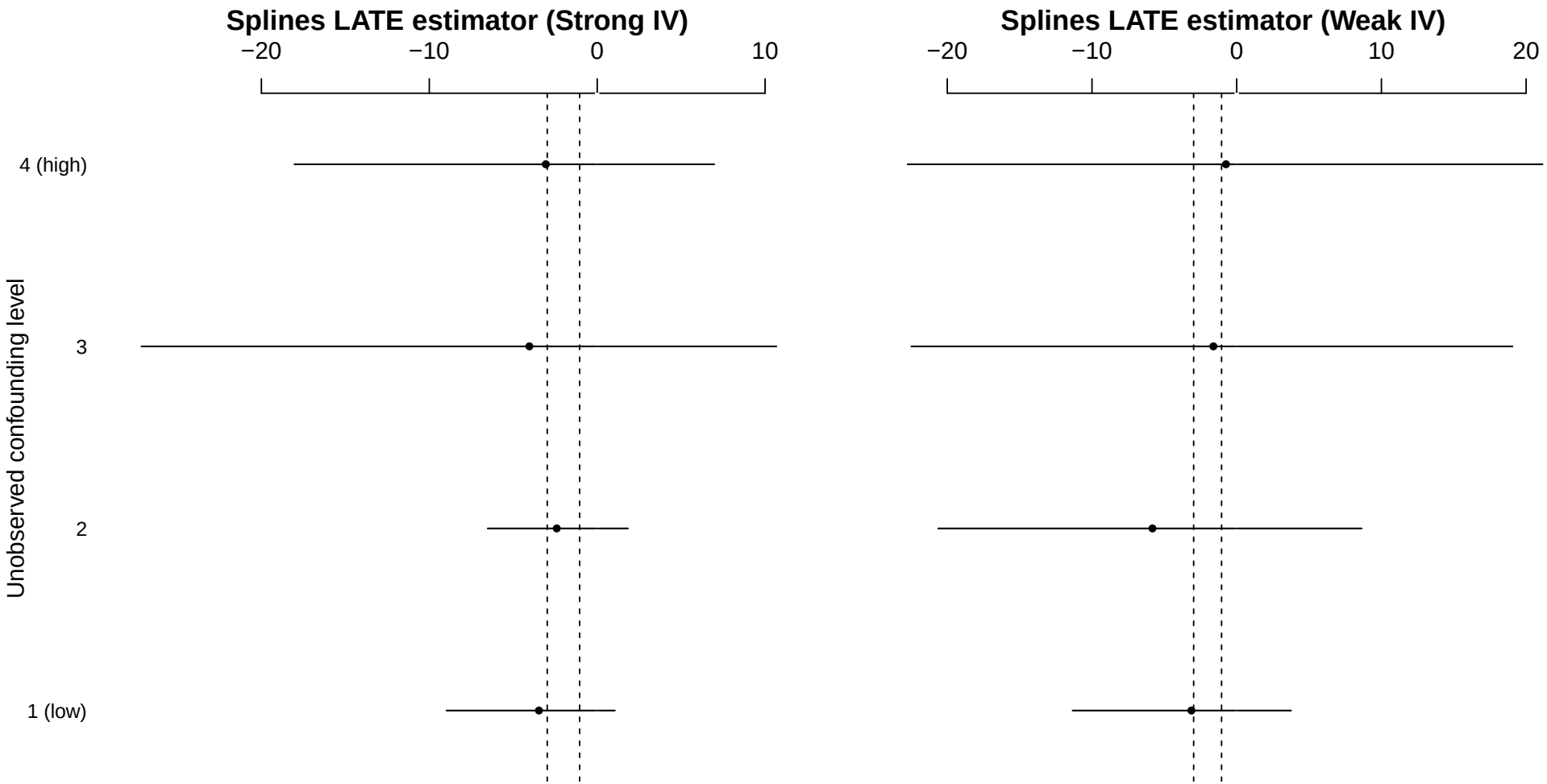
Results (cont'd)

Bandwidth $= 0.05$, Treatment effect size $\sim$ Normal$(-2, 0.5^2)$



Results (cont'd)

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



Conclusions

- The flexible prior does well in recovering the treatment effect when the conditions to apply the RD are valid
  - It is as good as the other estimators for low confounding (even if weak IV) or if strong IV (even if confounding is high)
  - It does not work when the instrument is weak and confounding is high

Conclusions

- The flexible prior does well in recovering the treatment effect when the conditions to apply the RD are valid
  - It is as good as the other estimators for low confounding (even if weak IV) or if strong IV (even if confounding is high)
  - It does not work when the instrument is weak and confounding is high
- The fixed prior can nearly always recover a sensible result
  - But of course this is strongly influenced by the very strict prior
  - It fails to flag scenarios when the RD is not applicable, as the strong prior still induces "reasonable" results

Conclusions

- The flexible prior does well in recovering the treatment effect when the conditions to apply the RD are valid
  - It is as good as the other estimators for low confounding (even if weak IV) or if strong IV (even if confounding is high)
  - It does not work when the instrument is weak and confounding is high
- The fixed prior can nearly always recover a sensible result
  - But of course this is strongly influenced by the very strict prior
  - It fails to flag scenarios when the RD is not applicable, as the strong prior still induces "reasonable" results
- Care is needed in applying "flexible" models (*eg* splines)
  - They can be too flexible and can adapt too well to the idiosyncrasies of the data, resulting in very variable estimations

References

Dawid, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance (with Discussion). In P. Green, N. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 45–81. Oxford University Press.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615 – 635. The regression discontinuity design: Theory and applications.

NICE (2008). *Quick reference guide: Statins for the prevention of cardiovascular events*. NICE.

Thistlethwaite, D. and D. Campbell (1960). Regression-Discontinuity Analysis - An alternative to the ex-post-facto experiment. *Journal of Educational Psychology 51*(6), 309–317.

van der Klaauw, G. (2008). Regression-discontinuity analysis: A survey of recent developments in economics. *Labour 22*(2), 219–245.

Ward, S., L. Jones, A. Pandor, M. Holmes, R. Ara, A. Ryan, W. Yeo, and N. Payne (2007). A systematic review and economic evaluation of statins for the prevention of coronary events. *Health Technology Assessment 11*(14).

Thank you!

Results

Bandwidth = 0.05, Treatment Effect Size ∼ Normal$(-2, 0.5^2)$

		ATE Estimators		LATE Estimators			
IV	*Confounding*	Δ_β^{freq}	Δ_β^{sip}	LATE_{unct}	LATE_{flex}	LATE_{cnst}	*Spline*
Strong	1	-1.74	-2.15	-2.42	-2.42	-2.41	
		(-1.98, -1.50)	(-2.25, -2.05)	(-2.54, -2.29)	(-2.55, -2.29)	(-2.54, -2.29)	
	2	-1.43	-1.86	-2.51	-2.52	-2.51	
		(-1.67, -1.19)	(-1.97, -1.76)	(-2.68, -2.35)	(-2.68, -2.35)	(-2.69, -2.35)	
	3	-0.72	-1.20	-2.88	-2.88	-2.32	
		(-1.08, -0.36)	(-1.32, -1.09)	(-3.24, -2.54)	(-3.25, -2.54)	(-2.58, -2.08)	
	4	-0.92	-1.45	-2.61	-2.61	-2.29	
		(-1.26, -0.58)	(-1.57, -1.33)	(-2.87, -2.36)	(-2.87, -2.36)	(-2.51, -2.08)	
Weak	1	-1.01	-1.44	-2.73	-2.72	-2.29	
		(-1.31, -0.72)	(-1.55, -1.33)	(-3.01, -2.46)	(-3.00, -2.45)	(-2.49, -2.09)	
	2	-1.08	-1.53	-2.67	-2.67	-2.32	
		(-1.33, -0.84)	(-1.63, -1.42)	(-2.93, -2.43)	(-2.92, -2.43)	(-2.51, -2.14)	
	3	0.05	-0.36	33.87	-66.28	-2.10	
		(-0.16, 0.25)	(-0.46, -0.26)	(-944.90, 657.63)	(-293.68, 58.97)	(-2.77, -1.48)	
	4	0.08	-0.36	-104.83	-53.01	-2.05	
		(-0.13, 0.29)	(-0.46, -0.26)	(-426.90, 243.85)	(-154.97, 4.56)	(-2.72, -1.43)	

Bandwidth = 0.25, Treatment Effect Size ∼ Normal$(-2, 0.5^2)$

		ATE Estimators		LATE Estimators			
IV	*Confounding*	Δ_β^{freq}	Δ_β^{sip}	LATE_{unct}	LATE_{flex}	LATE_{cnst}	*Spline*
Strong	1	-2.02	-2.17	-2.46	-2.46	-2.46	-3.47
		(-2.17, -1.86)	(-2.26, -2.08)	(-2.57, -2.35)	(-2.57, -2.35)	(-2.57, -2.35)	(-8.97, 1.05)
	2	-1.68	-1.87	-2.54	-2.55	-2.54	-2.41
		(-1.86, -1.51)	(-1.96, -1.78)	(-2.68, -2.41)	(-2.69, -2.41)	(-2.69, -2.41)	(-6.52, 1.83)
	3	-0.95	-1.20	-2.37	-2.38	-2.23	-4.04
		(-1.27, -0.62)	(-1.30, -1.10)	(-2.59, -2.16)	(-2.60, -2.16)	(-2.44, -2.03)	(-27.14, 10.67)
	4	-1.19	-1.46	-2.32	-2.32	-2.24	-3.06
		(-1.48, -0.90)	(-1.57, -1.36)	(-2.50, -2.14)	(-2.50, -2.14)	(-2.41, -2.07)	(-18.03, 6.97)
Weak	1	-1.25	-1.46	-2.48	-2.48	-2.26	-3.13
		(-1.46, -1.03)	(-1.56, -1.36)	(-2.67, -2.29)	(-2.67, -2.29)	(-2.42, -2.09)	(-11.34, 3.76)
	2	-1.31	-1.53	-2.54	-2.54	-2.32	-5.82
		(-1.45, -1.16)	(-1.63, -1.44)	(-2.73, -2.36)	(-2.72, -2.36)	(-2.47, -2.16)	(-20.63, 8.63)
	3	-0.20	-0.35	-45.08	-40.60	-4.77	-1.61
		(-0.31, -0.08)	(-0.44, -0.26)	(-79.69, -25.96)	(-76.06, -20.20)	(-6.18, -3.47)	(-22.48, 19.06)
	4	-0.15	-0.35	-24.12	-23.33	-4.31	-0.74
		(-0.27, -0.03)	(-0.44, -0.25)	(-36.52, -15.34)	(-35.53, -14.60)	(-5.62, -3.09)	(-22.74, 21.13)