Bayesian Regression Discontinuity Design With unknown cutoff

Julia Kowalska, Amsterdam Causality Meeting 23.11.2023

Amsterdam UMC
Universitair Medische Centra

The team

Stéphanie van der Pas

Mark van de Wiel

Regression Discontinuity Design

Why?

- First time applied in the 60's to assess effect of receiving certificate of merit on students future academic career
- Problem: Intervention assigned based on cutoff criterion → no overlap between groups with scores below and above the cutoff

Regression Discontinuity Design

Why?

- First time applied in the 60’s to assess effect of receiving certificate of merit on students future academic career
- Problem: Intervention assigned based on cutoff criterion → no overlap between groups with scores below and above the cutoff
- Solution: Overlap very close to the cutoff → local causal effect

RDD - basic model

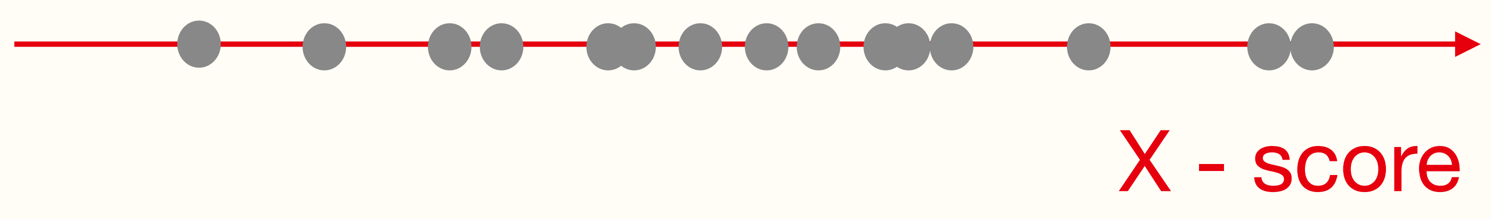
Framework for causal inference for observational data

X - score variable (age, blood pressure, exam score)

X - score

RDD - basic model

Framework for causal inference for observational data

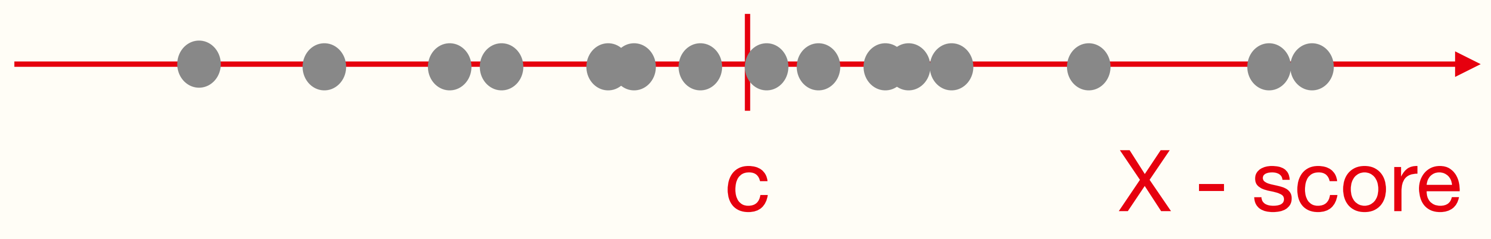


X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

RDD - basic model

Framework for causal inference for observational data



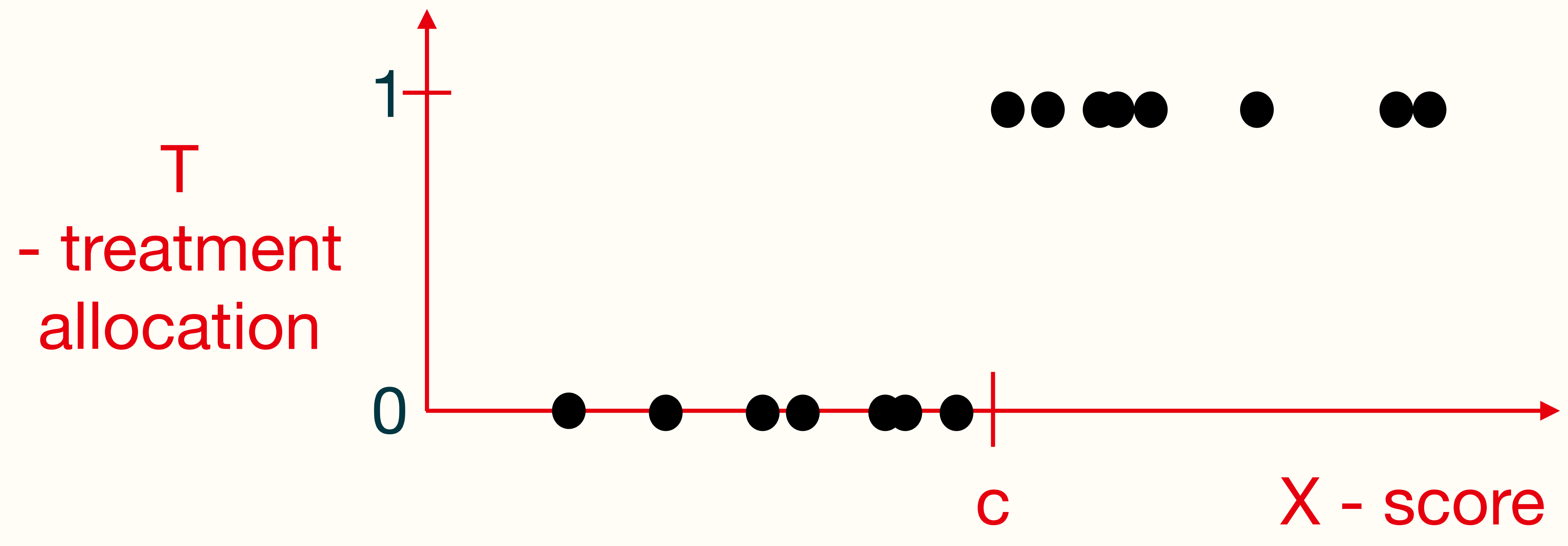
X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

c - cutoff point

RDD - basic model

Framework for causal inference for observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

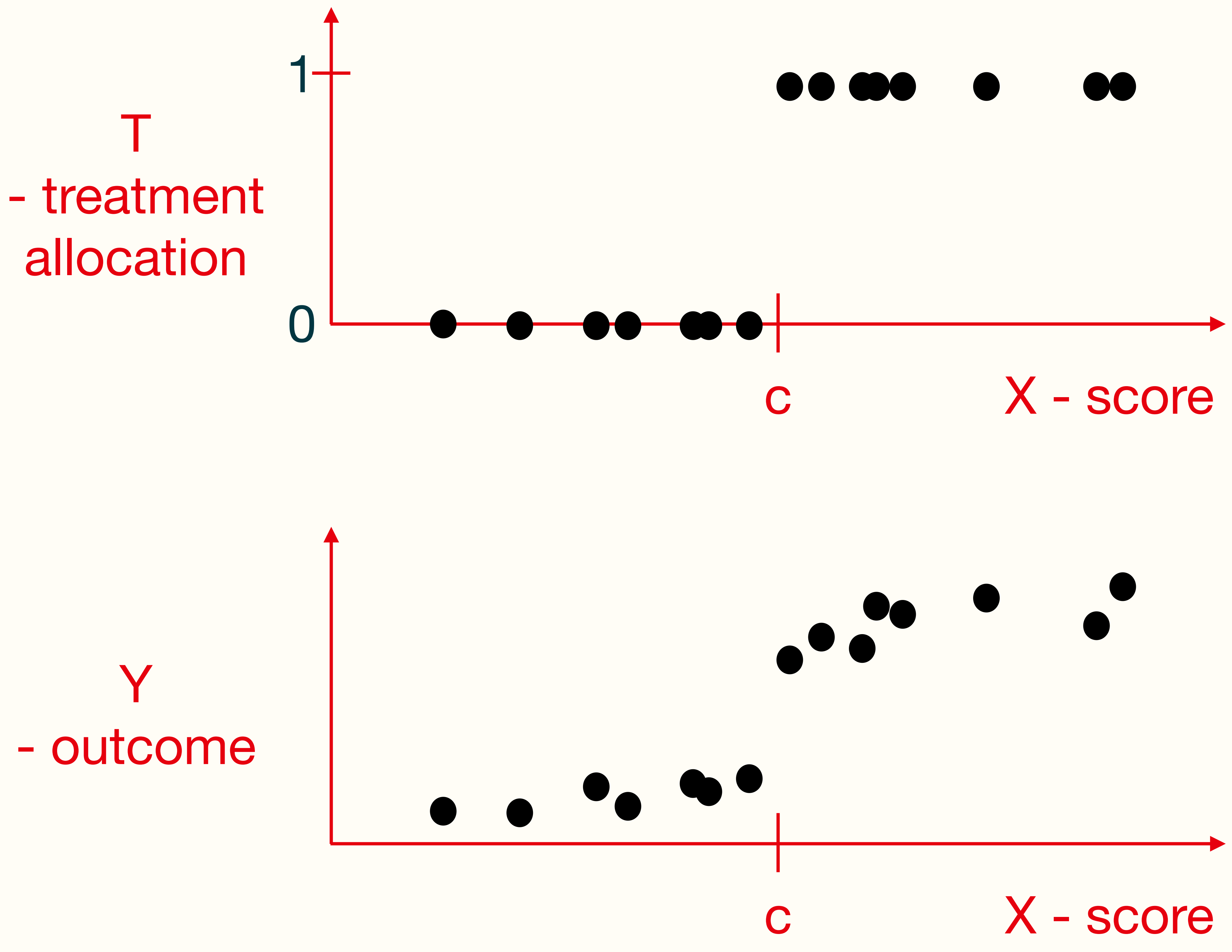
c - cutoff point

Units with scores above c get treatment, units with scores below c don’t get treatment

T - treatment allocation

RDD - basic model

Framework for causal inference for observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

c - cutoff point

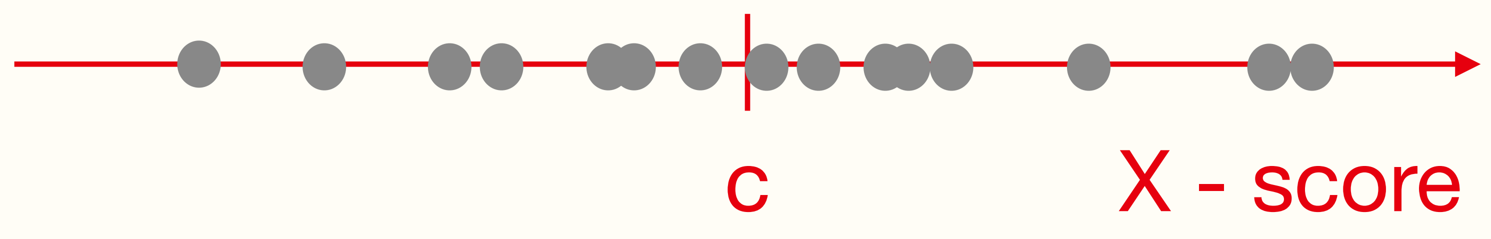
Units with scores above c get treatment, units with scores below c don't get treatment

T - treatment allocation

Y - outcome variable

RDD - fuzzy model

Framework for causal inference for observational data



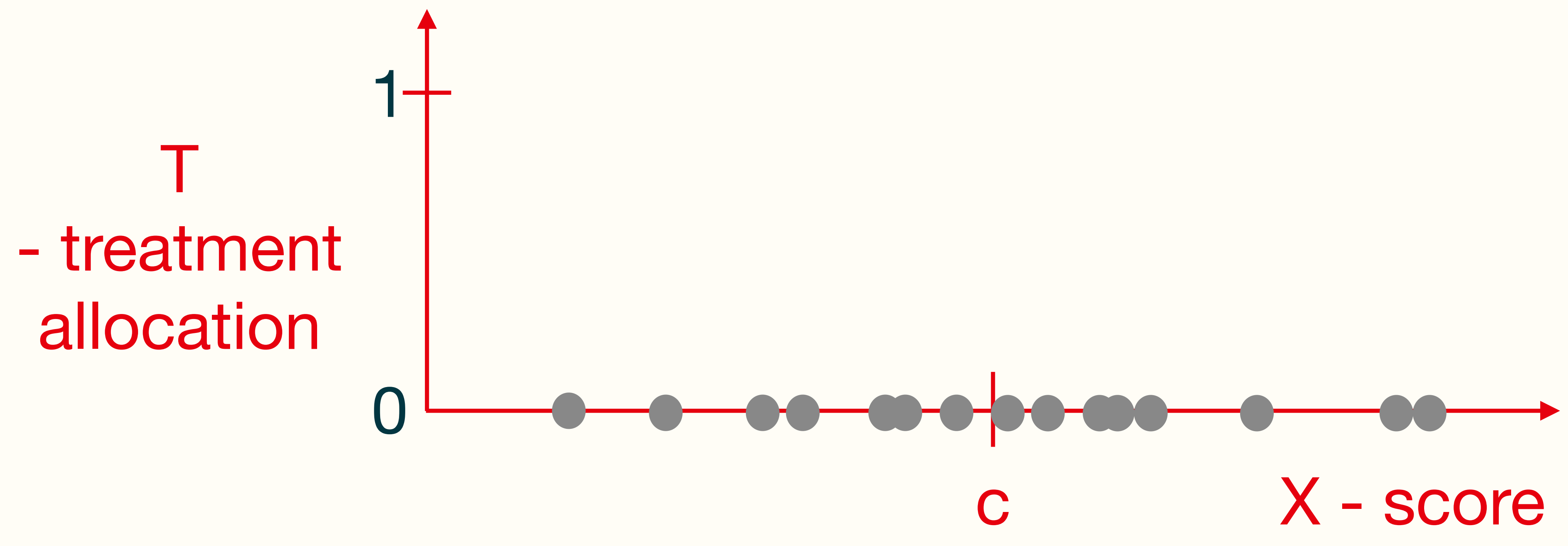
X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

c - cutoff point

RDD - fuzzy model

Framework for causal inference for observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

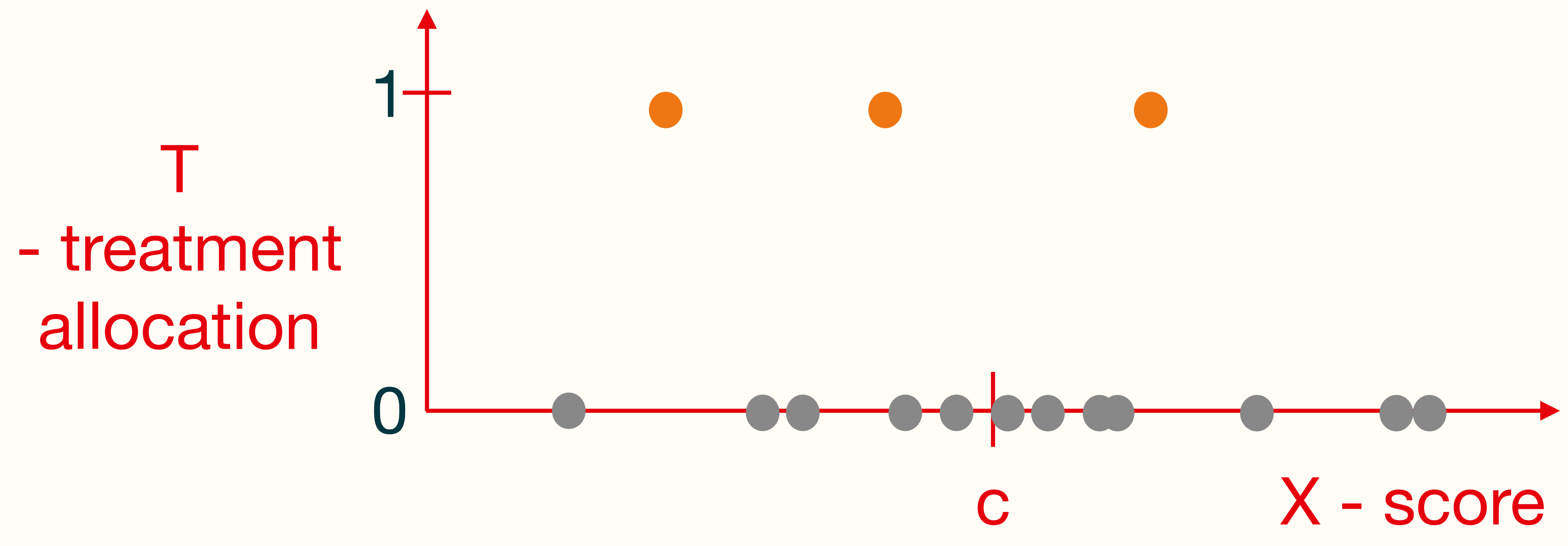
c - cutoff point

Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

RDD - fuzzy model

Framework for causal inference for observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

c - cutoff point

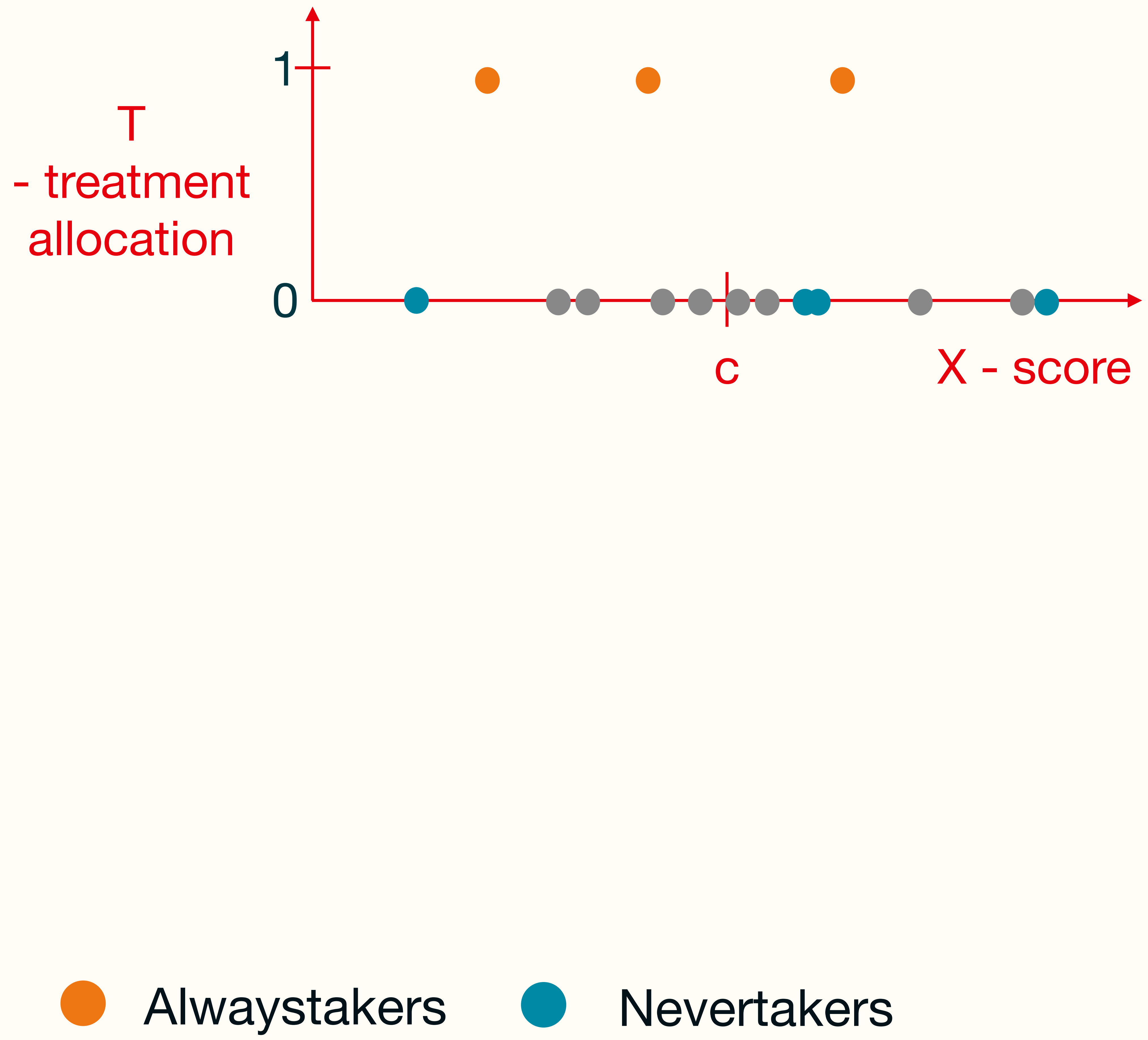
Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

Alwaystakers

RDD - fuzzy model

Framework for causal inference from observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

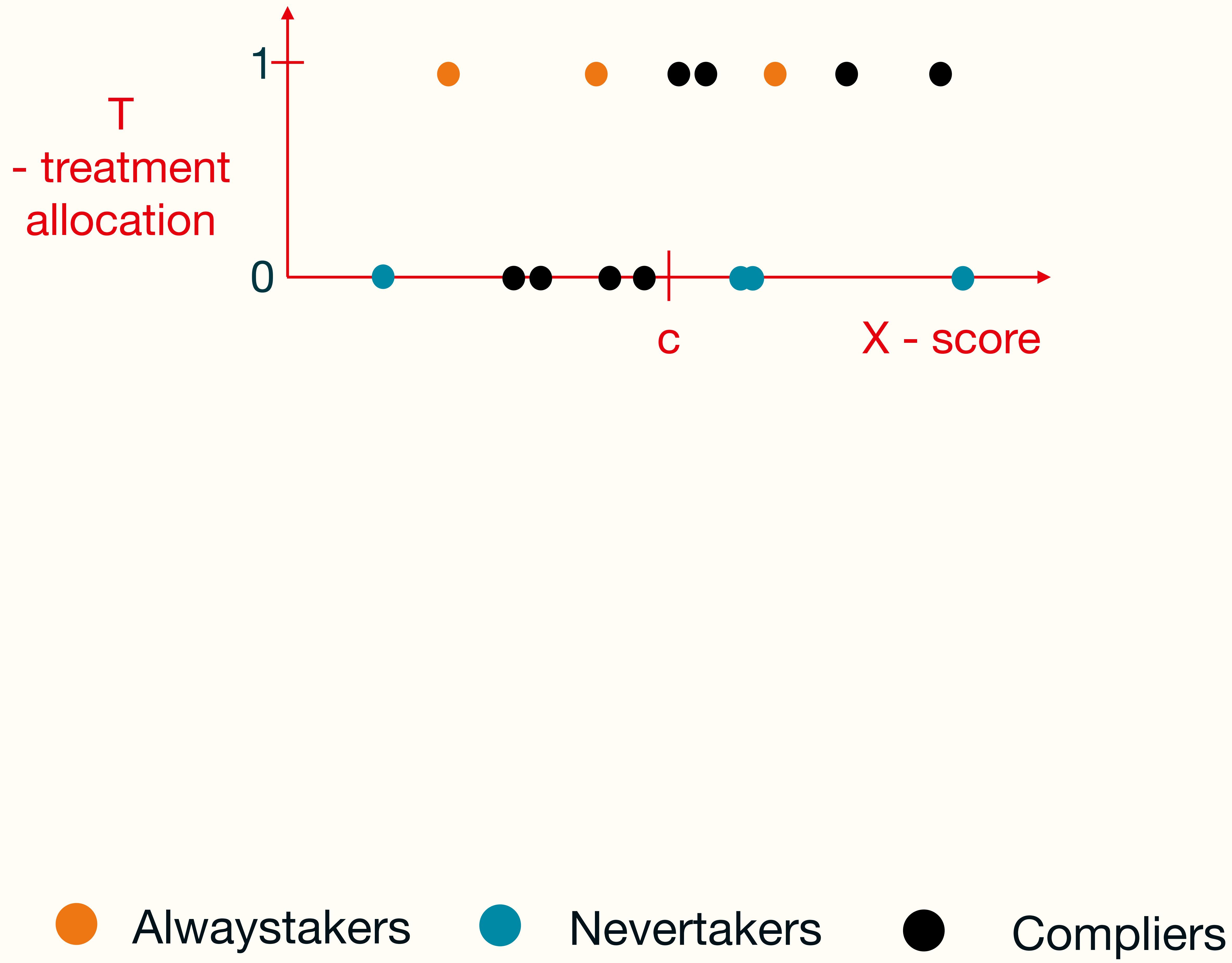
c - cutoff point

Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

RDD - fuzzy model

Framework for causal inference from observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

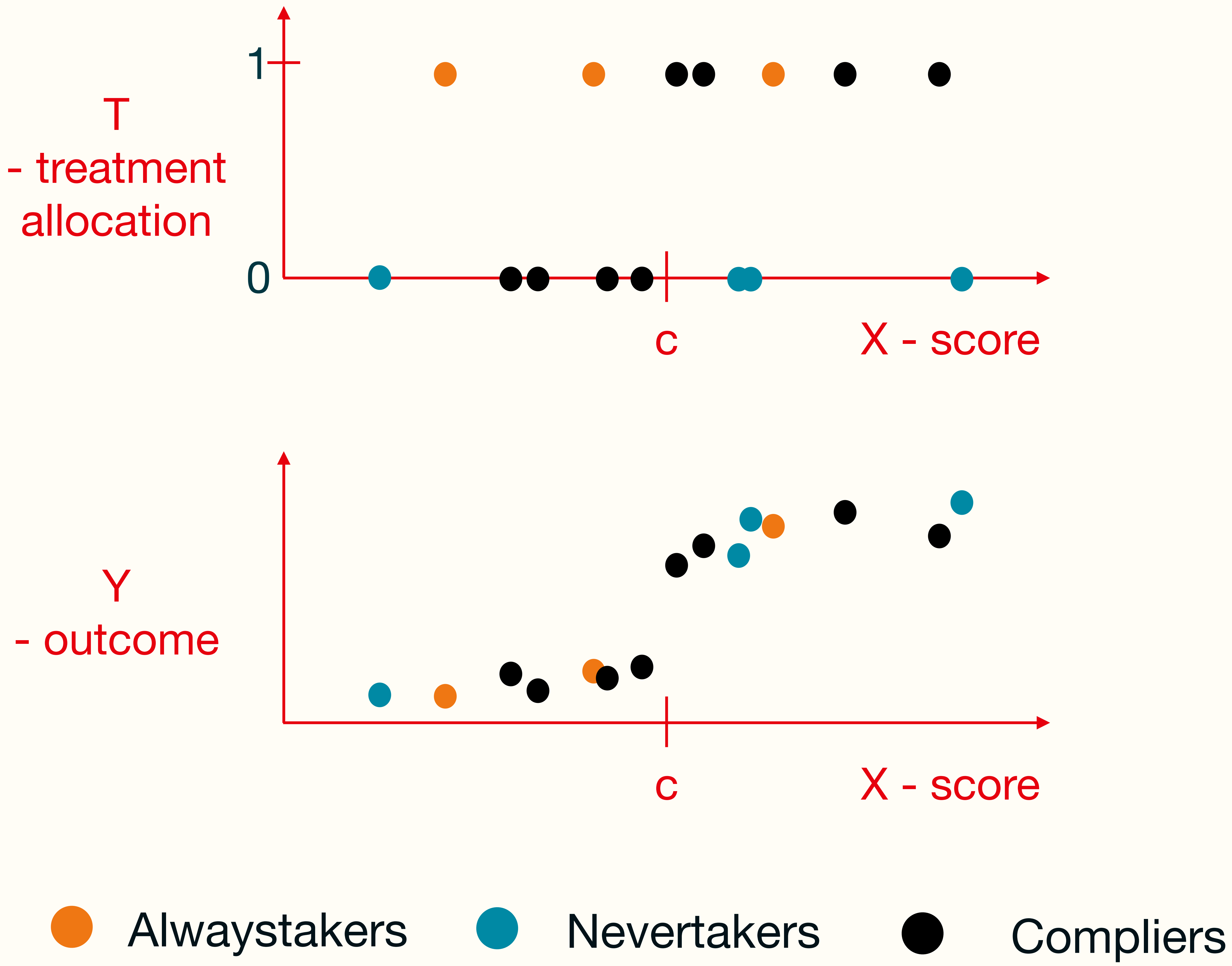
c - cutoff point

Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

RDD - fuzzy model

Framework for causal inference from observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

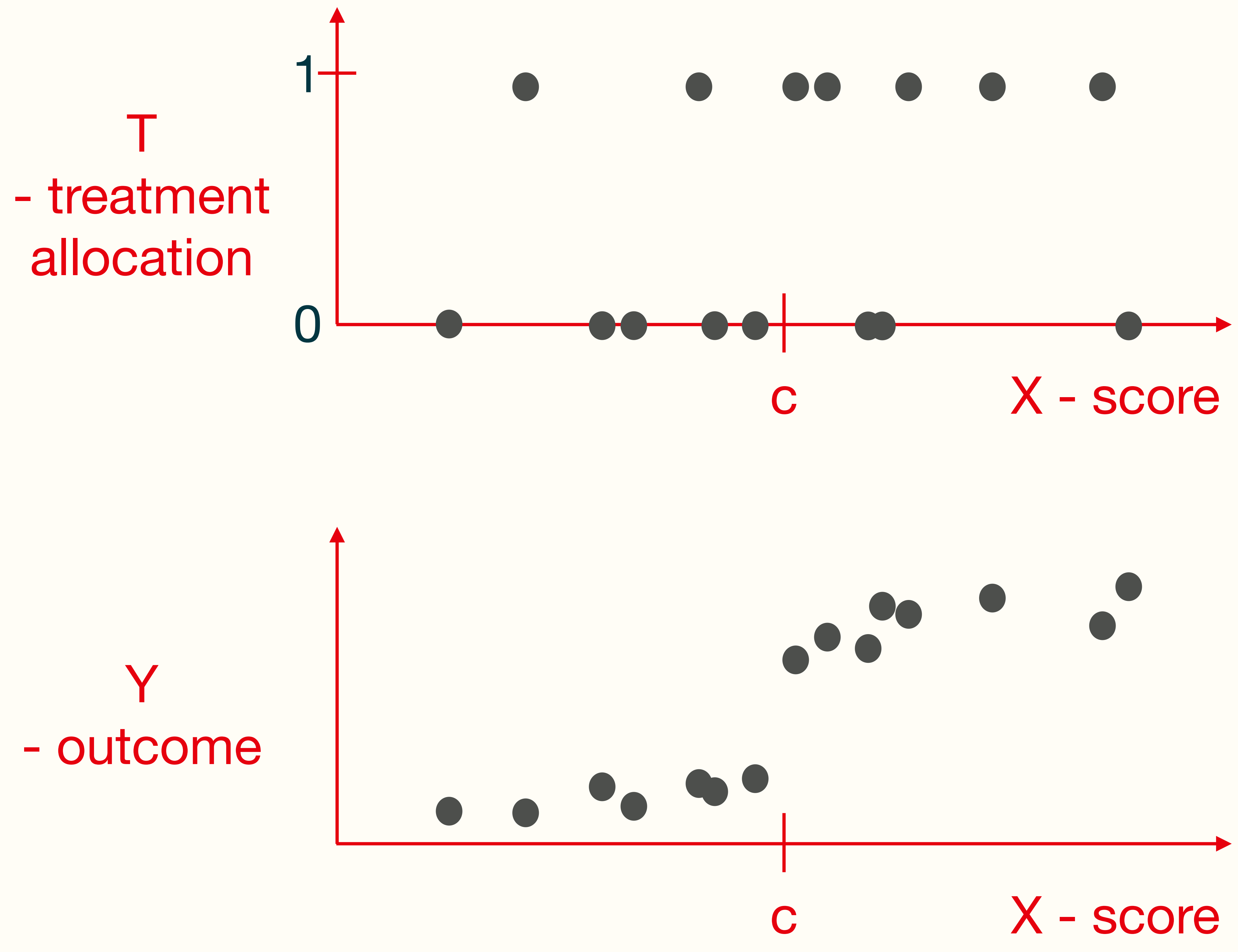
c - cutoff point

Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

RDD - fuzzy model

Framework for causal inference from observational data



X - score variable (age, blood pressure, exam score)

Score is assigned to each unit (patients, student)

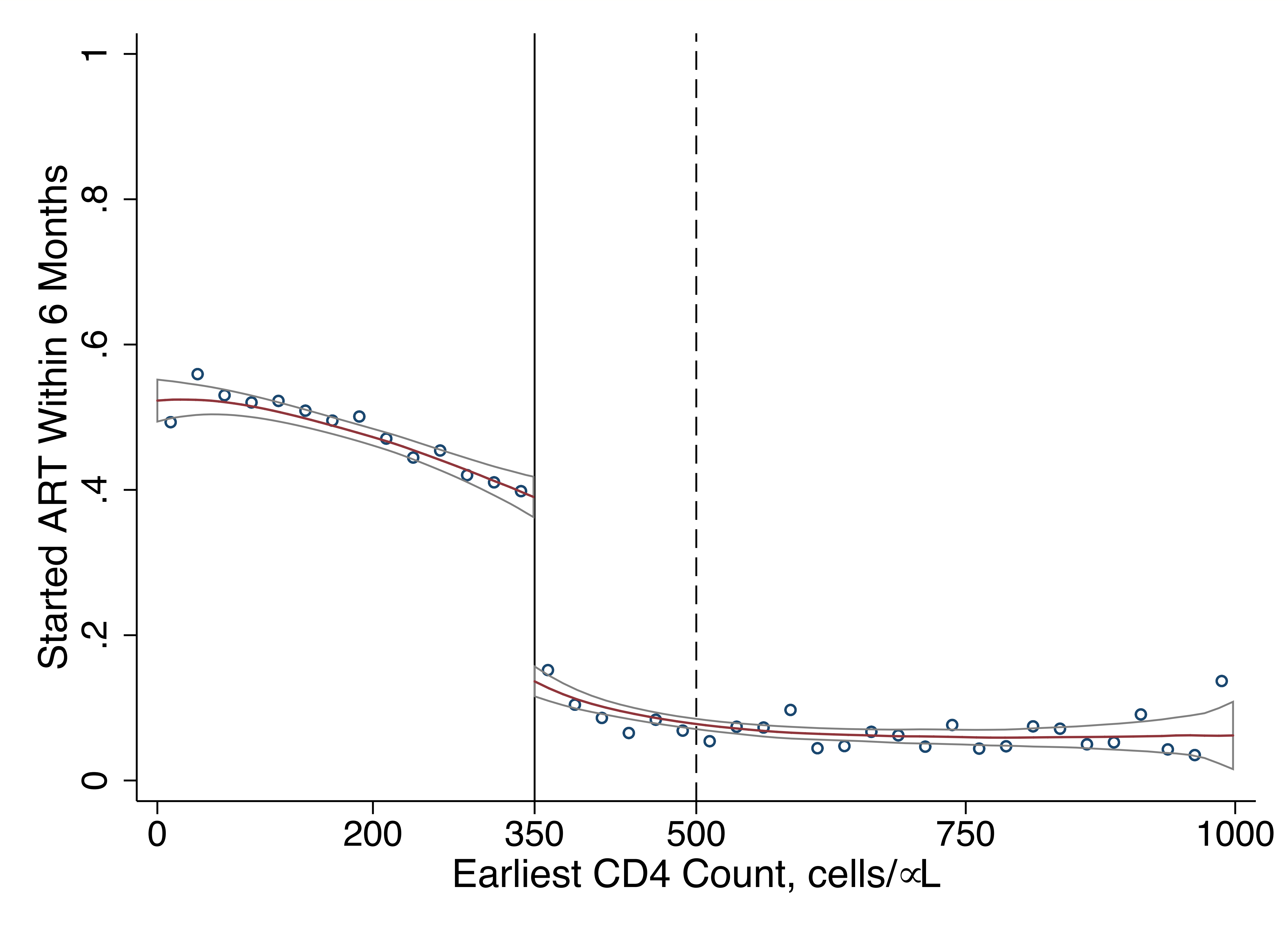
c - cutoff point

Units with score above c have significantly higher probability of getting treatment than units with scores below c

T - treatment allocation

RDD - fuzzy model, example

Hlabisa HIV Treatment and Care Programme



Impact on immediate antiretroviral therapy (ART) on retention in care.

X - CD4 count

c - 350 cells/ml

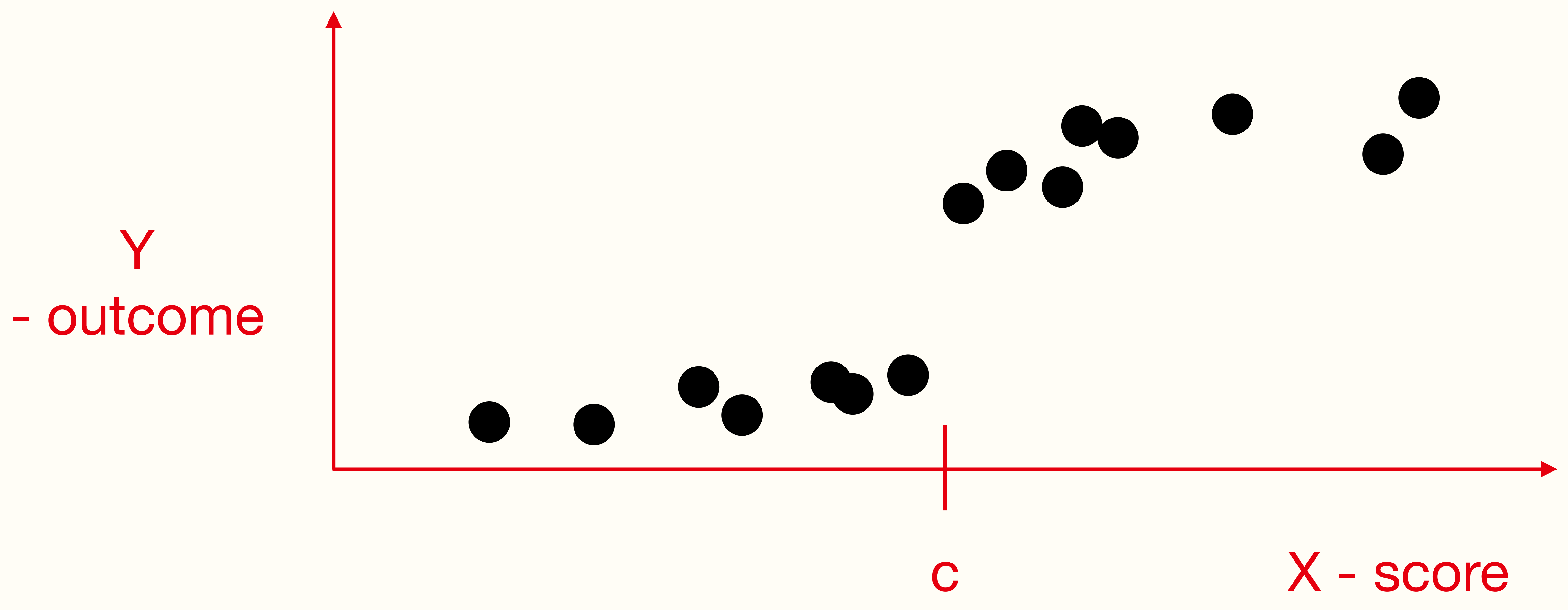
T - ART initiation

Y - binary outcome: 1 if evidence for retention in care

Source:Bor J,et.al. Effect of eliminating CD4-count thresholds on HIV treatment initiation in South Africa: An empirical modeling study. PLoS One. 2017 Jun 15;12(6):e0178249. doi: 10.1371/journal.pone.0178249. PMID: 28617805; PMCID: PMC5472329.

RDD - basic model

Causal treatment effect estimand



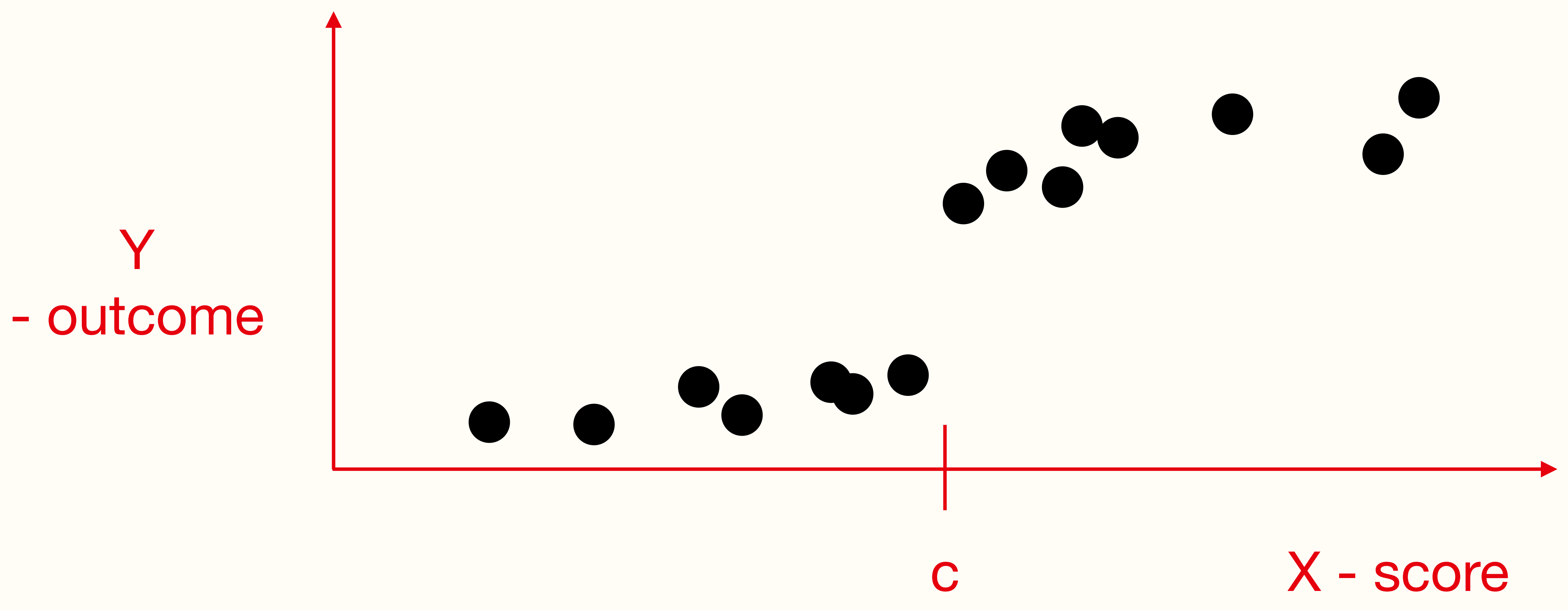
Idea behind RDD:

Units close to the cutoff are similar

- ‘local randomised experiment’

RDD - basic model

Causal treatment effect estimand



Idea behind RDD:

Units close to the cutoff are similar

- 'local randomised experiment'

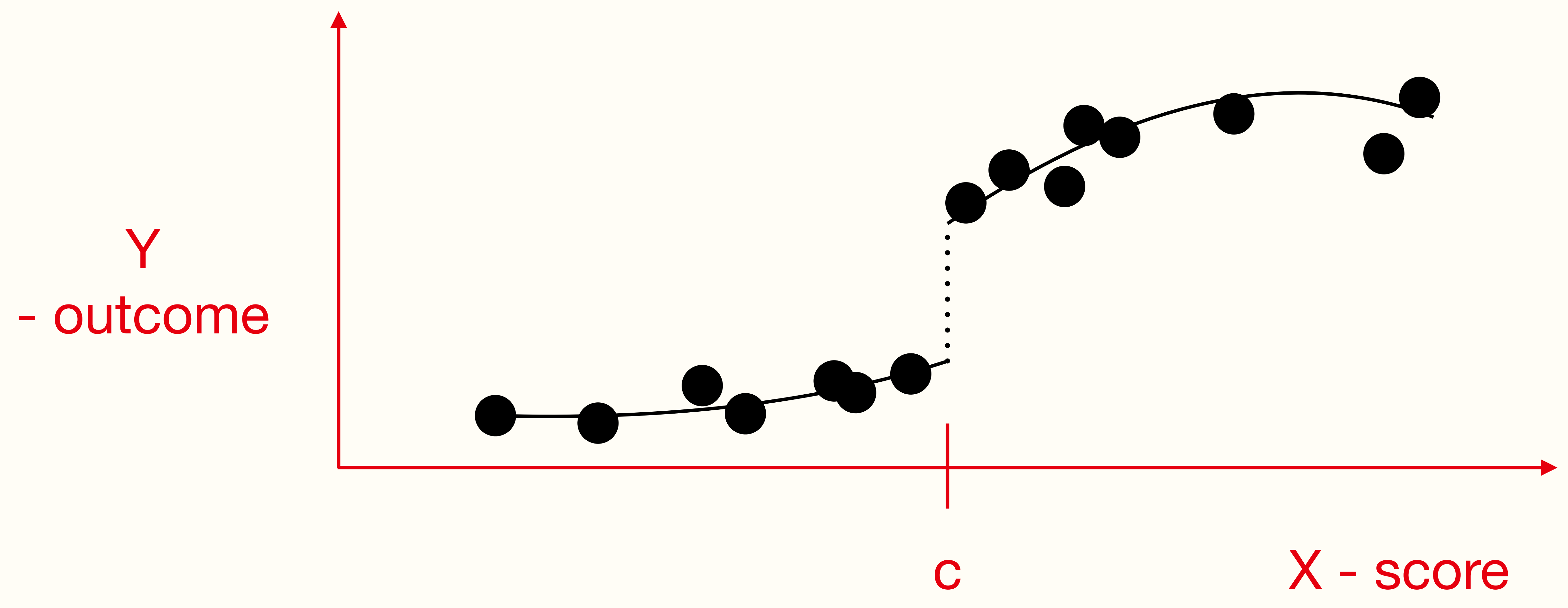
$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c\right]$$

$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

Treatment effect at the cutoff is identifiable (under some assumptions)

RDD - basic model

Causal treatment effect estimand



Idea behind RDD:

Units close to the cutoff are similar

- ‘local randomised experiment’

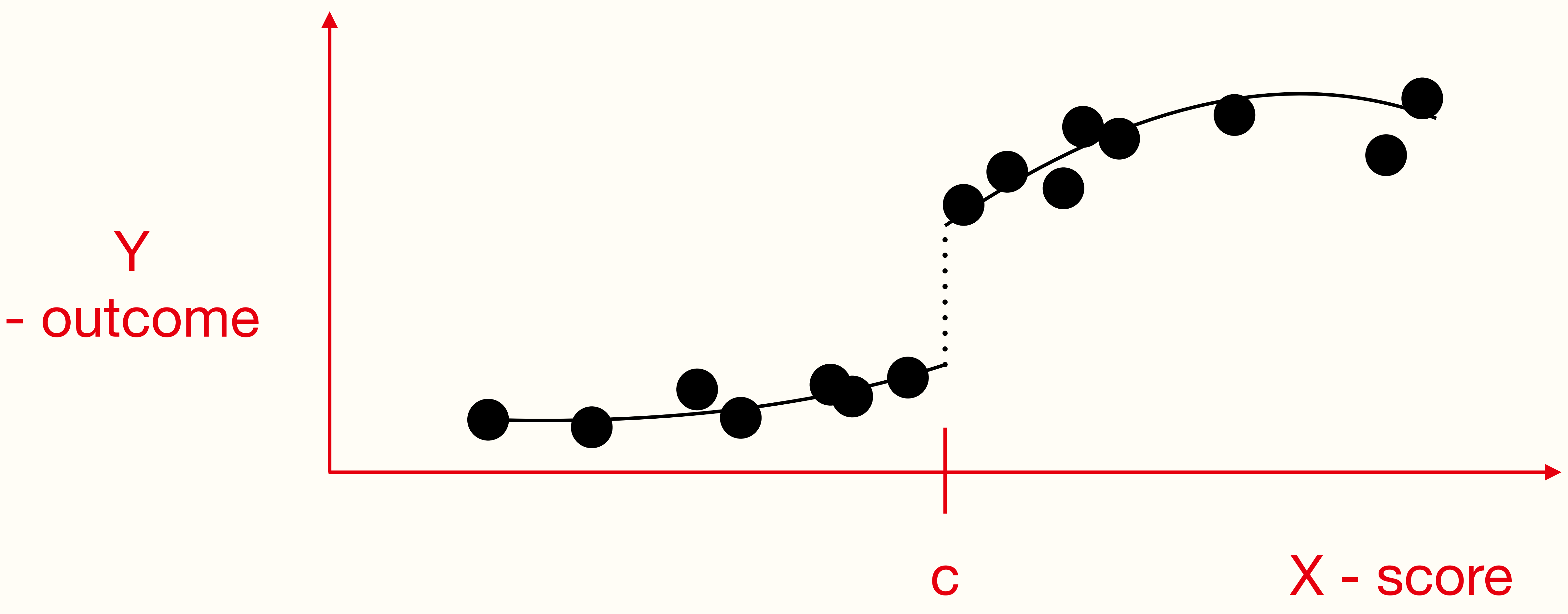
$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c\right]$$

$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

Treatment effect at the cutoff is identifiable (under some assumptions)

RDD - basic model

Causal treatment effect estimand



Idea behind RDD:

Units close to the cutoff are similar

- ‘local randomised experiment’

$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c\right]$$

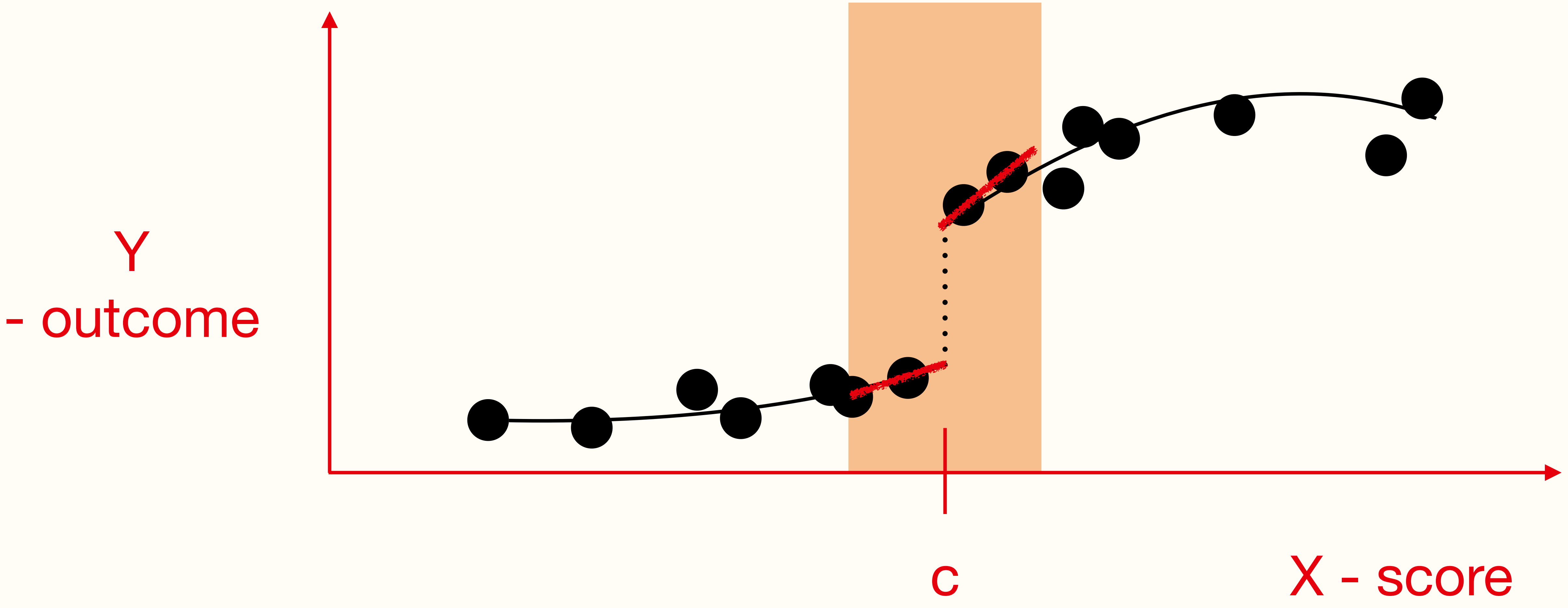
$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

$$\tau = \lim_{x \downarrow c} \mathbb{E}[Y \,|\, X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \,|\, X = x]$$

Treatment effect at the cutoff is identifiable (under some assumptions)

RDD - basic model

Causal treatment effect estimand



Idea behind RDD:

Units close to the cutoff are similar

- 'local randomised experiment'

$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c\right]$$

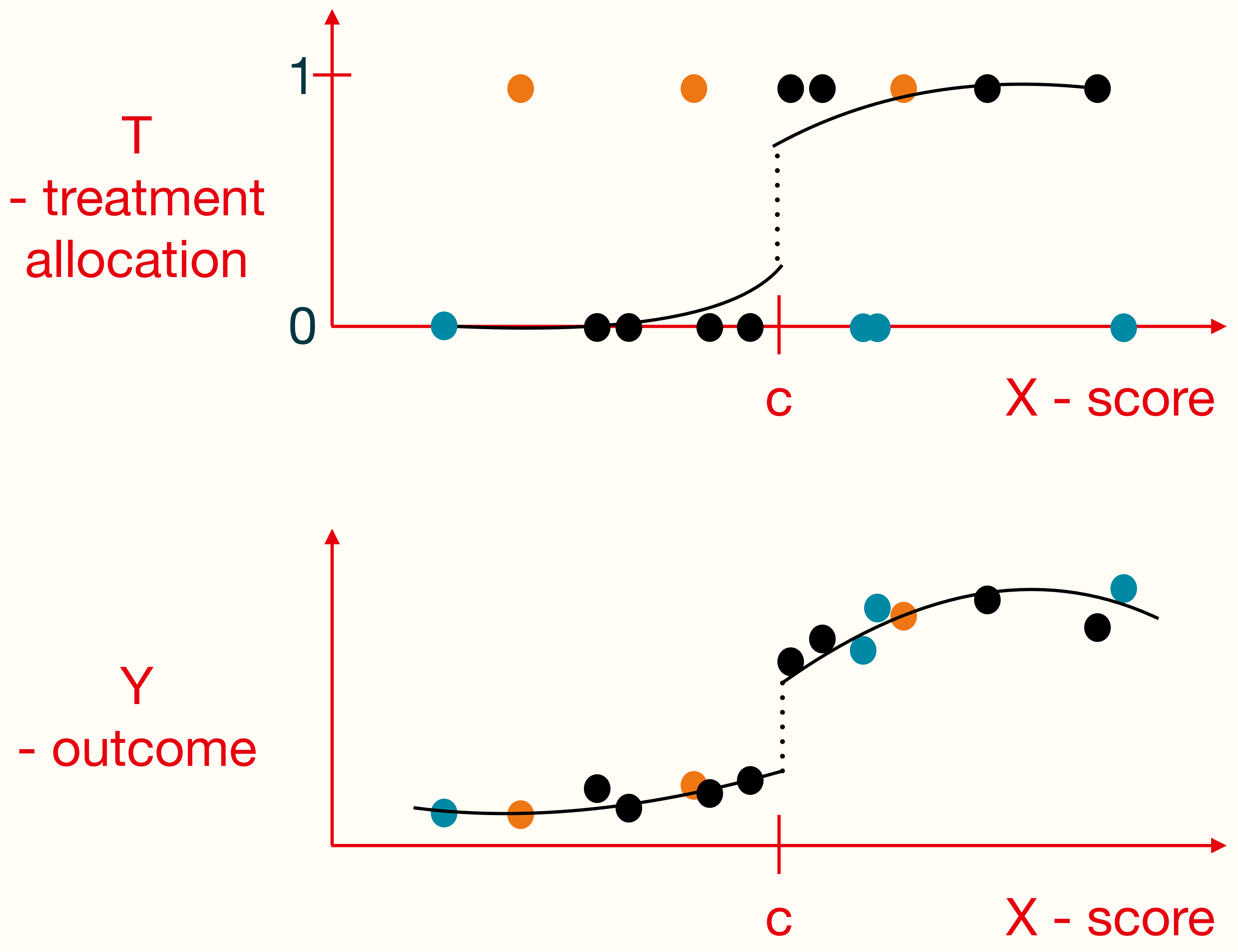
$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

$$\tau = \lim_{x \downarrow c} \mathbb{E}[Y \,|\, X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \,|\, X = x]$$

Treatment effect at the cutoff is identifiable (under some assumptions)

RDD - fuzzy model

Causal treatment effect estimand



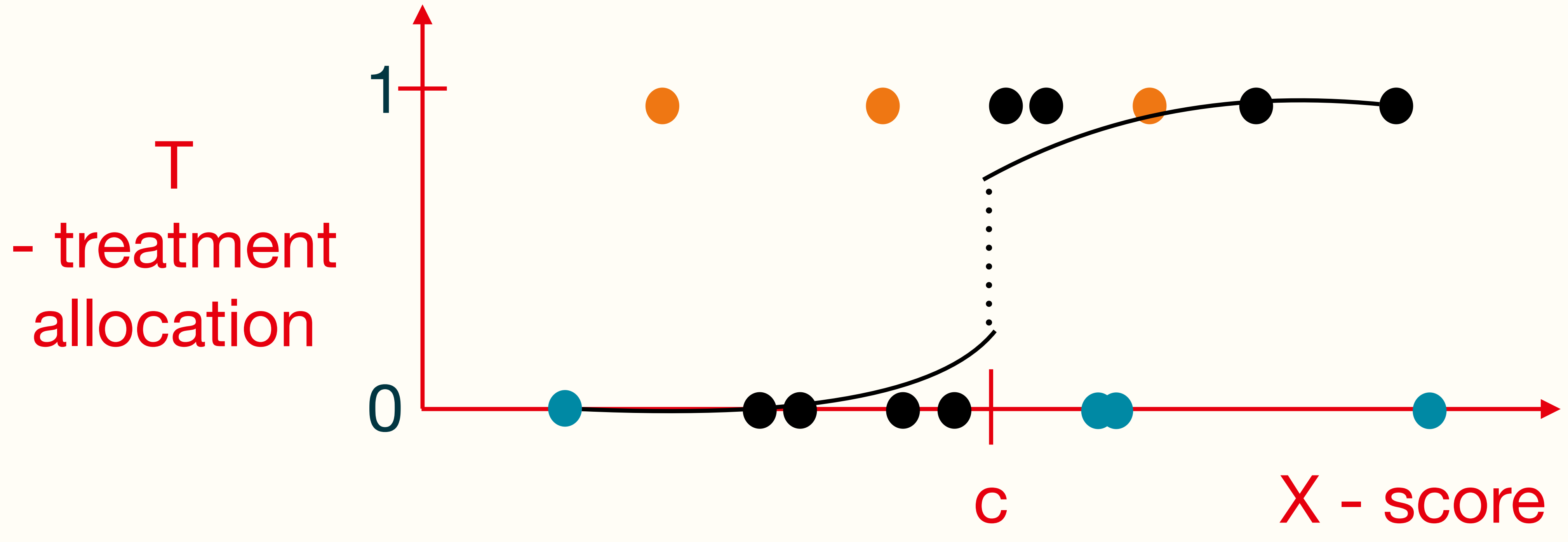
We can only estimate treatment effect for compliers at the cutoff:

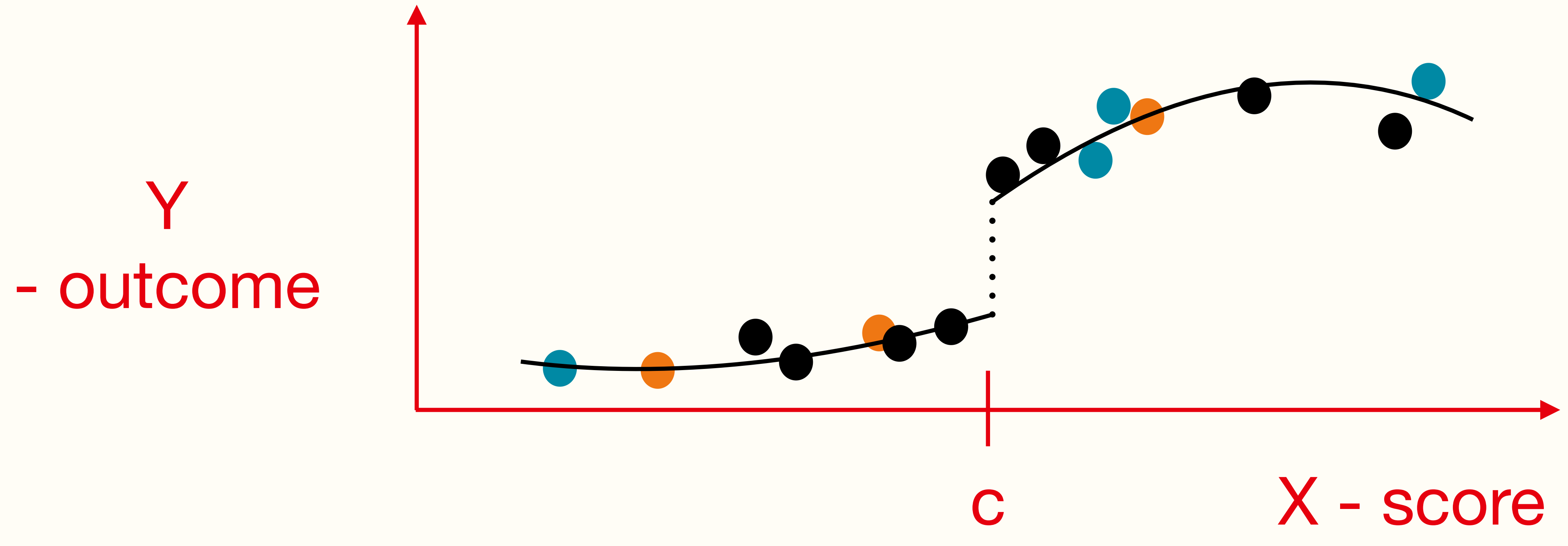
$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c, ct = C\right]$$

$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

RDD - fuzzy model

Causal treatment effect estimand





We can only estimate treatment effect for compliers at the cutoff:

$$\mathbb{E}\left[Y^{(1)} - Y^{(0)} \,|\, X = c, ct = C\right]$$

$Y^{(1)}, Y^{(0)}$ - potential outcome under treatment/ no treatment

$$\tau = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \,|\, X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \,|\, X = x]}{\lim_{x \downarrow c} \mathbb{P}[T = 1 \,|\, X = x] - \lim_{x \uparrow c} \mathbb{P}[T = 1 \,|\, X = x]}$$

$$\tau = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \,|\, X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \,|\, X = x]}{\boxed{\lim_{x \downarrow c} \mathbb{P}[T = 1 \,|\, X = x] - \lim_{x \uparrow c} \mathbb{P}[T = 1 \,|\, X = x]}}$$

We don’t want the denominator to be small:

- Philosophical problem: there is almost no compliance so it puts the whole design into question
- Mathematical problem: analysis is unstable

$$\tau = \frac{\lim_{x \downarrow c} \mathbb{E}[Y \mid X = x] - \lim_{x \uparrow c} \mathbb{E}[Y \mid X = x]}{\boxed{\lim_{x \downarrow c} \mathbb{P}[T = 1 \mid X = x] - \lim_{x \uparrow c} \mathbb{P}[T = 1 \mid X = x]}}$$

We don't want the denominator to be small:

- Philosophical problem: there is almost no compliance so it puts the whole design into question
- Mathematical problem: analysis is unstable

Cutoff

experiment design perspective	**mathematical perspective**
Guideline imposed through a policy	Point with a significant jump in the treatment probability

Common assumption: they are the same

- Cutoff is unknown if and only if the guideline is unknown
- Compliance assessed at the cutoff

Common assumption: they are the same

- Cutoff is unknown if and only if the guideline is unknown
- Compliance assessed at the cutoff

Common assumption: they are the same

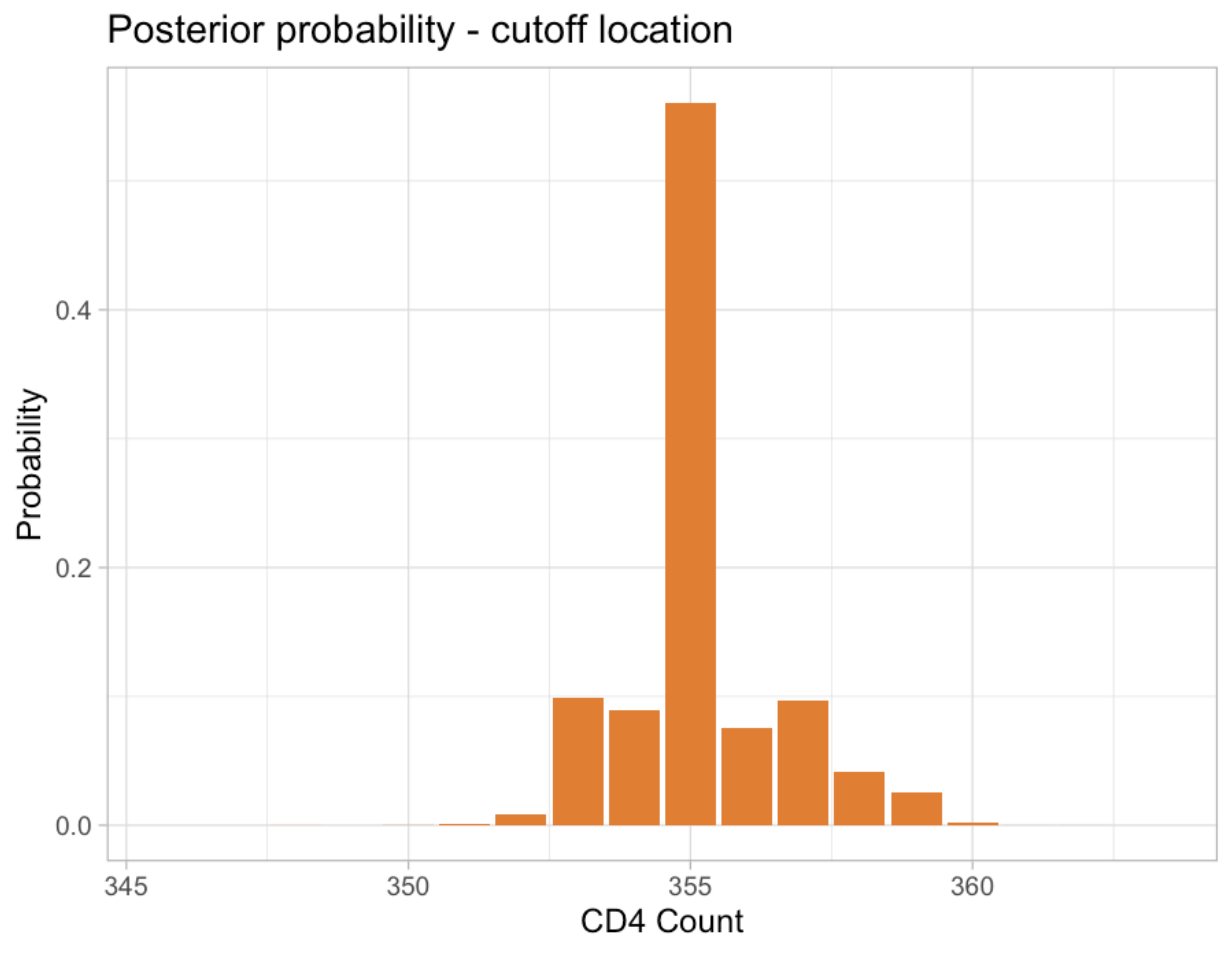
- Cutoff is unknown if and only if the guideline is unknown
- Compliance assessed at the cutoff

RDD - fuzzy model, example revisited

c = 350 cells/ml is not the cutoff!

RDD - fuzzy model, example revisited

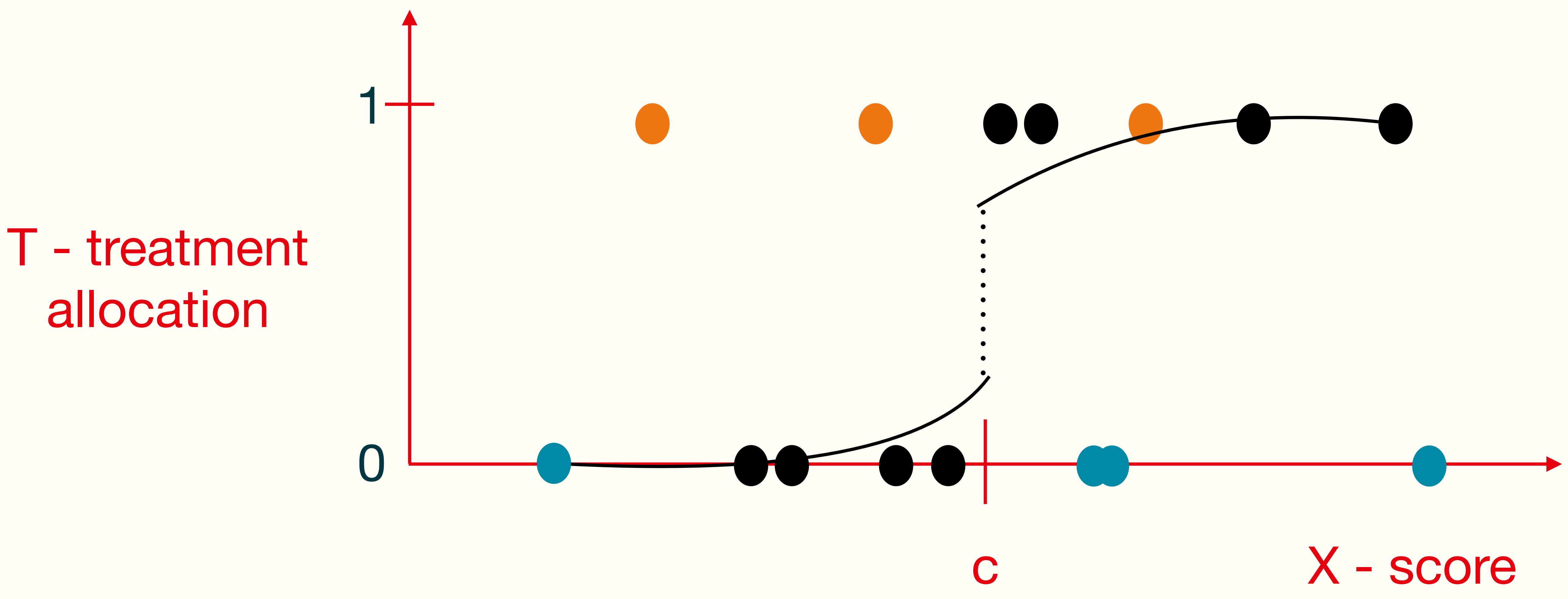
c = 350 cells/ml is not the cutoff!



- Misleading clustering:
  - Points on the two sides of the cutoff clustered together
- Misleading calculations:
  - Continuity assumptions violated
- Probable cutoff at c=355 close to the guideline

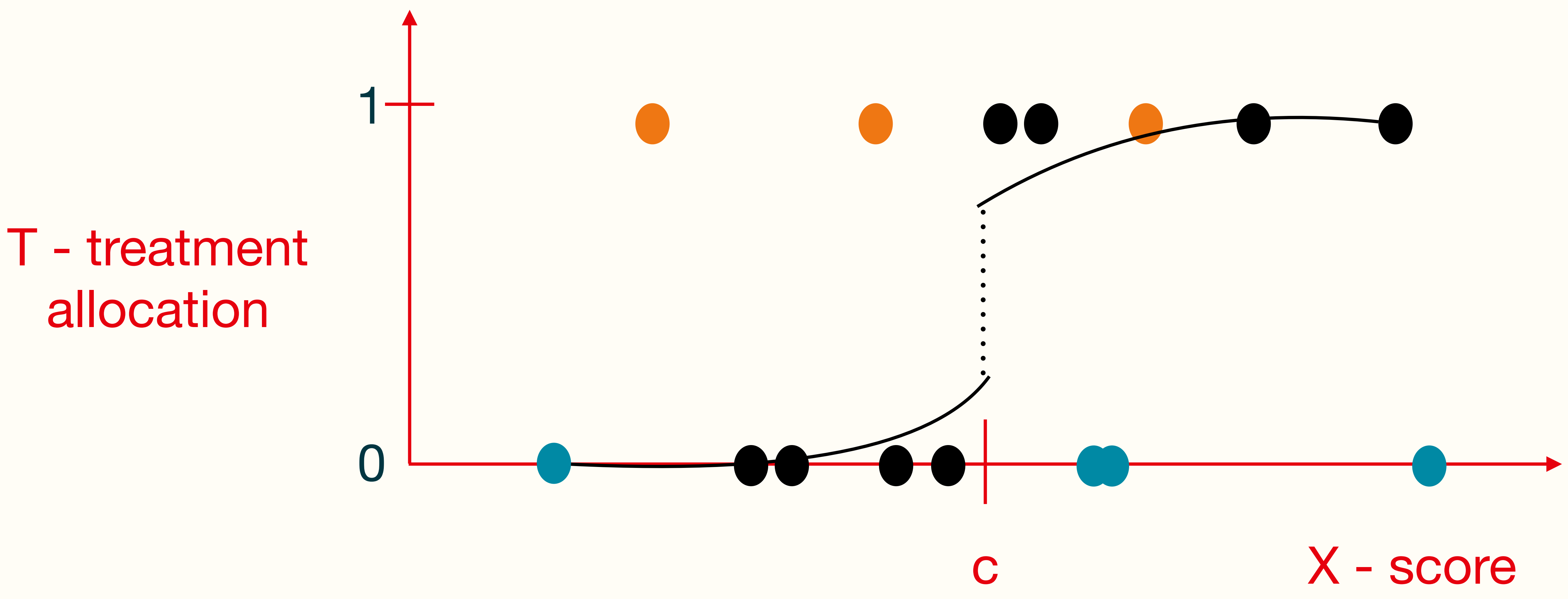
Treatment probability

Bayesian approach



Treatment probability

Bayesian approach

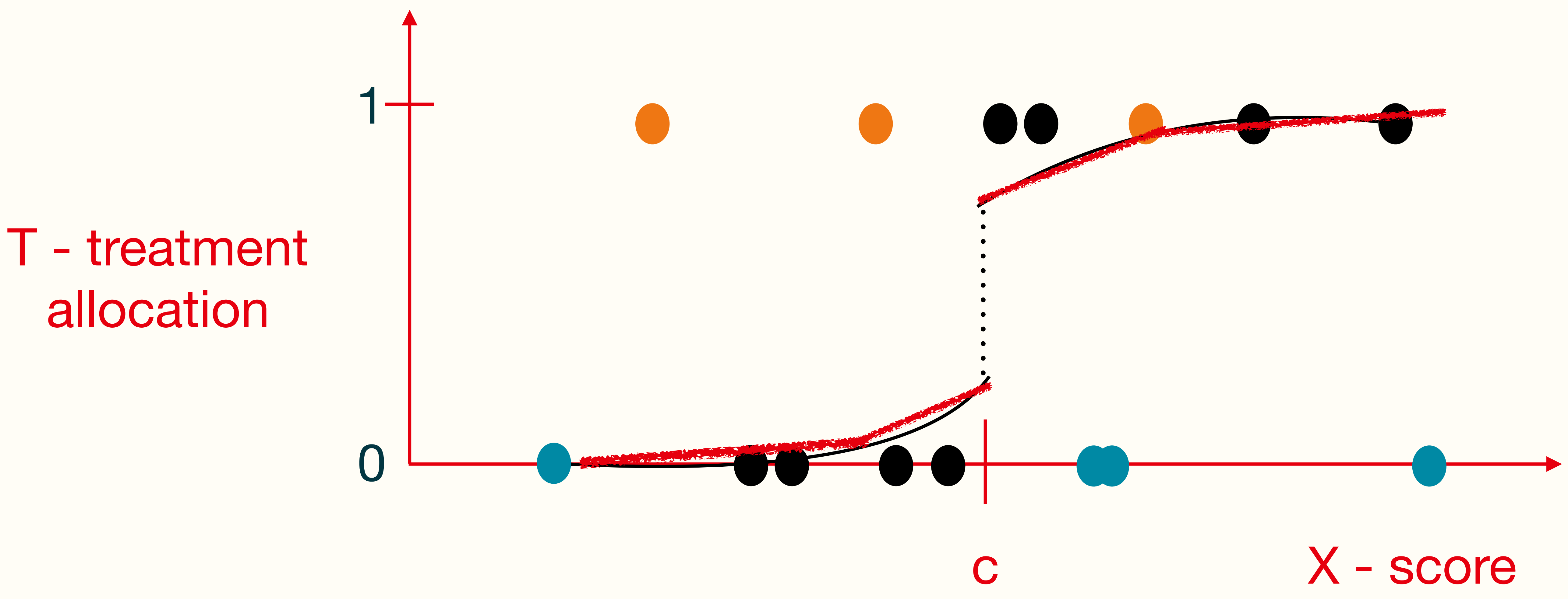


Assumptions:

- Function is increasing
- Convex/concave on each side of the cutoff

Treatment probability

Bayesian approach



Assumptions:

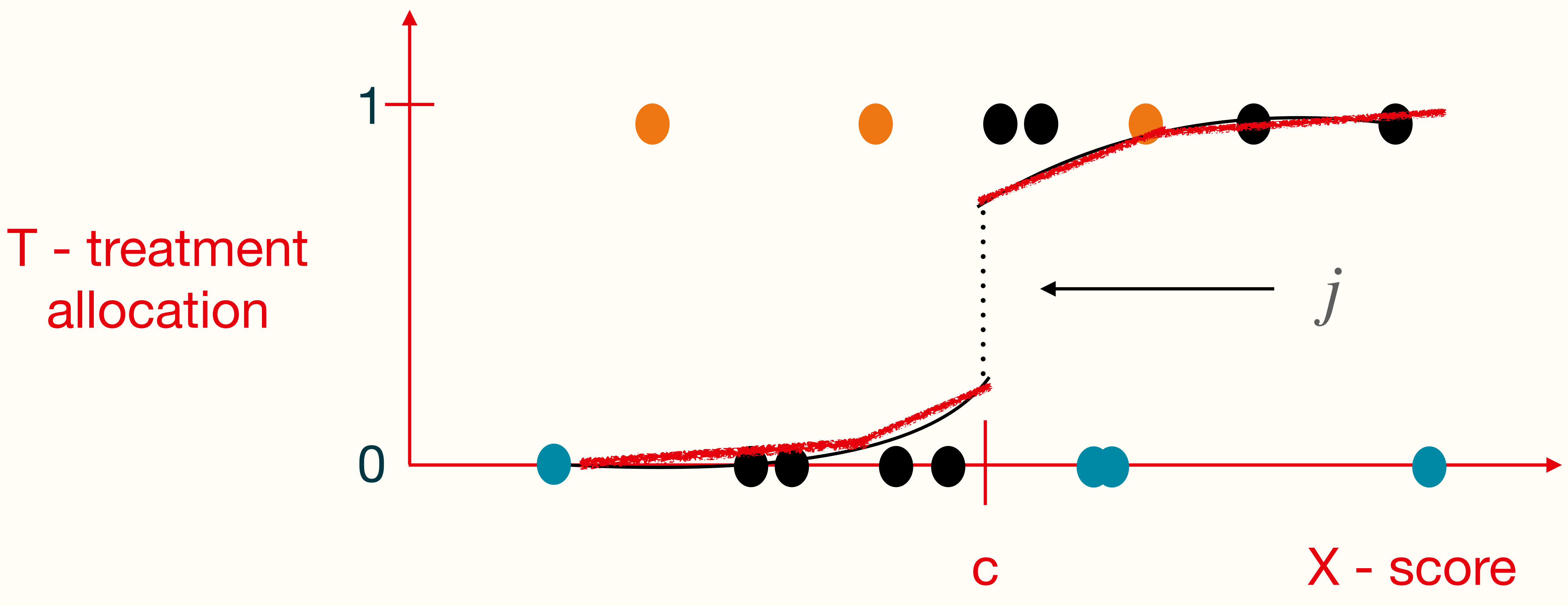
- Function is increasing
- Convex/concave on each side of the cutoff

$$T|_{X=x} \sim ber(p(x))$$

- p(x) : two connected linear functions on each side of the cutoff

Treatment probability

Bayesian approach



Assumptions:

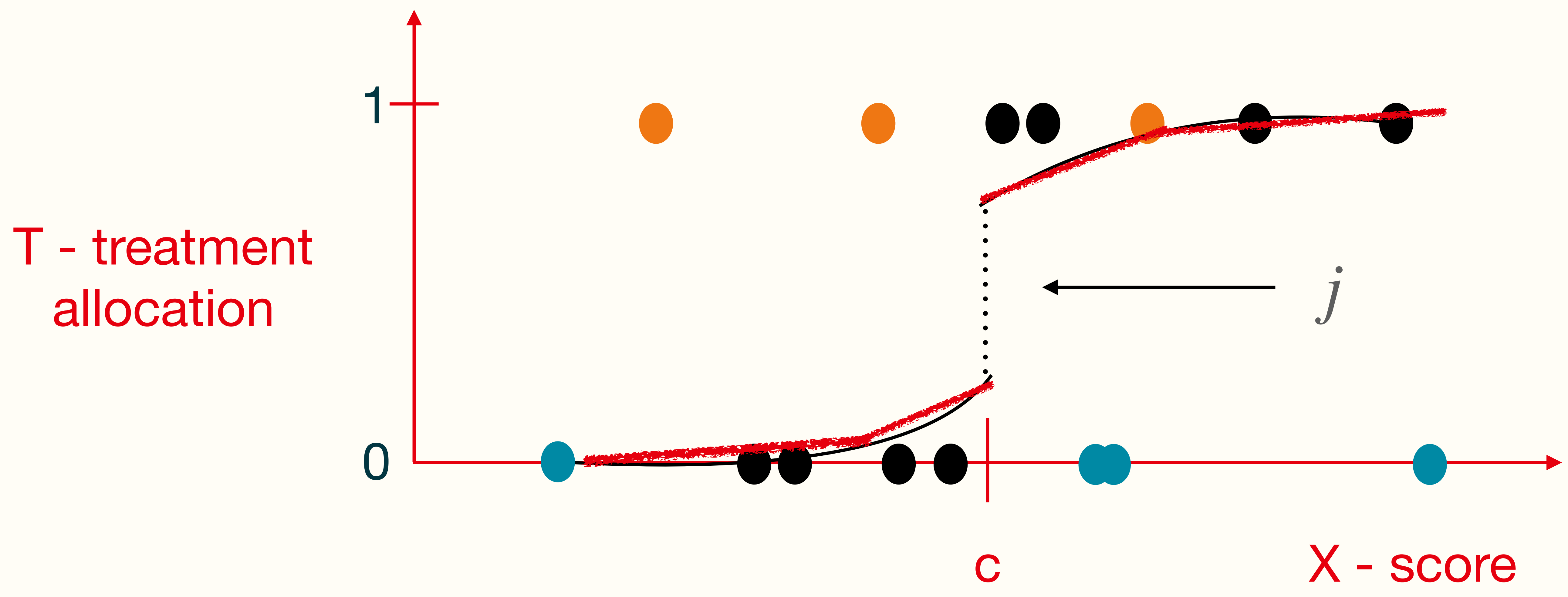
- Function is increasing
- Convex/concave on each side of the cutoff

$$T|_{X=x} \sim ber(p(x))$$

- p(x) : two connected linear functions on each side of the cutoff
- Additional requirement: minimal jump size

Treatment probability

Bayesian approach



$$a_{0,l}, b_{0,l}, a_{1,l}, b_{1,l}, a_{0,r}, b_{0,r}, a_{1,r}, b_{1,r}, j, c$$

Uniform priors

Uniform prior

Informative prior

Assumptions:

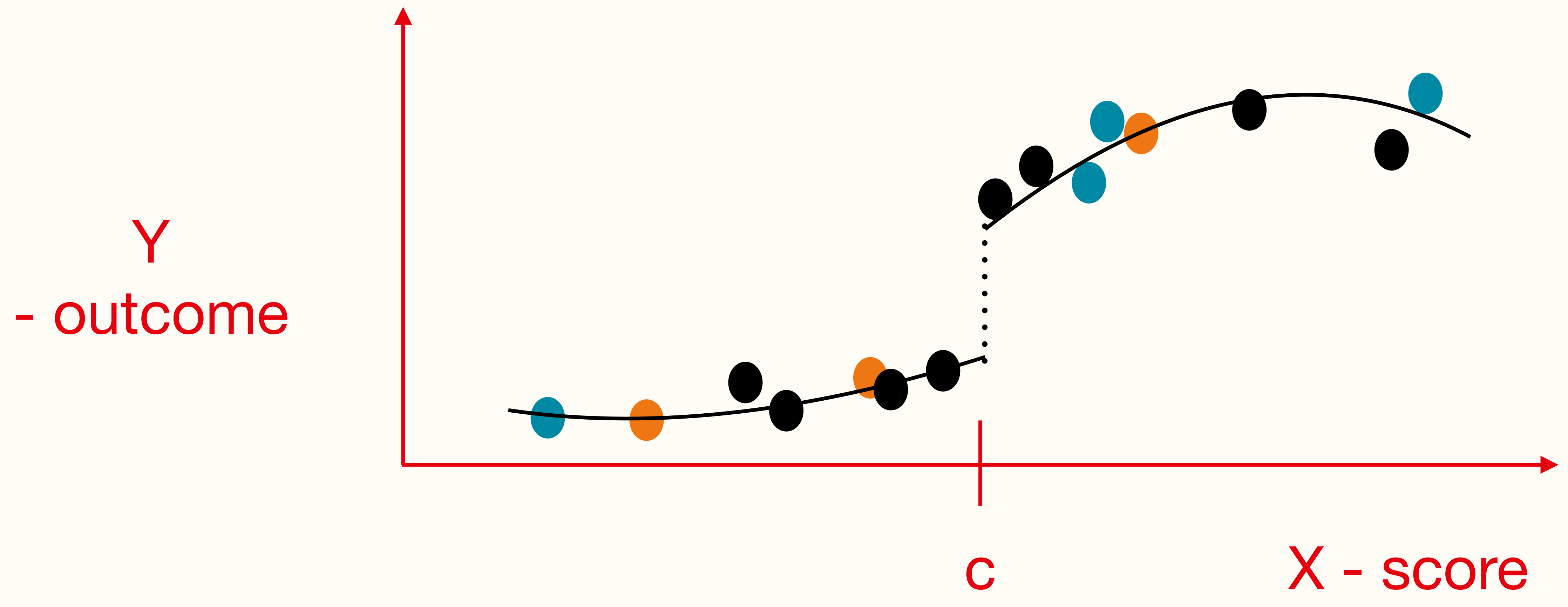
- Function is increasing
- Convex/concave on each side of the cutoff

$$T|_{X=x} \sim ber(p(x))$$

- p(x) : two connected linear functions on each side of the cutoff
- Additional requirement: minimal jump size

Outcome function

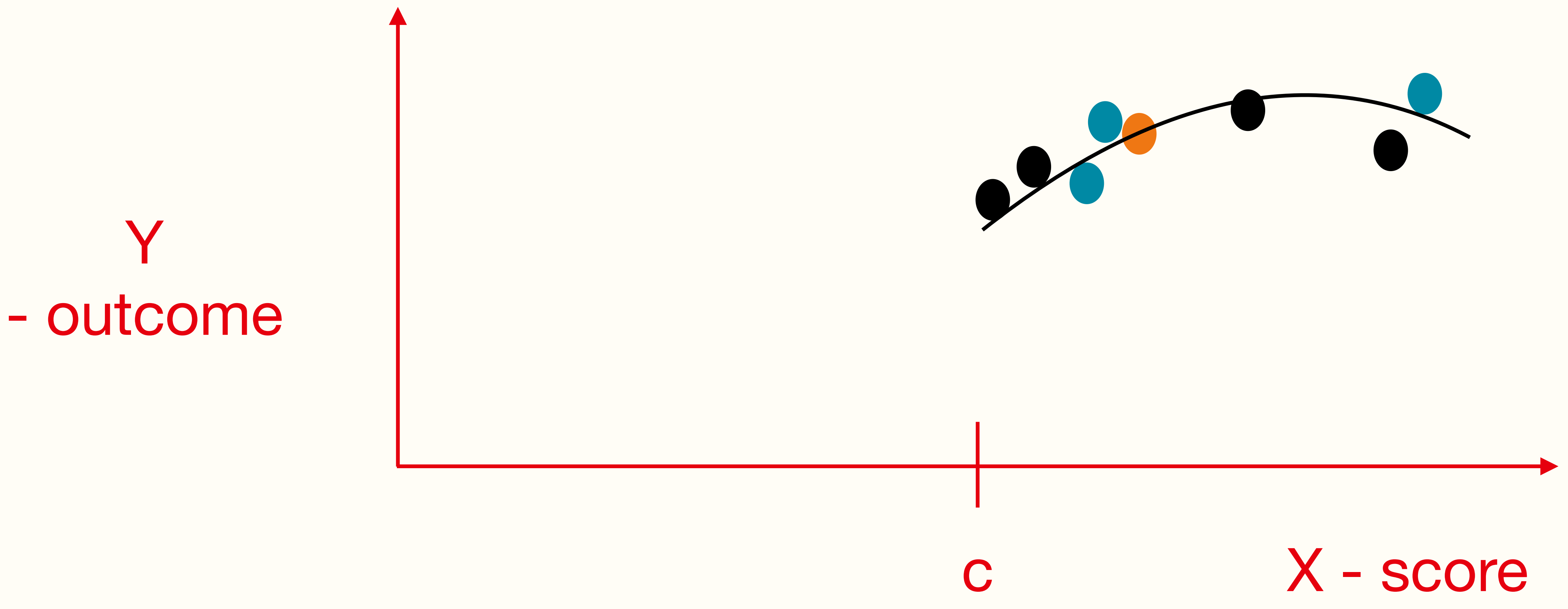
How to make Bayesian inference (more) local?



Problem: Bayesian model fits function to the whole data.

Outcome function

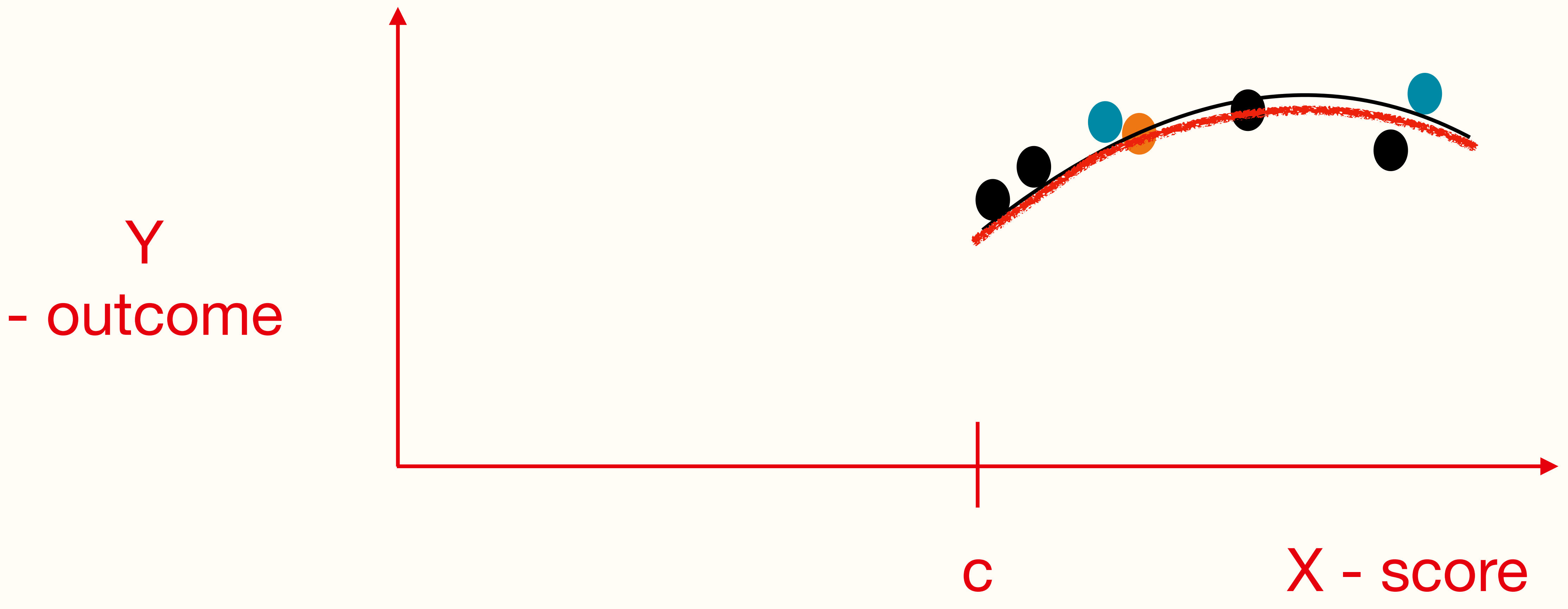
How to make Bayesian inference (more) local?



Problem: Bayesian model fits function to the whole data.

Outcome function

How to make Bayesian inference (more) local?

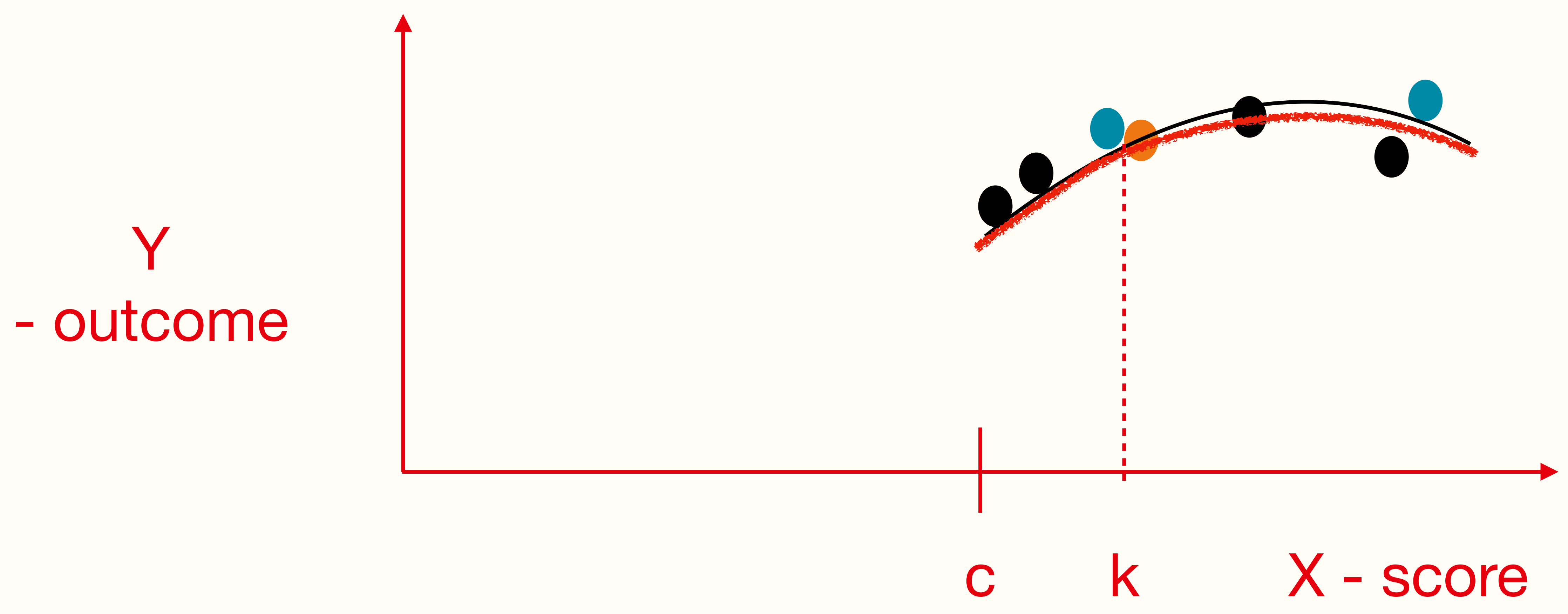


Problem: Bayesian model fits function to the whole data.

$$Y(X) = \mathcal{N}(0,\sigma^2) + \begin{cases} a_{0,r} + a_{1,r}(X-k) & \text{for } c \leq X < k \\ a_{0,r} + a_{1,r}(X-k) + a_{2,r}(X-k_2)^2 + a_{3,r}(X-k)^3 & \text{for } c < k \leq X \end{cases}$$

Outcome function

How to make Bayesian inference (more) local?



Problem: Bayesian model fits function to the whole data.

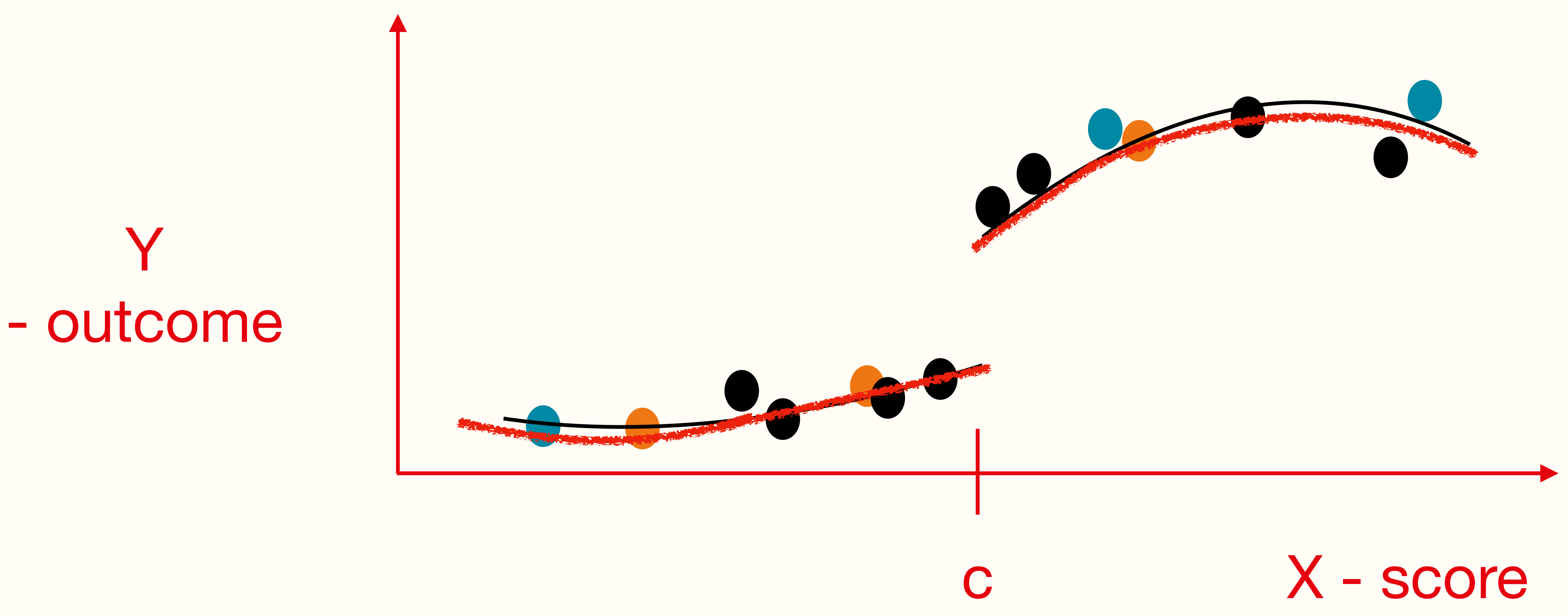
Idea: turn a disadvantage into an advantage.

More information, less uncertainty.

$$Y(X) = \mathcal{N}(0,\sigma^2) + \begin{cases} a_{0,r} + a_{1,r}(X-k) & \text{for } c \leq X < k \\ a_{0,r} + a_{1,r}(X-k) + a_{2,r}(X-k_2)^2 + a_{3,r}(X-k)^3 & \text{for } c < k \leq X \end{cases}$$

Outcome function

How to make Bayesian inference (more) local?



Problem: Bayesian model fits function to the whole data.

Idea: turn a disadvantage into an advantage.

More information, less uncertainty.

$$Y(X) = \mathcal{N}(0,\sigma^2) + \begin{cases} a_{0,l} + a_{1,l}(X - k_1) & \text{for } k_1 < X < c \\ a_{0,l} + a_{1,l}(X - k_1) + a_{2,l}(X - k1)^2 + a_{3,l}(X - k1)^3 & \text{for } X \leq k_1 < c \\ a_{0,r} + a_{1,r}(X - k_2) & \text{for } c \leq X < k_2 \\ a_{0,r} + a_{1,r}(X - k_2) + a_{2,r}(X - k_2)^2 + a_{3,r}(X - k_2)^3 & \text{for } c < k_2 \leq X \end{cases}$$

How to combine the two models?

Cut posterior approach: first sample cutoff location, then treatment effect

Joint model: sample all parameters jointly,
feedback between the outcome function and probability function

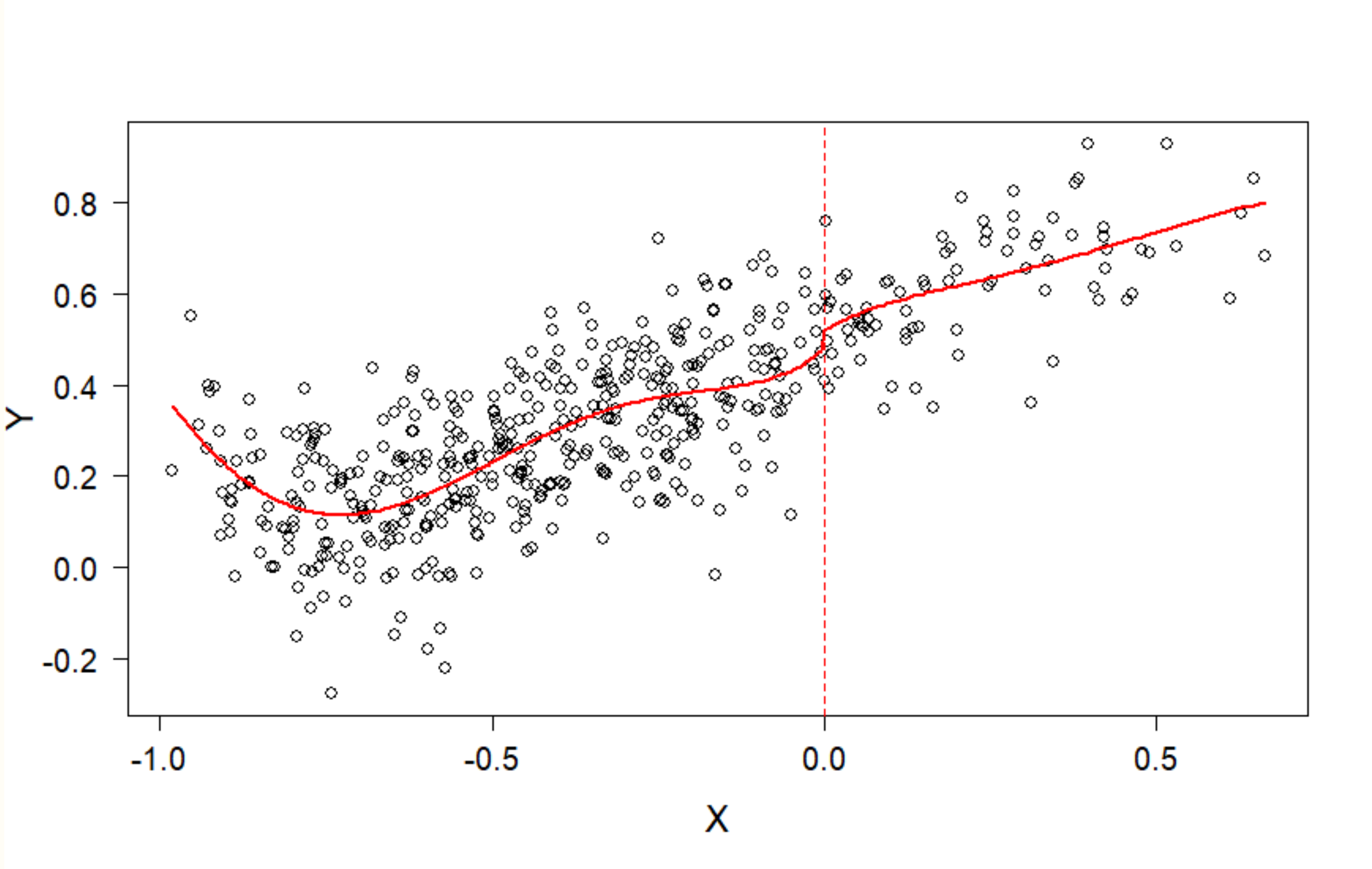
How to combine the two models?

~~**Cut posterior approach:** first sample cutoff location, then treatment effect~~

Joint model: sample all parameters jointly,
feedback between the outcome function and probability function

Simulation results

Lee function



4th degree polynomial

Small jump of 0.04

$$X \sim 2beta(2,4) - 1$$

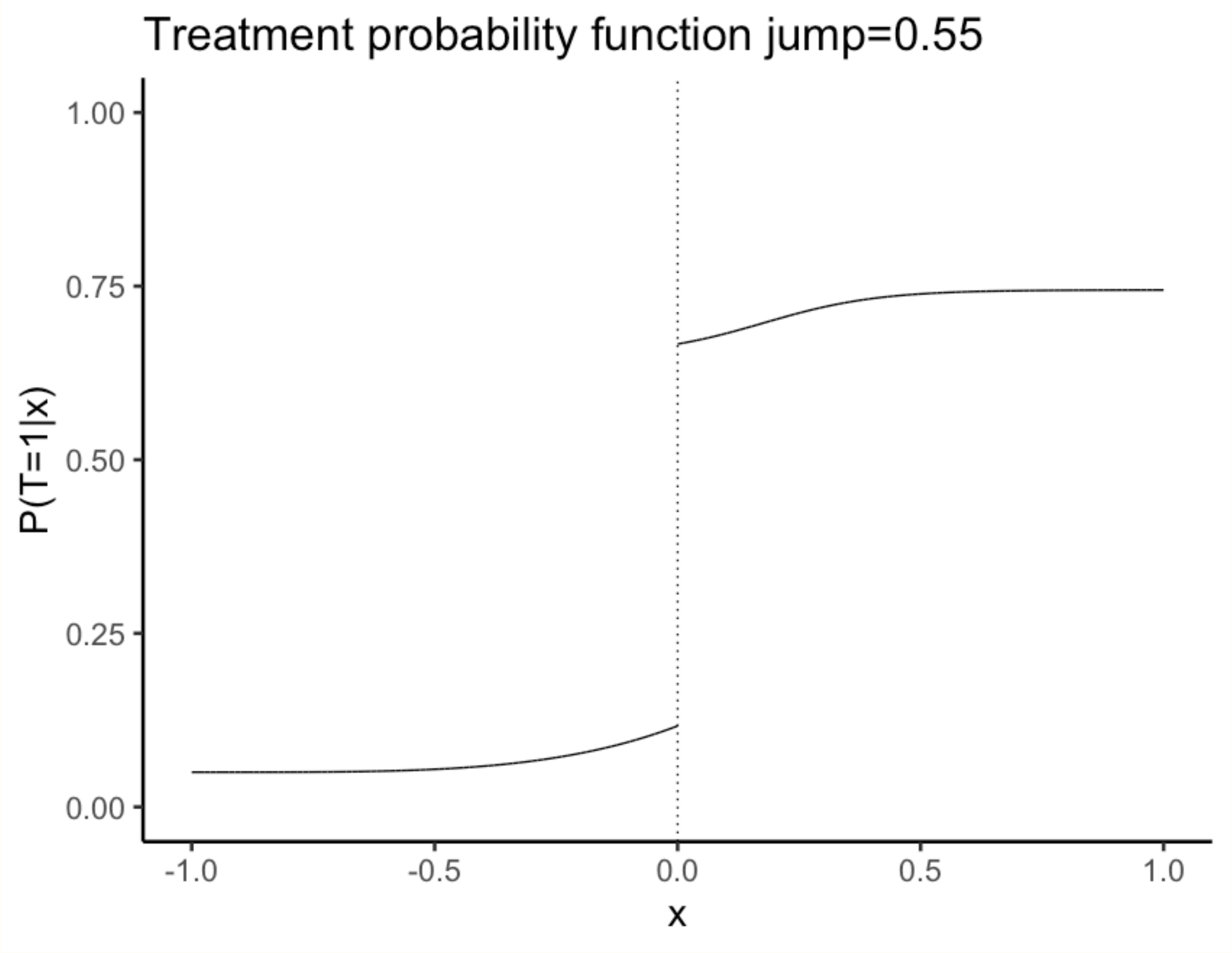
Simulation results

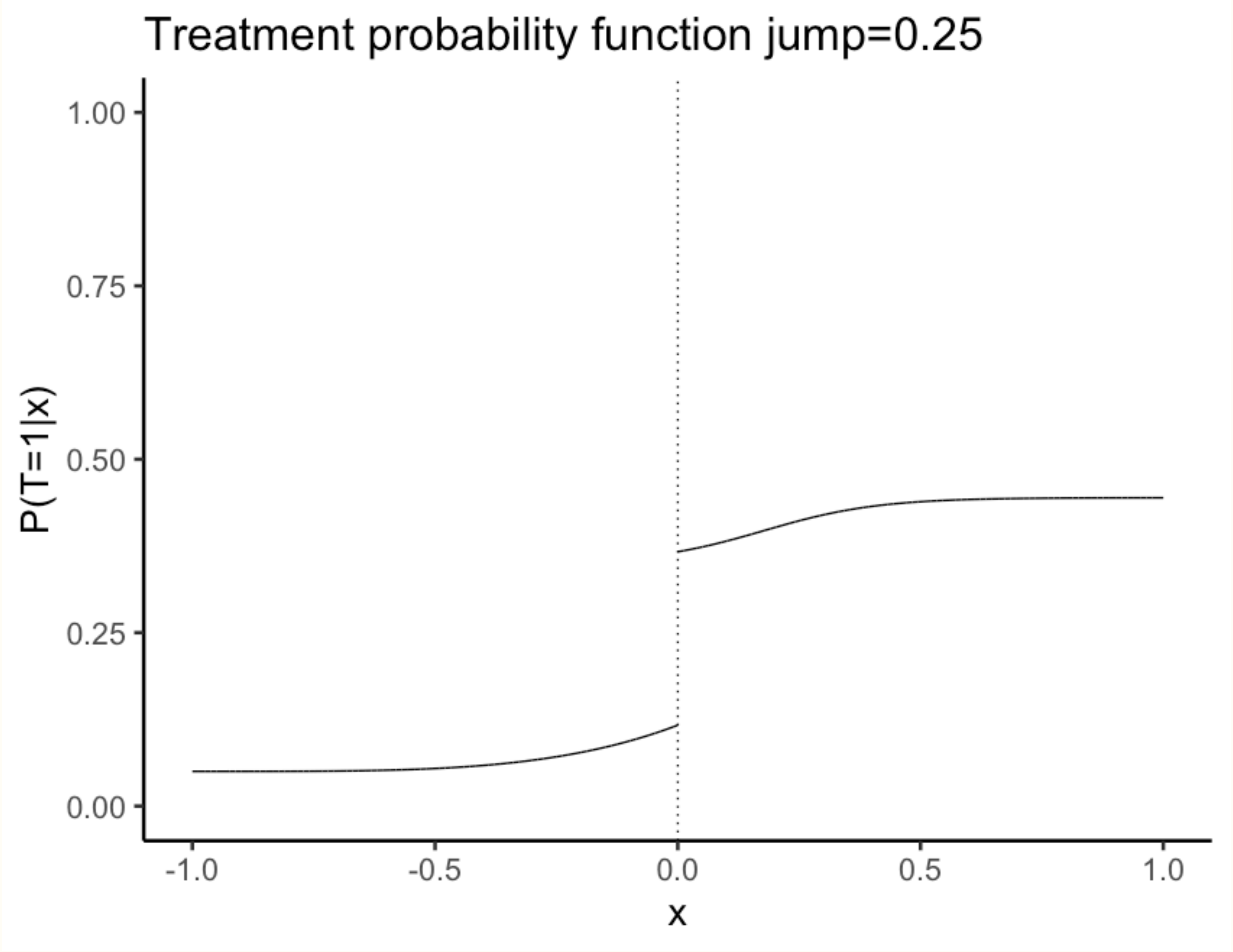
Treatment effect = 0.04, 100 simulations, 500 data points

	Local linear regression	Localised Cubic	Cubic
Absolute error	61	46	0.09
95% CI length	0.25	0.15	0.2
95% CI coverage	0.89	0.93	0.59

Simulation results

Let's add fuzziness





Simulation results

Treatment effect approx 0.07, j=0.55, 50 simulations

	Local linear regression	Localised Cubic Known c	Localised Cubic Unknown c
Absolute error	0.14	0.1	0.11
95% CI length	0.81	0.34	0.37
95% CI coverage	0.96	0.92	0.92

Simulation results

Treatment effect approx 0.07, j=0.25, 50 simulations

	Local linear regression	Localised Cubic Known c	Localised Cubic Unknown c
Absolute error	0.2335484	0.1683406	0.5739337
95% CI length	1.636036	0.5632158	2.852051
95% CI coverage	0.96	0.96	0.94

Coming up: arXiv manuscript, r package

Do you have some interesting dataset?

Reach out!
j.m.kowalska@amsterdamumc.nl