# STAT6061/STAT5008 – Causal Inference

# Part 5. Sensitivity analysis

**An-Shun Tai**

[1]*Department of Statistics*
*National Cheng Kung University*

[2]*Institute of Statistics and Data Science*
*National Tsing Hua University*

# Exchangeability

**Assumption (ignorability or exchangeability)**

$$Y_i(a) \perp A_i | C_i \text{ for } a = 0, 1$$

- Exchangeability is fundamentally untestable without additional assumptions, because the counterfactual outcome distributions are not directly observed in the data.

- The following discusses two approaches to assess exchangeability in observational studies.

- Here, the term "assess" reflects a weaker notion than "test":

- Assessment refers to supplementary analyses that support or cast doubt on the initial assumptions.

- Testing, in contrast, refers to formal statistical procedures.

# Assessing exchangeability: negative outcomes

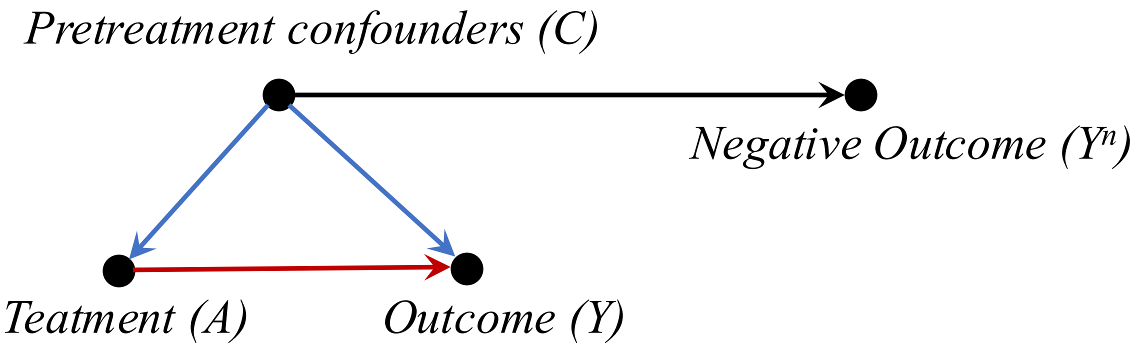
- Negative outcomes denoted as $Y^n$ is an outcome similar to $Y$ and ideally, shares the same confounding structure as $Y$.

- If the exchangeability holds, that is $Y(a) \perp A|C$, then we have $Y^n(a) \perp A|C$.

- The causal effect of $A$ on $Y^n$ is

$$\mathbb{E}\big(Y^n(1) - Y^n(0)\big)$$

which should be zero.



# Assessing exchangeability: negative outcomes (Examples)

**Example 1: Cornfield et al. (1959) – Smoking and Lung Cancer**

- Investigated the causal relationship between cigarette smoking and lung cancer using observational data.
- Controlled for many confounders, but unmeasured confounding remained a concern.
- As a negative outcome, they examined the effect of smoking on car accidents, where no biological link is expected.
- Found no association between smoking and car accidents, supporting the validity of the smoking–lung cancer link.

**Example 2: Imbens and Rubin (2015) – Lagged Outcomes**

- Proposed using the lagged outcome as a negative control outcome, assuming it shares similar confounding structure.
- Since the lagged outcome occurs before treatment, its causal effect must be zero.
- Useful as a diagnostic check—any significant effect indicates possible confounding.
- Caution: Lagged outcomes are often used as covariates, not negative controls, so interpretation must be careful.

**Example 3: Jackson et al. (2006) – Flu Vaccination Studies**

- Observational studies suggested that influenza vaccination reduced hospitalization and all-cause mortalityamong the elderly.
- Jackson et al. were skeptical due to the large effect sizes and used pre-influenza season mortality as a negative outcome.
- Found significant effects before influenza season, when vaccines should have no biological impact, indicating residual confounding.

# Assessing exchangeability: negative exposures

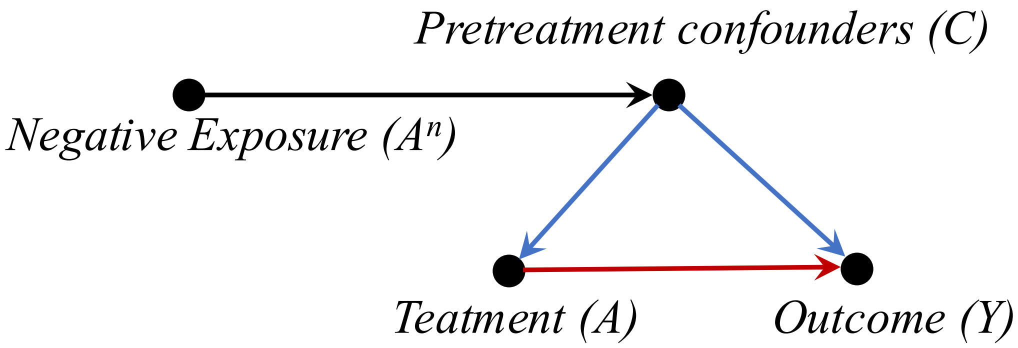
- Negative exposures denoted as $A^n$ is a variable similar to $A$ and shares the same confounding structure as $A$.

- If the exchangeability holds, that is $Y(a) \perp A|C$, then we have $Y(a) \perp A^n|C$.

- The causal effect of $A^n$ on $Y$ is

$$\mathbb{E}\big(Y(1^n) - Y(0^n)\big)$$

which should be zero.



# Assessing exchangeability: negative exposures (Example)

**Example: Sanderson et al. (2017) – Intrauterine Exposure**

- Introduced the concept of negative exposures to assess the causal effect of maternal intrauterine exposures on later child outcomes.

- Strategy: Compare the association of a maternal exposure during pregnancy with the outcome to that of the paternal exposure, which should not have a direct intrauterine effect.

- Examples include:
  - Maternal vs. paternal smoking and offspring outcomes
  - Maternal vs. paternal BMI and later offspring BMI or autism spectrum disorder

- Key assumption:

If the effect is truly intrauterine, then the maternal association should be stronger than the paternal one.

- Interpretation:

A similar association for maternal and paternal exposures may indicate shared familial or environmental confounding, rather than a true causal effect from intrauterine exposure.

# Assessing exchangeability: E-value
(VanderWeele and Ding, 2017)

- The **E-value** is a sensitivity analysis metric that quantifies **how strong unmeasured confounding would need to be** to explain away a **causal effect estimate** from an observational study.

- Formula

- If the observed risk ratio (RR) is **greater than 1**, the E-value is:

$$\text{E-value} = RR + \sqrt{RR(RR-1)}$$

- Interpretation

- A **higher E-value** implies that a very strong unmeasured confounder would be needed to negate the result
→ more robust to unmeasured confounding

- A **lower E-value** suggests that the observed effect could be more easily explained away by a modest confounder
→ less robust

# Assessing exchangeability: E-value

(VanderWeele and Ding, 2017)

### Theorem

Let $A$ denote a binary exposure, $Y$ a binary outcome, $X$ a set of observed covariates, and $U$ an unobserved confounder.

Assume the following conditions hold:

- The observed conditional risk ratio satisfies $\mathrm{RR}^{\mathrm{obs}}_{AY|x} > 1$,
- The unmeasured confounder is positively associated with both the exposure and the outcome, conditional on $X$:
  $\mathrm{RR}_{AU|x} > 1$ and $\mathrm{RR}_{UY|x} > 1$,
- The exposure is conditionally independent of the outcome given both observed and unobserved covariates:
  $A \perp\!\!\!\perp Y \mid (X, U)$.

Then, the observed risk ratio is bounded above by:

$$\mathrm{RR}^{\mathrm{obs}}_{AY|x} \leq \frac{\mathrm{RR}_{AU|x} \cdot \mathrm{RR}_{UY|x}}{\mathrm{RR}_{AU|x} + \mathrm{RR}_{UY|x} - 1}.$$

**Proof of Theorem**

We can decompose $\mathrm{RR}^{\mathrm{obs}}_{AY|x}$ as:

$$\mathrm{RR}^{\mathrm{obs}}_{AY|x} = \frac{\Pr(Y=1 \mid A=1, X=x)}{\Pr(Y=1 \mid A=0, X=x)}$$

$$= \frac{\Pr(U=1 \mid A=1, X=x)\Pr(Y=1 \mid A=1, U=1, X=x) + \Pr(U=0 \mid A=1, X=x)\Pr(Y=1 \mid A=1, U=0, X=x)}{\Pr(U=1 \mid A=0, X=x)\Pr(Y=1 \mid A=0, U=1, X=x) + \Pr(U=0 \mid A=0, X=x)\Pr(Y=1 \mid A=0, U=0, X=x)}$$

Using the conditional independence assumption $A \perp\!\!\!\perp Y \mid (X, U)$, we simplify:

$$= \frac{\Pr(U=1 \mid A=1, X=x)\Pr(Y=1 \mid U=1, X=x) + \Pr(U=0 \mid A=1, X=x)\Pr(Y=1 \mid U=0, X=x)}{\Pr(U=1 \mid A=0, X=x)\Pr(Y=1 \mid U=1, X=x) + \Pr(U=0 \mid A=0, X=x)\Pr(Y=1 \mid U=0, X=x)}$$

Letting

$f_{1,x} = \Pr(U=1 \mid A=1, X=x)$,

$f_{0,x} = \Pr(U=1 \mid A=0, X=x)$,

and

$\mathrm{RR}_{UY|x} = \frac{\Pr(Y=1|U=1,X=x)}{\Pr(Y=1|U=0,X=x)}$,

we obtain:

$$= \frac{f_{1,x} \cdot \mathrm{RR}_{UY|x} + 1 - f_{1,x}}{f_{0,x} \cdot \mathrm{RR}_{UY|x} + 1 - f_{0,x}} = \frac{(\mathrm{RR}_{UY|x} - 1) f_{1,x} + 1}{(\mathrm{RR}_{UY|x} - 1) f_{0,x} + 1}$$

Now let

$\mathrm{RR}_{AU|x} = \frac{f_{1,x}}{f_{0,x}}$,

and we get:

$$\mathrm{RR}^{\mathrm{obs}}_{AY|x} = \frac{(\mathrm{RR}_{UY|x} - 1) f_{1,x} + 1}{(\mathrm{RR}_{UY|x} - 1) f_{1,x} / \mathrm{RR}_{AU|x} + 1}$$

This expression is **increasing in** $f_{1,x}$.

Setting $f_{1,x} = 1$ gives the upper bound:

$$\mathrm{RR}^{\mathrm{obs}}_{AY|x} \le \frac{(\mathrm{RR}_{UY|x} - 1) + 1}{\frac{\mathrm{RR}_{UY|x} - 1}{\mathrm{RR}_{AU|x}} + 1} = \frac{\mathrm{RR}_{AU|x} \cdot \mathrm{RR}_{UY|x}}{\mathrm{RR}_{AU|x} + \mathrm{RR}_{UY|x} - 1}$$

# E-value

**Lemma**

Let $\beta(w_1, w_2) = \dfrac{w_1 w_2}{w_1 + w_2 - 1}$, defined for $w_1 > 1$ and $w_2 > 1$. Then:

1. $\beta(w_1, w_2)$ is **symmetric** in $w_1$ and $w_2$;
2. $\beta(w_1, w_2)$ is **monotone increasing** in both $w_1$ and $w_2$;
3. $\beta(w_1, w_2) \leq \min(w_1, w_2)$;
4. $\beta(w_1, w_2) \leq \dfrac{w^2}{2w - 1}$, where $w = \max(w_1, w_2)$.

Using **Theorem** and **Lemma**, we have:

$$\mathrm{RR}_{AU|x} \geq \mathrm{RR}^{\mathrm{obs}}_{AY|x}, \quad \mathrm{RR}_{UY|x} \geq \mathrm{RR}^{\mathrm{obs}}_{AY|x},$$

or, equivalently,

$$\min\left(\mathrm{RR}_{AU|x}, \mathrm{RR}_{UY|x}\right) \geq \mathrm{RR}^{\mathrm{obs}}_{AY|x}.$$

# E-value

If we define

$$w = \max(\mathrm{RR}_{AU|x}, \mathrm{RR}_{UY|x}),$$

then by applying **Theorem** and **Lemma**, we obtain:

$$\frac{w^2}{2w-1} \geq \beta(\mathrm{RR}_{AU|x}, \mathrm{RR}_{UY|x}) \geq \mathrm{RR}^{\mathrm{obs}}_{AY|x},$$

which implies:

$$w^2 - 2\mathrm{RR}^{\mathrm{obs}}_{AY|x} w + \mathrm{RR}^{\mathrm{obs}}_{AY|x} \geq 0,$$

This is a quadratic inequality. One root,

$$\mathrm{RR}^{\mathrm{obs}}_{AY|x} - \sqrt{\mathrm{RR}^{\mathrm{obs}}_{AY|x}(\mathrm{RR}^{\mathrm{obs}}_{AY|x} - 1)},$$

is always less than or equal to 1. Thus, we have:

$$w = \max(\mathrm{RR}_{AU|x}, \mathrm{RR}_{UY|x}) \geq \mathrm{RR}^{\mathrm{obs}}_{AY|x} + \sqrt{\mathrm{RR}^{\mathrm{obs}}_{AY|x}(\mathrm{RR}^{\mathrm{obs}}_{AY|x} - 1)}.$$

# Positivity assumption

**Assumption (positivity)**

$$0 < Pr(A = 1 \mid C) < 1$$

➢ **The Tension Between Exchangeability and Positivity**

- Including more covariates typically enhances the plausibility of the **exchangeability (unconfoundedness)** assumption by better accounting for confounding.

- However, this comes at the cost of weakening the **positivity (overlap)** assumption, as treatment assignment may become increasingly deterministic and less variable across covariate strata.

## Assumption (Strict Overlap)

There exists a constant $\eta \in (0, 1/2)$ such that the propensity score satisfies:

$$\eta \leq e(X) \leq 1 - \eta,$$

where $e(X) = \Pr(A = 1 \mid X)$.

---

## Theorem (Adapted from D'Amour et al., 2021)

Under the strict overlap assumption above, for a covariate vector $X = (X_1, \ldots, X_p)$, let $X_k$ denote the $k$-th component of $X$, and let $e = \Pr(A = 1)$ denote the marginal probability of treatment. Then:

$$\frac{1}{p} \sum_{k=1}^{p} |\mathbb{E}(X_k \mid A = 1) - \mathbb{E}(X_k \mid A = 0)| \leq p^{-1/2} C^{1/2} \left\{ e \lambda_1^{1/2} + (1 - e) \lambda_0^{1/2} \right\},$$

where

$$C = \frac{(e - \eta)(1 - e - \eta)}{e^2 (1 - e)^2 \eta (1 - \eta)}$$

is a positive constant depending only on $(e, \eta)$, and $\lambda_1$, $\lambda_0$ are the largest eigenvalues of the covariance matrices $\operatorname{cov}(X \mid A = 1)$ and $\operatorname{cov}(X \mid A = 0)$, respectively.

- Convergence of the left-hand side of the inequality to zero implies that it is not possible for all components of $X$ to exhibit persistent, non-negligible differences in means between the treatment and control groups.

  → **Strong requirment**

# Causal inference with no overlap (positivity): regression discontinuity (Thistlethwaite and Campbell, 1960)

- ➢ Thistlethwaite and Campbell (1960) first introduced the idea of regression-discontinuity analysis.

- ➢ They studied whether receiving a **certificate of merit** influenced **students' career aspirations**.

- Treatment:
  - Receiving the certificate of merit

- Outcomes:
  - *I-I′*: % planning ≥ 3 years of graduate study (PhD or MD)
  - *J-J′*: % planning to become a college teacher or scientific researcher

- ➢ **Running variable:** the variable used to assign treatment based on whether its value is above or below a pre-specified cutoff.

- Students with aptitude test scores above a predetermined cutoff were awarded a Certificate of Merit.

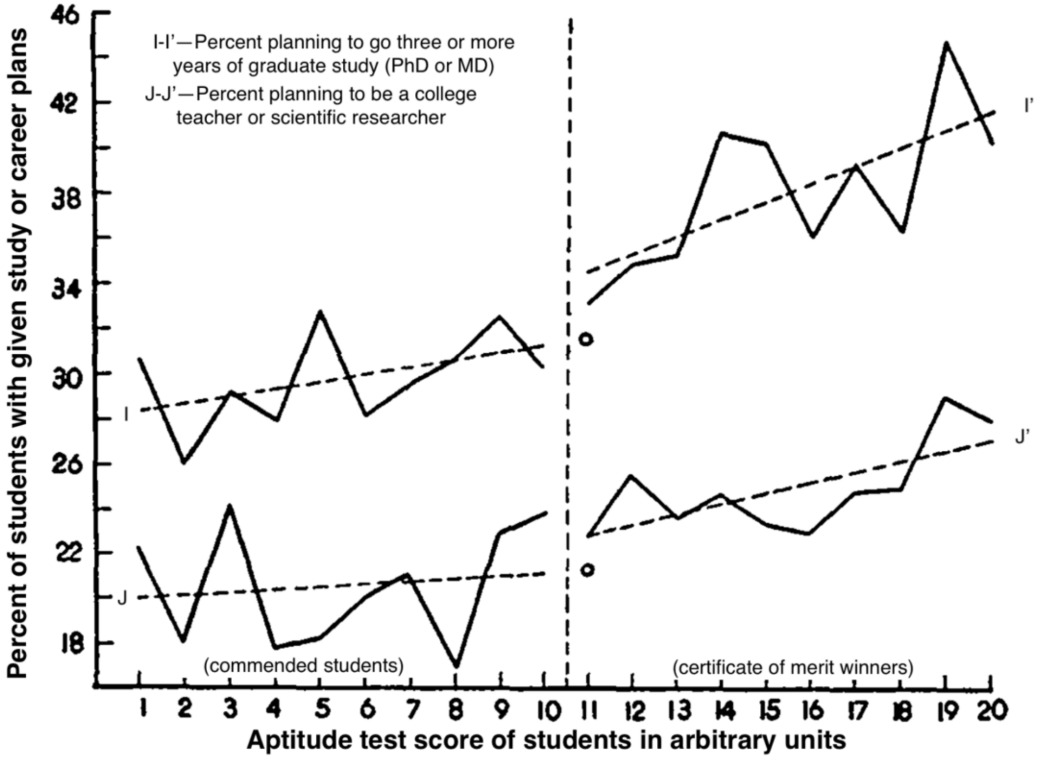



**Figure.** A graph from Thistlethwaite and Campbell (1960) with minor modifications of the unclear text in the original paper (Ding, 2024)

# Regression discontinuity design (RDD)

- RDD is a quasi-experimental method for estimating causal effects

- Random assignment is not feasible, but
- Treatment is instead determined by a known, **deterministic rule**—specifically, whether a continuous running variable exceeds a pre-specified cutoff.

- Intuition:

- Treatment is assigned according to a fixed, deterministic rule: $A = I(X \geq x_0)$
- The exchangeability assumption holds automatically:

$$A \perp (Y(1), Y(0))|X$$

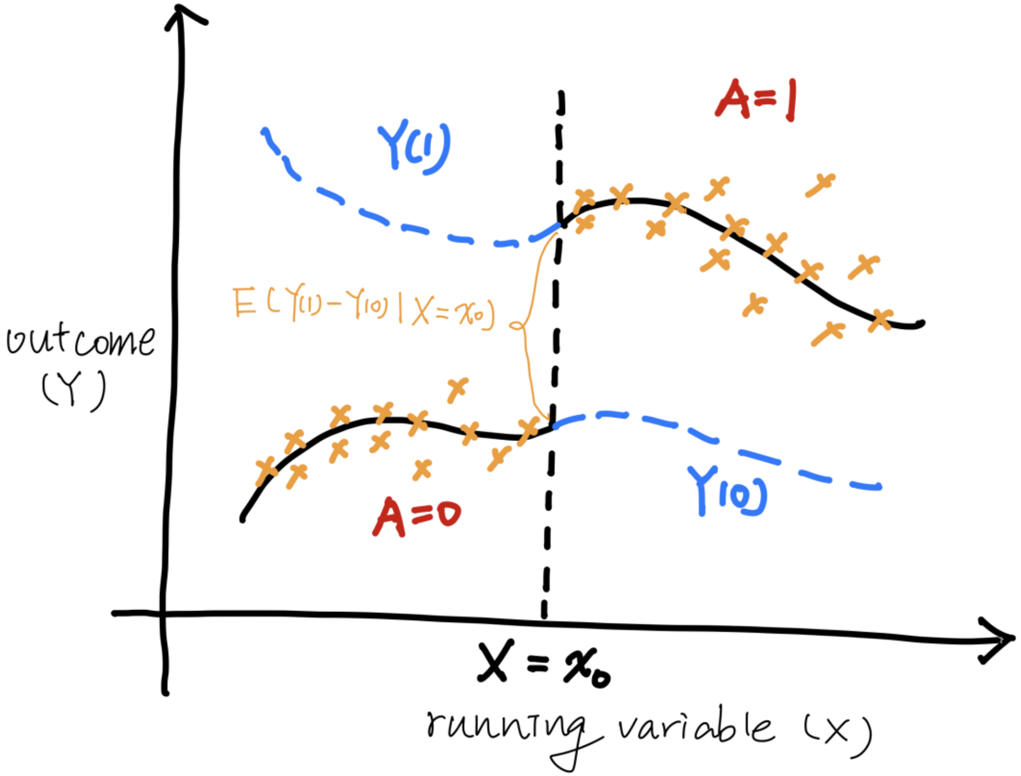
- However, in RDDs, the positivity assumption does not hold by design:

Propensity score: $e(X) = \Pr(Y = 1|X) = I(X \geq x_0) = \begin{cases} 1 & \text{if } X \geq x_0 \\ 0 & \text{if } X < x_0 \end{cases}$

# Regression discontinuity design (RDD) (cont.)

- Intervention ↔ Discontinuity at Threshold

Assuming no other factors cause discontinuous changes in the outcome around the cutoff, the observed variation in the outcome near the threshold ($X = x_0$) can be reasonably attributed to the change in the probability of receiving the treatment.



# Identification in RDDs

➢ **Local ATE at the cutoff point ($X = x_0$):**

$$\tau_{RDD}(x_0) = \mathbb{E}(Y(1) - Y(0)|X = x_0)$$

➢ **Identification Assumption:**

**1. Deterministic assumption:**

$$A = I(X \geq x_0)$$

**2. Continuity assumption:**

$\mathbb{E}(Y(1)|X = x_0)$ is continous from the right to $x_0$ and $\mathbb{E}(Y(0)|X = x_0)$ is continous from the left to $x_0$.

**Theorem 5.1**

*If both the deterministic treatment assignment rule and the continuity assumption hold, then the local ATE at the cutoff $X = x_0$ is identified by*

$$\tau_{RDD}(x_0) = \lim_{\varepsilon \to 0^+} \mathbb{E}(Y|A = 1, X = x_0 + \varepsilon) - \lim_{\varepsilon \to 0^-} \mathbb{E}(Y|A = 0, X = x_0 - \varepsilon)$$

# Estimation in RDDs

- Parametric/global method

  Step 1. Assume that

$$\mathbb{E}(Y|A=1, X=x) = \gamma_1 + \beta_1 x$$
$$\mathbb{E}(Y|A=0, X=x) = \gamma_0 + \beta_0 x$$

  Step 2. Run OLS based on treated and control data to obtain the fitted line $\hat{\gamma}_1 + \hat{\beta}_1 x$ and $\hat{\gamma}_0 + \hat{\beta}_0 x$.

  Step 3. Estimate the local ATE at the cutoff as

$$\hat{\tau}_{RDD}(x_0) = (\hat{\gamma}_1 - \hat{\gamma}_0) + (\hat{\beta}_1 - \hat{\beta}_0)x_0$$

- The estimate $\hat{\tau}_{RDD}(x_0)$ can also be obtained as the coefficient on $A_i$ from the following the OLS regression model:

$$Y_i \sim \{1, A_i, X_i - x_0, A_i(X_i - x_0)\}$$

# Estimation in RDDs

➢ Nonparametric/local method

- It focus on fitting flexible models **only near the threshold**, reducing reliance on functional form assumptions.

- Local Linear Regression (LLR) is the most commonl used nonparametric approach.

  Step 1. Separate regression lines are fit on either side of the cutoff using a kernel-weighted sample that gives more weight to observations closer to the threshold.

  Step 2. Estimates the treatment effect as the differenc in intercepts at the cutoff.

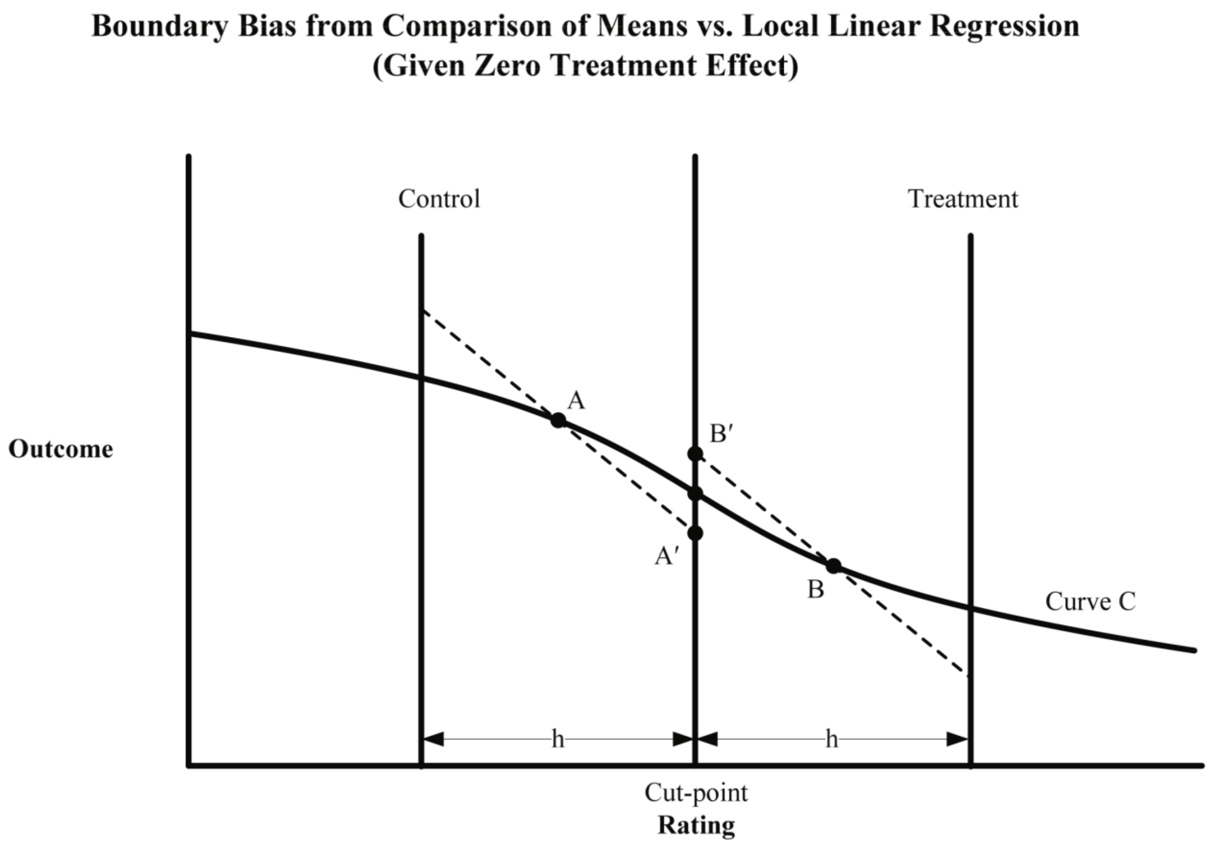



Figure adapted from Jacob et al. (2012), *A Practical Guide to Regression Discontinuity*. MDRC.

# RDD in Action: Lee (2008) on Incumbency Advantage

## ➢ Study Overview

- Objective: Estimate the causal effect of incumbency on electoral success in U.S. House elections.

- Challenge: Incumbents are inherently those who previously won; their advantage may reflect unobserved political strength, not just the fact of being in office.

## ➢ Why RDD?

- As Lee notes: "Incumbents are, by definition, those politicians who were successful in the previous election... If what makes them successful is persistent over time, they should be more successful in re-election."

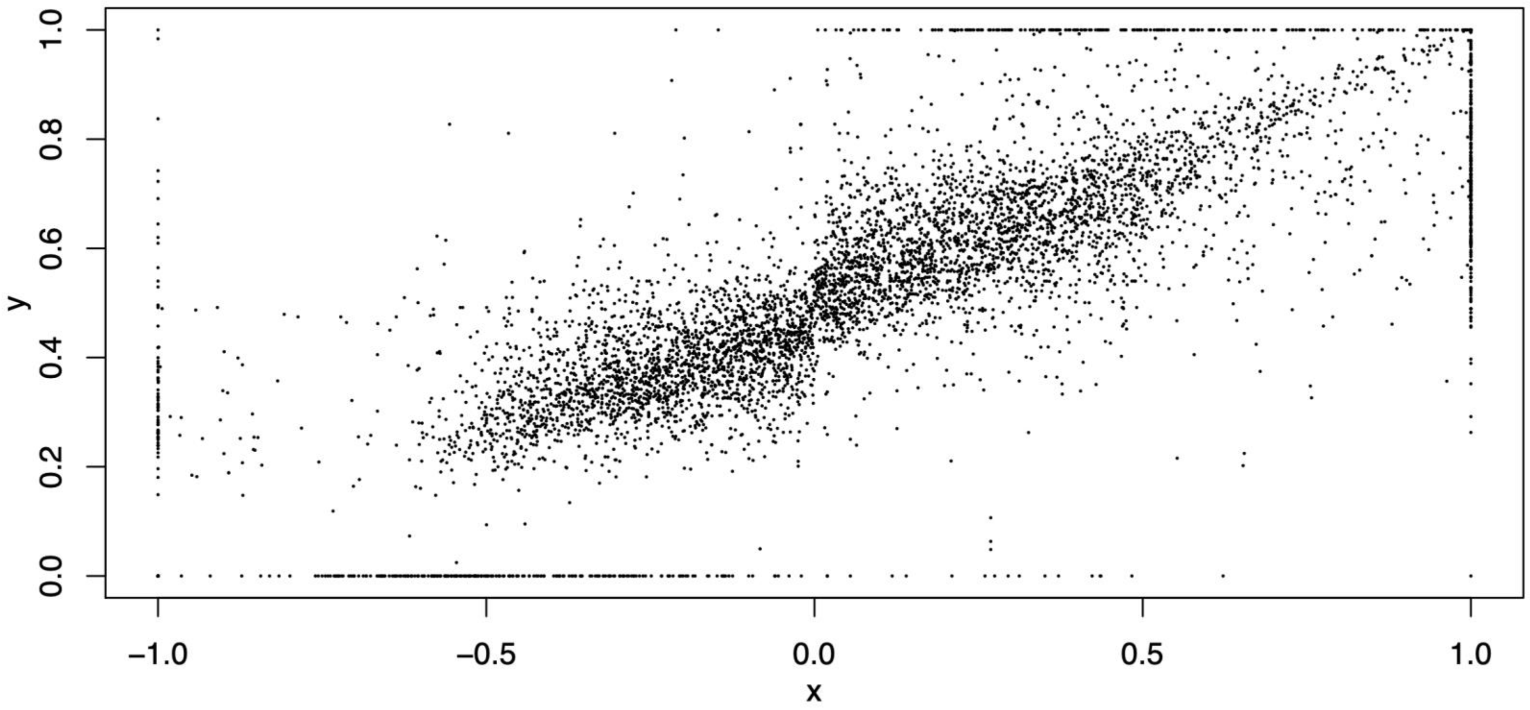
- Thus, naive comparisons of incumbents and non-incumbents are likely confounded.

## ➢ Variables

| | |
|---|---|
| **Running Variable** | Vote margin in the previous election, centered at 0 |
| **Cutoff** | 0 — determines whether the party holds the seat (i.e., becomes the incumbent) |
| **Treatment** | Indicator for whether the current candidate is from the incumbent party |
| **Outcome** | Vote share in the current election |
| **Unit of Analysis** | Congressional district |

# RDD in Action: Lee (2008) on Incumbency Advantage

- In Lee (2008), the author analyzes U.S. House elections from 1946 to 1998 using a dataset comprising 6,558 observations across 435 congressional districts. Each observation corresponds to a candidate in a specific district-year election.




Raw data of Lee (2008)

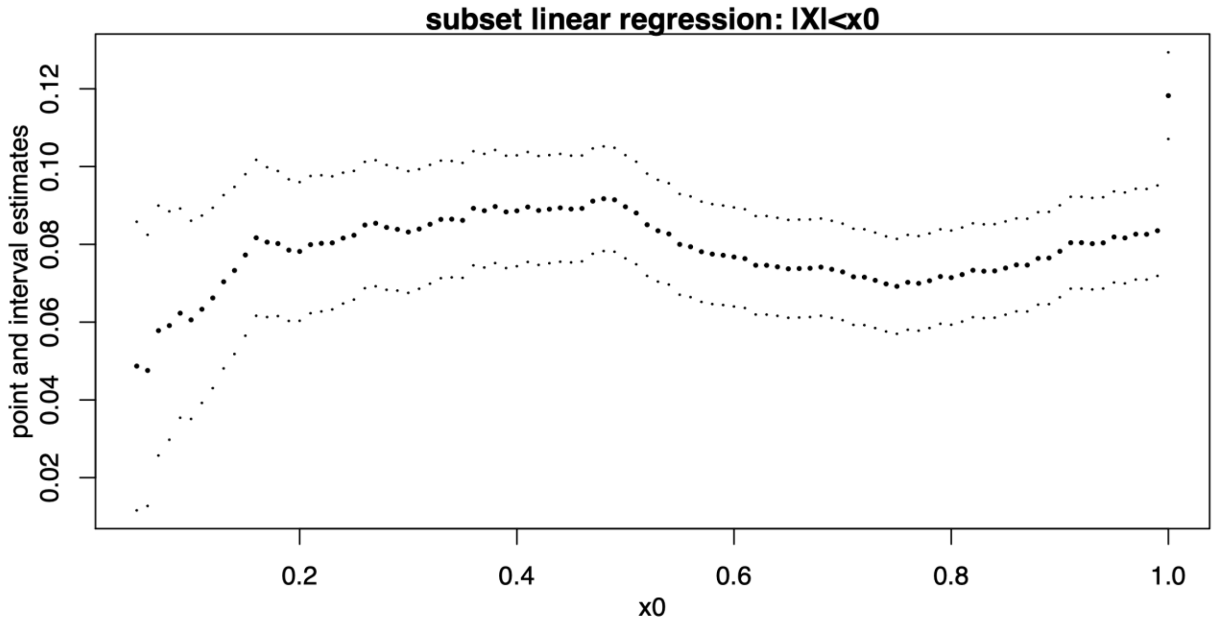
# RDD in Action: Lee (2008) on Incumbency Advantage

✓ Parametric/global method

```
# Load required packages
library(rdrobust)
library(rddtools)

# Load dataset
data(house)

# Estimate treatment effect using rdrobust
RDDest <- rdrobust(house$y, house$x)

# Extract coefficient estimates and confidence intervals
cbind(RDDest$coef, RDDest$ci)
```

| | Coefficient | CI Lower | CI Upper |
|---|---|---|---|
| Conventional | 0.0637 | 0.0422 | 0.0852 |
| Bias-Corrected | 0.0594 | 0.0379 | 0.0808 |
| Robust | 0.0594 | 0.0348 | 0.0839 |

✓ Nonparametric/local method




Estimates based on local linear regressions

# References

D'Amour, A., Ding, P., Feller, A., Lei, L., & Sekhon, J. (2021). Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics*, *221*(2), 644-654.

Jacob, R., Zhu, P., Somers, M. A., & Bloom, H. (2012). A practical guide to regression discontinuity. *MDRC*.

Lee, D. S. (2008). Randomized experiments from non-random selection in US House elections. *Journal of Econometrics*, *142*(2), 675-697.

Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, *51*(6), 309.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: introducing the E-value. *Annals of internal medicine*, *167*(4), 268-274.