# Regression Discontinuity Design

May 2025

# Reading this week

- MixTape chapter on Regression Discontinuity
- Applied paper:
  - Card et al. *The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare*

# Regression Discontinuity Design (RDD)

- **Goal:** Estimate some causal effect of a treatment D on some outcome Y
- Cannot compare treated and untreated directly because of self-selection (selection bias)
- What if treatment assignment is based on **meeting some threshold?**
  - Under some assumptions can identify causal effects
  - RDD formalized this

# Example from Card et al.

- **Research Question:** What is the effect of universal insurance coverage (D) on health care utilization (Y)?
  - Our "treatment": Universal insurance coverage (Medicare)
  - Our outcome Y: Health care Utilization

- Cannot compare those with and without insurance directly:
  - Insurance coverage is "**endogenous**" (e.g., healthier individuals more likely to have insurance, which affects utilization) – Selection bias!

- Age threshold creates an **objective change** in insurance status that is not influenced by confounding factors ("credible source of **exogenous variation**")
  - When an individual turns 65, they become eligible for Medicare – a change that occurs independently of individual's circumstances

# At 65

Two indicators for health insurance coverage:

- Overall **coverage** increase from at age 65
- **Coverage generosity:** Proportion with 2 or more insurance policies increased

- The “treatment” assignment (Medicare eligibility) suddenly changes right at the cutoff

- We expect that there will be a **discontinuity** (a jump) in the outcome right at the cutoff

# Changes in “treatment” (insurance coverage) at 65

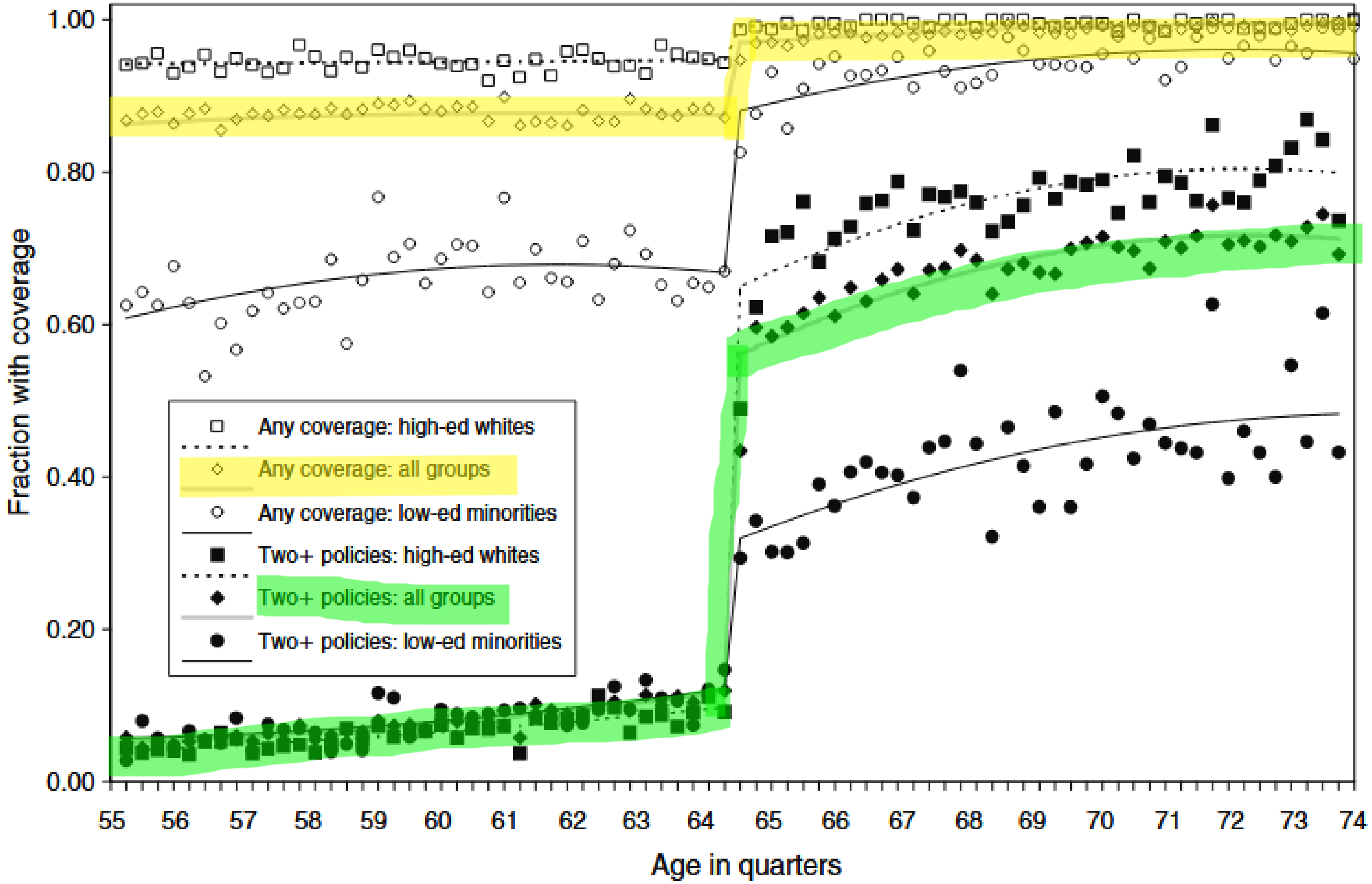



FIGURE 1. COVERAGE BY ANY INSURANCE AND BY TWO OR MORE POLICIES, BY AGE AND DEMOGRAPHIC GROUP

# Discontinuity of outcome (admissions) at cutoff

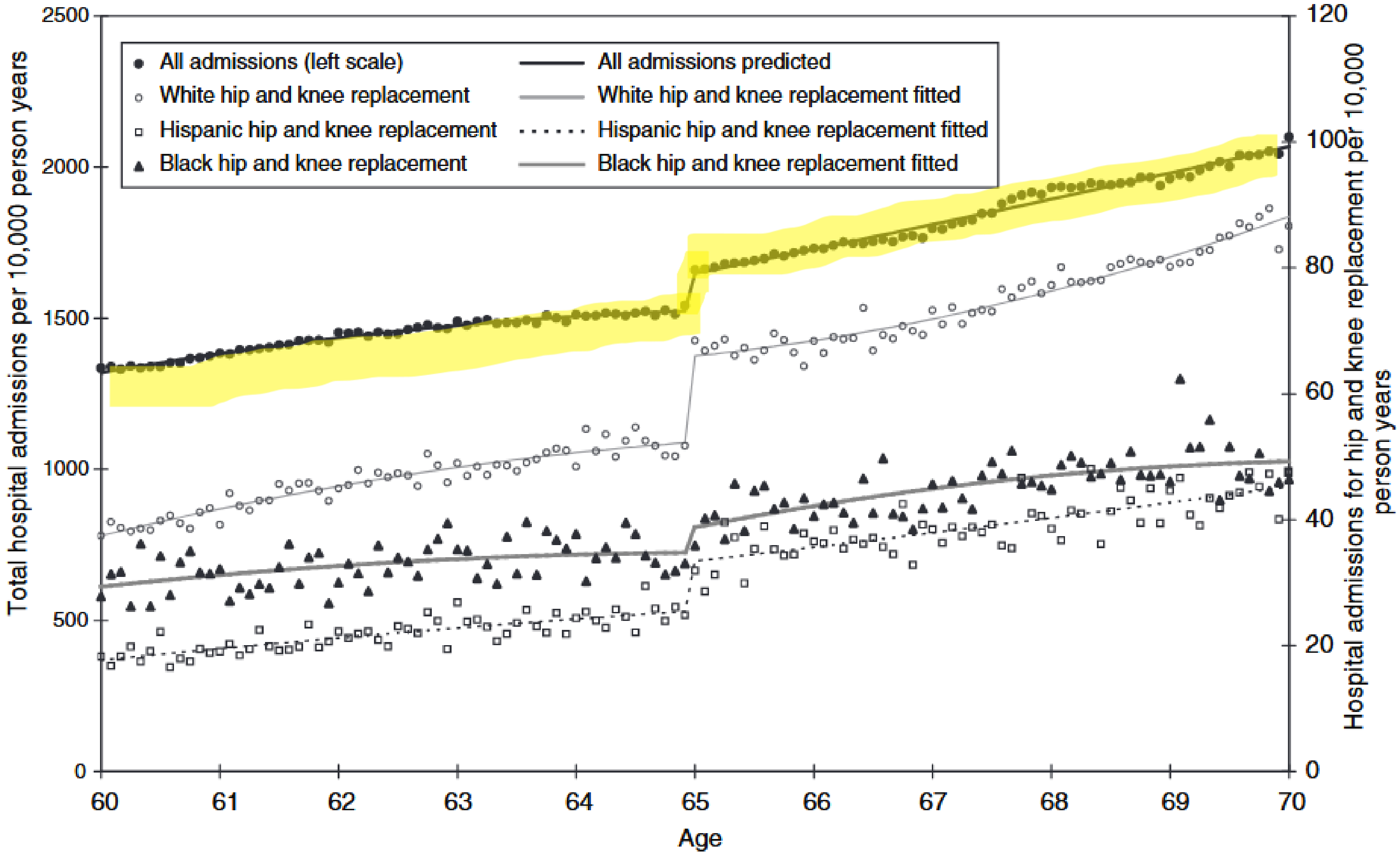



FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

# First, terminology

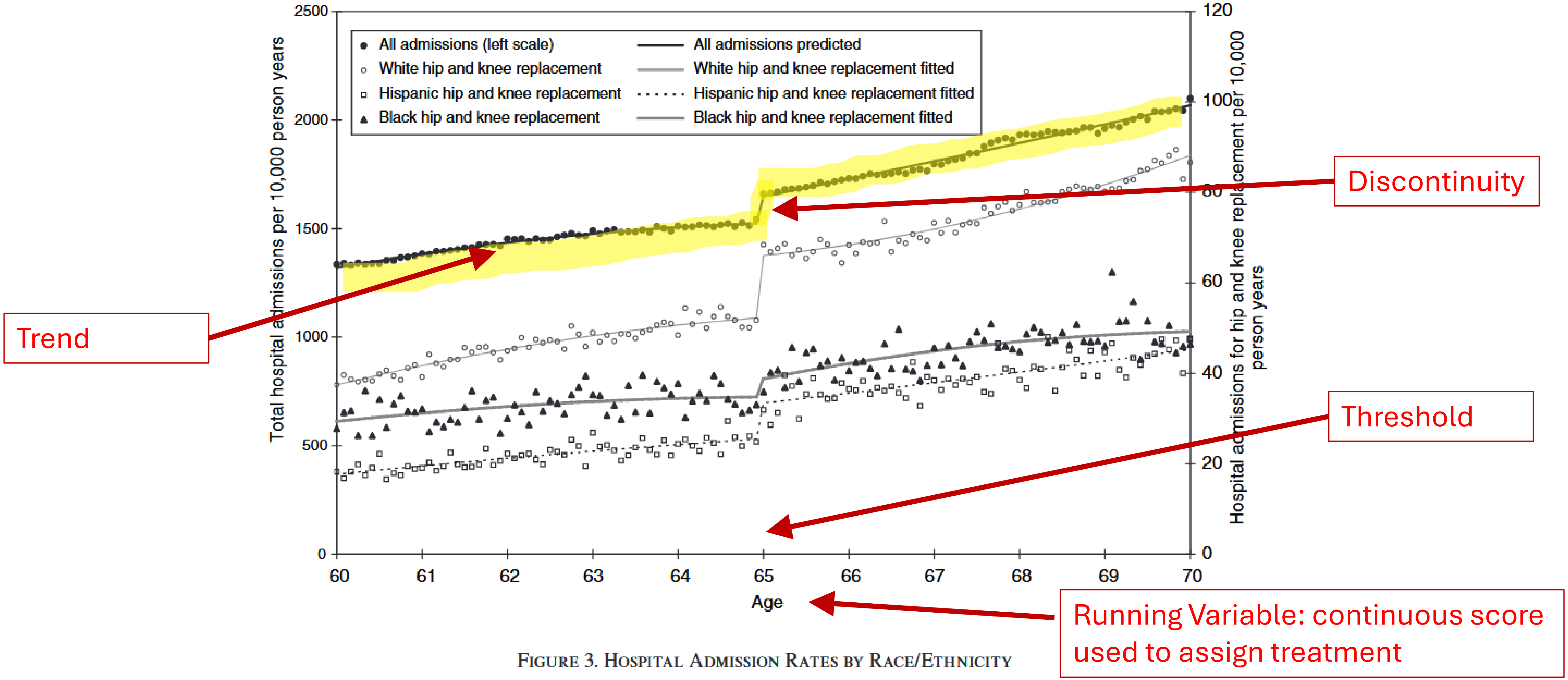



FIGURE 3. HOSPITAL ADMISSION RATES BY RACE/ETHNICITY

# Switching Equation

1. Potential outcomes: $Y_i^1$, $Y_i^0$
   - Hypotheticals in worlds with or without treatment
2. One of these two is chosen when a treatment, $D_i$ , is chosen according to the switching equation:

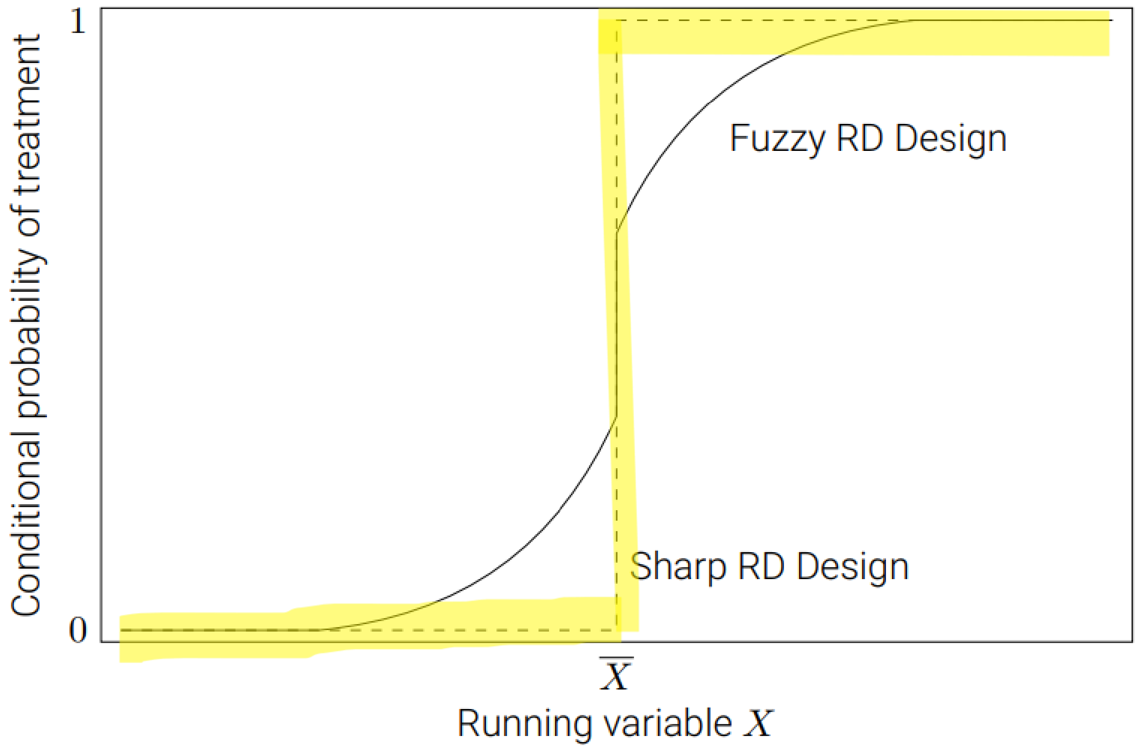
$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

$Y_i$ is the **observed outcomes**

**we are moving from potential outcomes to observed outcomes based on treatment assignment

# Sharp RDD

- Treatment is a deterministic and discontinuous function of running variable, X.
- Example: Medicare Coverage at 65




$$D_i = \begin{cases} 1 \text{ if} & X_i \geq c_0 \\ 0 \text{ if} & X_i < c_0 \end{cases}$$

# Sharp RDD

- **Common support** problems: there will NEVER be a unit of treatment and control across the running variable
- Requires **extrapolation** (prediction beyond data support) using models like regression and nonparametric methods by comparing units just below and above the cutoff
  - !! Sensitive to trends, bandwidths, and number of observations
  - !! Model functional form is important

## Definition of treatment effect

The treatment effect parameter, $\delta$, is the discontinuity in the conditional expectation function:

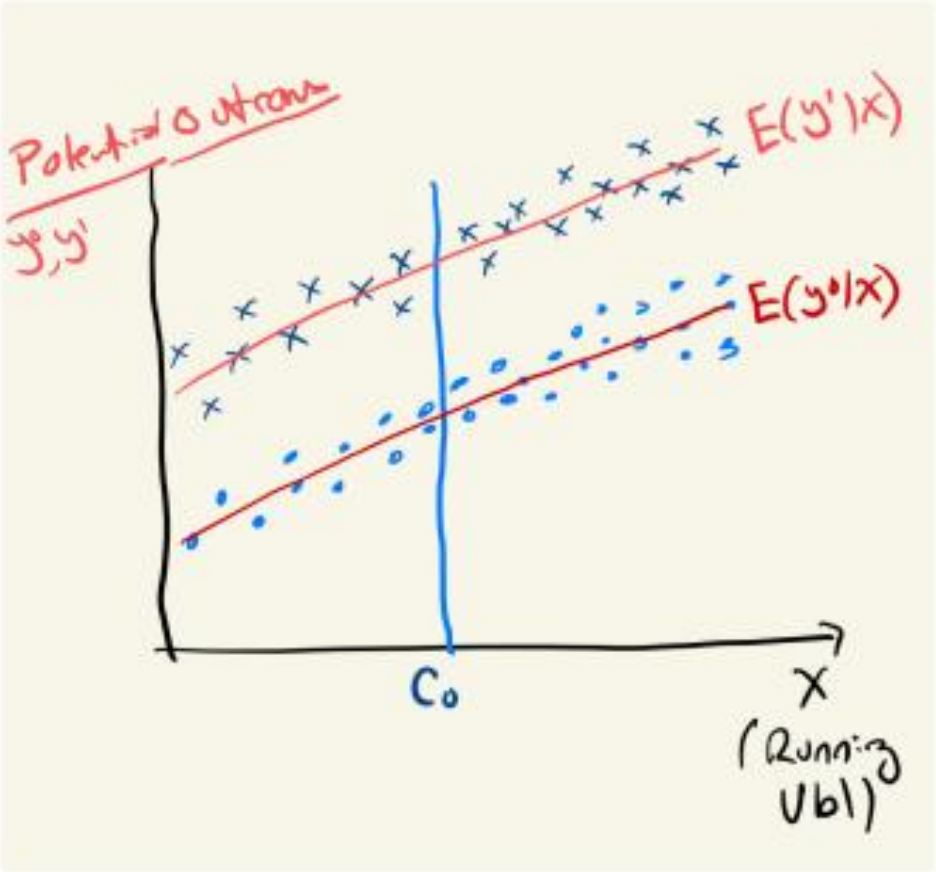
$$\begin{aligned}\delta &= lim_{X_i \rightarrow c_0} E[Y_i^1 | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i^0 | X_i = c_0] \\ &= lim_{X_i \rightarrow c_0} E[Y_i | X_i = c_0] - lim_{c_0 \leftarrow X_i} E[Y_i | X_i = c_0]\end{aligned}$$

The sharp RDD estimation is interpreted as an average causal effect (LATE) of the treatment ($D$) at the discontinuity ($c_0$)

$$\delta_{SRD} = E[Y_i^1 - Y_i^0 | X_i = c_0]$$

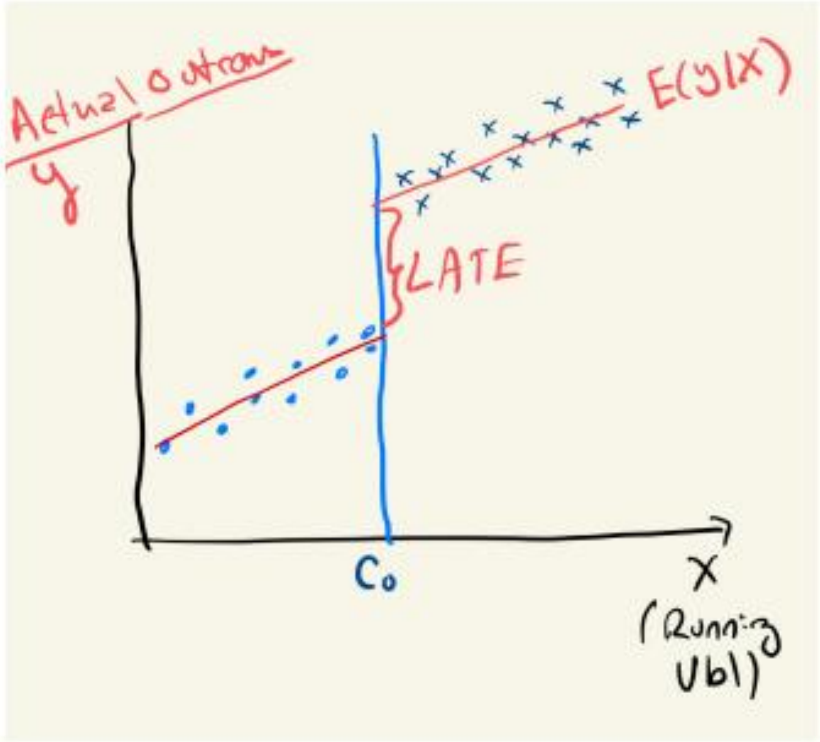
# Identifying Assumption: Smoothness (continuity)

- Smoothness of conditional expected potential outcome functions through the cutoff
- Potential outcomes $E(Y^1)$ and $E(Y^0)$are on average **smoothly changing across the threshold**
- Implies that <u>the confounders are also evolving smoothly across the cutoff</u>



# Smoothness permits extrapolation

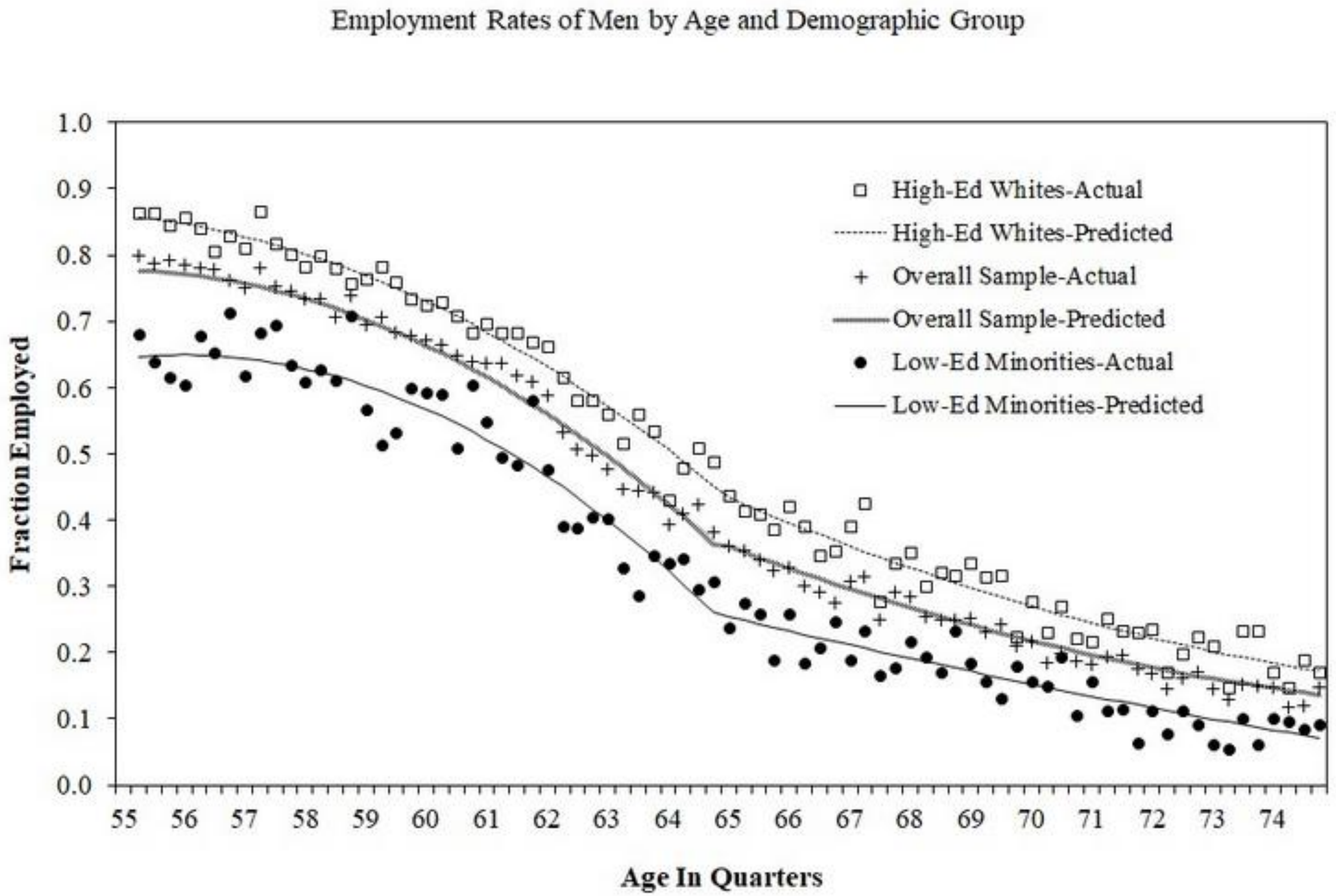
- Smoothness justifies the use of regression models to extrapolate missing potential outcomes from one side of the cutoff to the other
  - NOTE: smoothness does not imply linear
- Average causal effect is defined at the cutoff
- Estimation uses data left and right around the cutoff



# Back to the Medicare example...

- Is the smoothness (continuity) assumption valid?
- What else is changing at age 65 other than Medicare eligibility?
  - Retirement – change in employment → change in utilization
- Need to check if any potential **confounders** (e.g. employment) is evolving smoothly across the threshold

# Test for discontinuities at age 65 for confounding variables (e.g. employment)



# Notes about evaluating smoothness

- Can only check jumps in observable covariables

- **Other possible violations of the smoothness assumption:**
  - Assignment rule is known in advance
  - Agents are interested in adjusting
  - Agents have time to adjust
    - Individuals can sort themselves on the running variable (manipulation)
    - Example:  people move just on the other side of the cutoff to be eligible to treatment
    - Solution:  density test

  - Nonrandom heaping along the running variable
    - Rounding to the nearest integer
    - Solution:  donut hole RDD

# Estimation

Two ways to estimate the treatment effect at X = $c_0$

1. **Global/local parametric strategy**: borrow information from observations for from the threshold to estimate the average outcome for observations near the threshold – Sensitive to the functional forms
   - Pick the model to fit the data

2. Nonparametric kernel methods and local linear regressions - less sensitive to functional forms

# Potential outcome and nonlinear running variable

- Potential outcomes may not be linear →higher order
- The non-linearities may be different for $E[Y^1]$ and $E[Y^0]$
- Require saturated models in which you include them both individually and **interacting them with D**

$$
\begin{aligned}
E[Y_i^0|X_i] &= \alpha + \beta_{01}\tilde{X}_i + \beta_{02}\tilde{X}_i^2 + \cdots + \beta_{0p}\tilde{X}_i^p \\
E[Y_i^1|X_i] &= \alpha + \delta + \beta_{11}\tilde{X}_i + \beta_{12}\tilde{X}_i^2 + \cdots + \beta_{1p}\tilde{X}_i^p
\end{aligned}
$$

where $X_i$ is the centered running variable (i.e., $X_i - c_0$).

- The switching equation:

$$E[Y \mid X] = E[Y^0 \mid X] + \left(E[Y^1 \mid X] - E[Y^0 \mid X]\right)D$$

  - Regression model you estimate is:

$$\begin{aligned} Y_i &= \alpha + \beta_{01}\tilde{x}_i + \beta_{02}\tilde{x}_i^2 + \cdots + \beta_{0p}\tilde{x}_i^p \\ &\quad + \delta D_i + \beta_1^* D_i \tilde{x}_i + \beta_2^* D_i \tilde{x}_i^2 + \cdots + \beta_p^* D_i \tilde{x}_i^p + \varepsilon_i \end{aligned}$$

where $\beta_1^* = \beta_{11} - \beta_{01}$, $\beta_2^* = \beta_{21} - \beta_{21}$ and $\beta_p^* = \beta_{1p} - \beta_{0p}$

- The parameter of interest, the treatment effect, is the coefficient at $c_0$ or δ

# Potential problems

- Overfitting with higher order polynomial series → difficult to ensure that the functional form is specified correctly over a large range of data → increased potential for bias (Gelman and Imbens, 2019)
  - Use local linear regressions if possible
- Data intensive

# Nonparametric Kernel Methods

- **Local linear nonparametric regressions**:
  - Weighted regression restricted to a window (“Bandwidth *h*”)
  - Kernel provides the weights

  - Pick the right data to fit a given model

# Kernels

- Choose bandwidth *h:* the neighborhood around the cutoff that will be used to fit the model
- The kernel function assigns non-negative weights to each re-centered observation based on the distance between each observation's running variable score Xi and the cutoff c
- Different kinds of kernel weights:
  - **Rectangular:** uniform weights equivalent to E[Y ] at a given bin on X
  - **Triangular** draws a straight line from the threshold to the edge of the bandwidth and weights along the line
    - Weights are maximized at the cutoff
  - **Epanechnikov** is similar but is more like a parabola

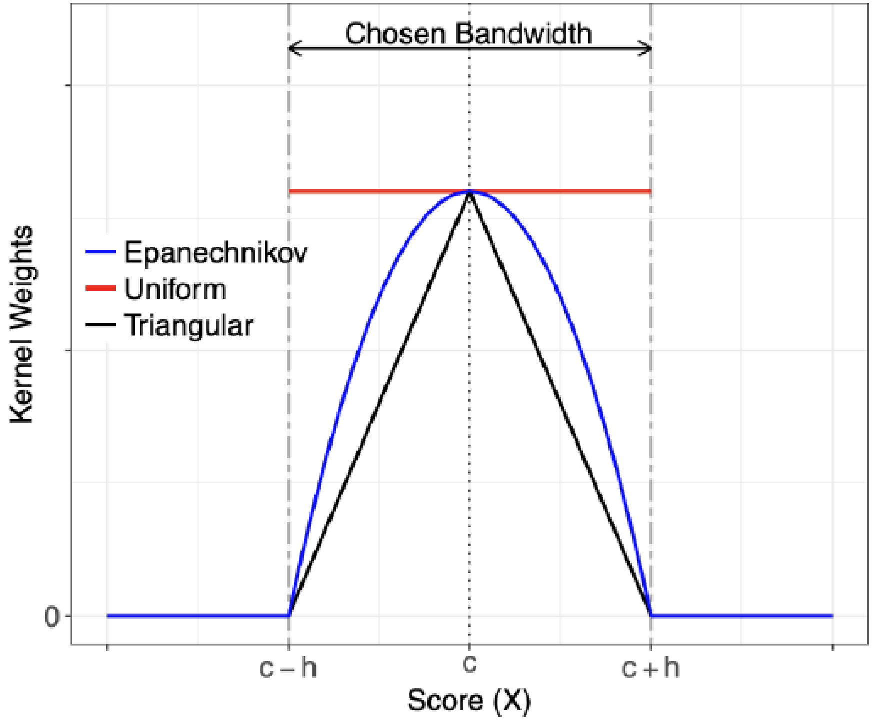



*Figure*: From Cattaneo, et al. (2019)

# Challenges

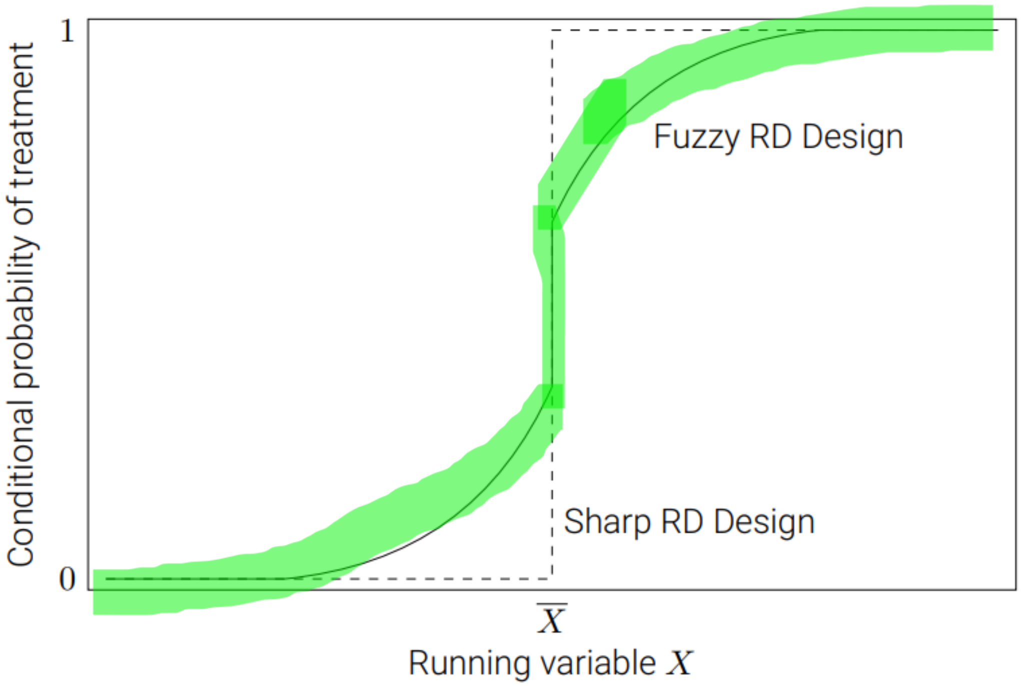
- Sensitive to the choice of bandwidth
  - Need to balance between bias and variance
  - Larger bandwidth:  better precision (more data pts), but increased potential for bias

- Choose the bandwidth that minimize the mean square error

$$MSE(\widehat{\delta}) = Bias^2(\widehat{\delta}) + Variance(\widehat{\delta})$$

# Fuzzy RDD

- Discontinuous "jump" in the **probability of treatment** when X > $c_0$.
- Not entirely deterministic – treatment assignment probabilities non-zero below threshold, and < 1 above.
- Cutoff is used as an **instrumental variable** for treatment.



# Example of fuzzy RDD

- Incentives to participate in a job training program may change discontinuously at the income-eligibility cutoff but are not powerful enough to move everyone from non-participation to participation.

# Pros of RDD

- Viewed as very credible among observational designs
  - Credible exogeneity of the treatment at threshold

- With larger dataset, possible to use shorter window (small bandwidth) for estimation → lower bias and lower variance

# Cons

- **Generalizability:** The average causal effect is for those who flipped over to treatment because their X > threshold near the threshold
  - If their treatment effects are profoundly different than anywhere else (even opposing sign), your ability to infer treatment effects elsewhere is limited
- **Smoothness assumption**
- **Manipulation** of the running variable
- Sensitivity to **sample size** and **functional form**