# An Introduction to Causal Inference and Applied Microeconometrics for Development Economics

**Organized by Omolola AK-Fasasi (Nee Lipede)**

**Olanrewaju Yusuff**

Department of Economics
Georgia State University

May 24, 2025

# Outline

# The Fundamental Question

$$Y_i = \alpha + \beta X_i + \epsilon_i \tag{1}$$

**What is the effect of X on Y?**

- What is the effect of health insurance on health outcomes?
- What is the effect of free maternal care on child mortality?
- What is the effect of education on wages?
- What is the effect of microcredit on poverty reduction?

# The Fundamental Question

$$Y_i = \alpha + \beta X_i + \epsilon_i \tag{1}$$

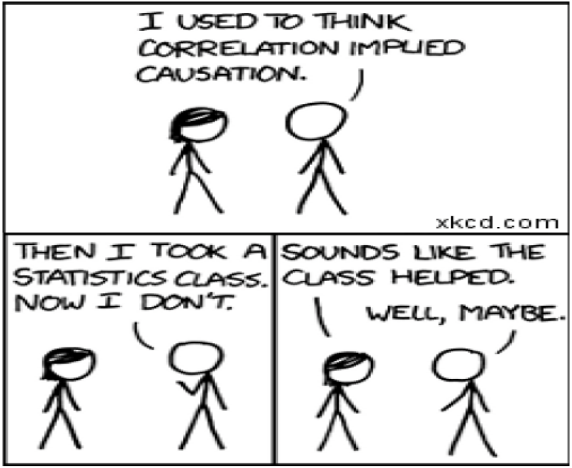
**What is the effect of X on Y?**

- What is the effect of health insurance on health outcomes?
- What is the effect of free maternal care on child mortality?
- What is the effect of education on wages?
- What is the effect of microcredit on poverty reduction?

## Correlation vs. Causation

- Regression can show correlation, not causation
- $\beta$ captures association, not causal effect
- Problem: Other factors affect Y besides X
- Violation of *ceteris paribus*
- Confounders lurk in the error term
- Need to account for these to get causal interpretation See Causal Graph

# Objective of the Workshop




*The classic correlation vs. causation dilemma*

# The Credibility Revolution in Economics

- Economics has undergone a methodological shift over the past 30 years
- Emphasis on research designs that can identify causal effects
- Movement away from structural modeling toward design-based approaches
- Pioneers: Card, Angrist, Imbens, Rubin, Heckman, among others
- The 2021 Nobel Prize in Economics awarded to Card, Angrist, and Imbens for methodological contributions to causal inference

*"Empirical work in economics is now more credible than it was in the past." (Angrist & Pischke, 2010)*

# Why Causal Inference Matters for Development Economics

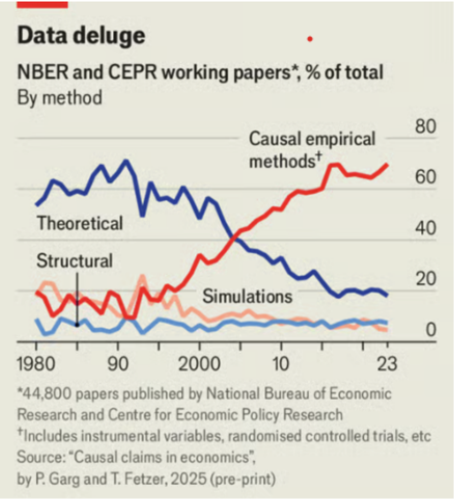



CHART: THE ECONOMIST

*Source: The Economist*

- Development economics focuses on policy interventions
- Limited resources require effective allocation
- Robust evidence needed for policy recommendations
- International organizations increasingly demand rigorous evaluation
- Special challenges in developing country contexts:
  - Data limitations
  - Institutional constraints
  - External validity concerns

# The Potential Outcomes Framework


*Source: Marginal Revolution University*

- Developed by Rubin (1974); also called the Rubin Causal Model
- Central concept: **Potential outcomes**
  - $Y_i(1)$ = outcome for unit $i$ if treated
  - $Y_i(0)$ = outcome for unit $i$ if not treated $\Rightarrow$ **The Counterfactual**
- The causal effect for unit $i$ is: $\tau_i = Y_i(1) - Y_i(0)$
- **Fundamental problem of causal inference:** We never observe both potential outcomes for the same unit
- So, how can we observe the counterfactual outcome?

# Observing the Counterfactual: The Clone Example

**The Ideal Experiment:**

- Imagine we could clone a person
- Send one clone to university, keep the other out
- Compare their wages 10 years later
- The difference = pure causal effect of education

# Observing the Counterfactual: The Clone Example

**The Ideal Experiment:**

- Imagine we could clone a person
- Send one clone to university, keep the other out
- Compare their wages 10 years later
- The difference = pure causal effect of education

**Why this works:**

- Both clones have identical:
  - Ability
  - Family background
  - Motivation
  - All other characteristics
- Only difference: education treatment

# Observing the Counterfactual: The Clone Example

## The Ideal Experiment:

- Imagine we could clone a person
- Send one clone to university, keep the other out
- Compare their wages 10 years later
- The difference = pure causal effect of education

## Why this works:

- Both clones have identical:
  - Ability
  - Family background
  - Motivation
  - All other characteristics
- Only difference: education treatment

## Ceteris Paribus in Causal Inference:

- To identify causal effects, we need to hold everything else constant
- Change only the treatment variable
- This ensures any **difference** in outcome is due to **treatment alone**

# Observing the Counterfactual: The Clone Example

**The Ideal Experiment:**

- Imagine we could clone a person
- Send one clone to university, keep the other out
- Compare their wages 10 years later
- The difference = pure causal effect of education

**Why this works:**

- Both clones have identical:
  - Ability
  - Family background
  - Motivation
  - All other characteristics
- Only difference: education treatment

**Ceteris Paribus in Causal Inference:**

- To identify causal effects, we need to hold everything else constant
- Change only the treatment variable
- This ensures any **difference** in outcome is due to **treatment alone**

**The Challenge:**

- We can't actually clone people
- We never observe both $Y_i(1)$ and $Y_i(0)$ for the same person
- Need methods that mimic this ideal
- Goal: Find "statistical twins" or create **comparable groups**

# Average Treatment Effects

Since we can't observe individual causal effects, we focus on average effects:

- **Average Treatment Effect (ATE)**:

$$\text{ATE} = E[Y_i(1) - Y_i(0)] \tag{2}$$

- **Average Treatment Effect on the Treated (ATT)**:

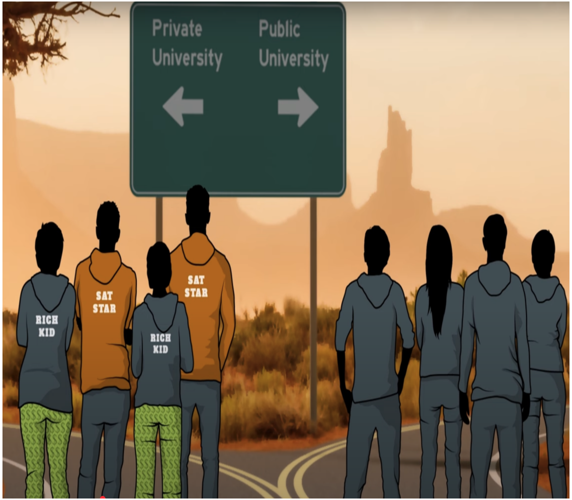
$$\text{ATT} = E[Y_i(1) - Y_i(0) | D_i = 1] \tag{3}$$

- **Average Treatment Effect on the Untreated (ATU)**:

$$\text{ATU} = E[Y_i(1) - Y_i(0) | D_i = 0] \tag{4}$$

These effects can differ when treatment effects are heterogeneous across units

# The Selection Problem ⇒: Critical to Causal Interpretations!!!




*Source: Marginal Revolution University*

**Selection Bias Decomposition:**

$$E[Y_i|D_i=1]-E[Y_i|D_i=0]=$$
$$\underbrace{E[Y_i(1)-Y_i(0)|D_i=1]}_{\text{ATT}}+$$
$$\underbrace{E[Y_i(0)|D_i=1]-E[Y_i(0)|D_i=0]}_{\text{Selection Bias}}$$

- Simple comparison of treated and untreated includes selection bias
- Example: Returns to education
  - Those who choose more education may be inherently different
  - Observed difference includes both education effect and underlying differences

# Understanding Selection Bias in Simple Terms

**Selection Bias:** $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$

**What does this measure?**

- "How different would the treated and untreated groups be, even if neither got treatment?"
- $E[Y_i(0)|D_i = 1]$ = What would happen to the **treated group** if they **didn't get treatment**
- $E[Y_i(0)|D_i = 0]$ = What actually happens to the **untreated group**

# Understanding Selection Bias in Simple Terms

$$\textbf{Selection Bias: } E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

**What does this measure?**

- "How different would the treated and untreated groups be, even if neither got treatment?"
- $E[Y_i(0)|D_i = 1]$ = What would happen to the **treated group** if they **didn't get treatment**
- $E[Y_i(0)|D_i = 0]$ = What actually happens to the **untreated group**

**The Problem:**

- If these are different, groups were already different before treatment
- Simple comparison includes both treatment effect AND pre-existing differences

# Understanding Selection Bias in Simple Terms

**Selection Bias:** $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$

**What does this measure?**

- "How different would the treated and untreated groups be, even if neither got treatment?"
- $E[Y_i(0)|D_i = 1]$ = What would happen to the **treated group** if they **didn't get treatment**
- $E[Y_i(0)|D_i = 0]$ = What actually happens to the **untreated group**

**The Problem:**

- If these are different, groups were already different before treatment
- Simple comparison includes both treatment effect AND pre-existing differences

**Education Example:**

- Treated: People who went to college
- Untreated: People who didn't go to college
- Selection bias asks: "Even if college students had NOT gone to college, would they still earn more?"
- If yes (due to ability, motivation, family connections), then selection bias ¿ 0

**Solution:** Use methods that eliminate selection bias (RCTs, DiD, RDD, IV, etc.)

The $1 Million Question in Economics:

# How do we move from CORRELATION to CAUSATION?

*Solving the selection bias puzzle...*

⇒ **Next: Our toolkit for causal identification**

# Randomized Controlled Trials: "The Experimental Gold Standard"

## Randomized Controlled Trials (RCTs)

- Random assignment ensures treatment is independent of potential outcomes
- Selection bias is eliminated by design
- Pioneered in development by J-PAL and others
- 2019 Nobel Prize to Banerjee, Duflo, Kremer for RCTs in development economics

## Simple regression with RCT: See Randomization Process

$$Y_i = \alpha + \beta T_i + \epsilon_i \tag{5}$$

- $\beta$ = Average Treatment Effect (ATE)
- No controls needed! Randomization does the work
- $E[\epsilon_i | T_i] = 0$ by design

# Randomized Controlled Trials: "The Experimental Gold Standard"

## Randomized Controlled Trials (RCTs)

- Random assignment ensures treatment is independent of potential outcomes
- Selection bias is eliminated by design
- Pioneered in development by J-PAL and others
- 2019 Nobel Prize to Banerjee, Duflo, Kremer for RCTs in development economics

## Simple regression with RCT:

See Randomization Process

$$Y_i = \alpha + \beta T_i + \epsilon_i \tag{5}$$

- $\beta$ = Average Treatment Effect (ATE)
- No controls needed! Randomization does the work
- $E[\epsilon_i | T_i] = 0$ by design

## Advantages:

- **Directly addresses selection bias**[a]
- Conceptually clear
- "Gold standard" of evidence
- $\beta$ has causal interpretation

# Randomized Controlled Trials: "The Experimental Gold Standard"

## Randomized Controlled Trials (RCTs)

- Random assignment ensures treatment is independent of potential outcomes
- Selection bias is eliminated by design
- Pioneered in development by J-PAL and others
- 2019 Nobel Prize to Banerjee, Duflo, Kremer for RCTs in development economics

## Simple regression with RCT:

See Randomization Process

$$Y_i = \alpha + \beta T_i + \epsilon_i \tag{5}$$

- $\beta$ = Average Treatment Effect (ATE)
- No controls needed! Randomization does the work
- $E[\epsilon_i | T_i] = 0$ by design

## Advantages:

- **Directly addresses selection bias**[a]
- Conceptually clear
- "Gold standard" of evidence
- $\beta$ has causal interpretation

## Challenges:

- Cost and logistics
- External validity
- Ethical considerations
- Political constraints

***Randomization ⇒ Selection bias = 0***

[a] Balance of covariates

Balance Table Example

# Quasi-Experimental Methods ⇒ When randomization is not possible

1. **Regression with Controls & Fixed Effects**
    - Control for observable confounders
    - Use fixed effects to control unobservables
    - Relies on selection on observables assumption
2. **Difference-in-Differences**
    - Exploits variation in timing of treatment
    - Compares before-after changes
    - Key assumption: Parallel trends
3. **Regression Discontinuity**
    - Exploits threshold rules in treatment
    - Compares units just above and below threshold
    - Local causal effects at the cutoff
4. **Instrumental Variables**
    - Uses source of exogenous variation in treatment
    - IV must satisfy exclusion restriction
    - Estimates LATE for "compliers"
5. **Synthetic Control Methods**
    - Creates artificial comparison group
    - Useful for single treated unit
    - Becoming popular for policy evaluation

# Regression with Controls: The Traditional Approach

**The Strategy:**

$$Y_{ist} = \alpha + \beta T_{ist} + X'_{ist}\gamma + \mu_i + \lambda_t + \delta_s + \epsilon_{ist} \tag{6}$$

**How it works:**

- Include control variables $X_{ist}$
- Add multiple fixed effects:
  - $\mu_i$: Individual fixed effects
  - $\lambda_t$: Time fixed effects
  - $\delta_s$: State/region fixed effects
- Balance covariates between groups
- Assume $E[\epsilon_{ist}|T_{ist}, X_{ist}] = 0$

# Regression with Controls: The Traditional Approach

**The Strategy:**

$$Y_{ist} = \alpha + \beta T_{ist} + X'_{ist}\gamma + \mu_i + \lambda_t + \delta_s + \epsilon_{ist} \tag{6}$$

**How it works:**

- Include control variables $X_{ist}$
- Add multiple fixed effects:
  - $\mu_i$: Individual fixed effects
  - $\lambda_t$: Time fixed effects
  - $\delta_s$: State/region fixed effects
- Balance covariates between groups
- Assume $E[\epsilon_{ist}|T_{ist}, X_{ist}] = 0$

**The Problem:**

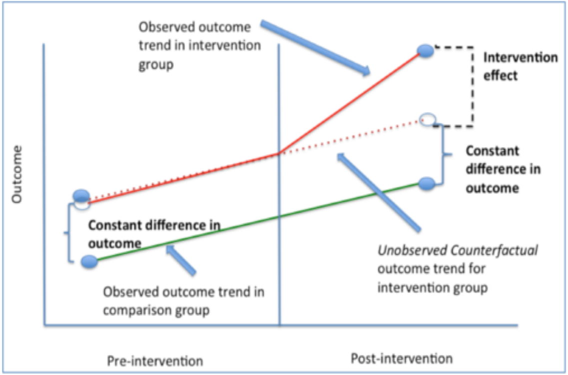
- **Omitted Variable Bias**
- Cannot control for unobservables
- Cannot control for time-varying unobservables

**Education Example:**

- Can control: income, test scores, state, year
- Cannot control: ambition, ability, motivation changes over time

**Why other methods perform better:** They exploit variation that mimics randomization

# Difference-in-Differences Method




*Source: Berkeley Economic Review*
*DiD Visualization of parallel trends assumption*

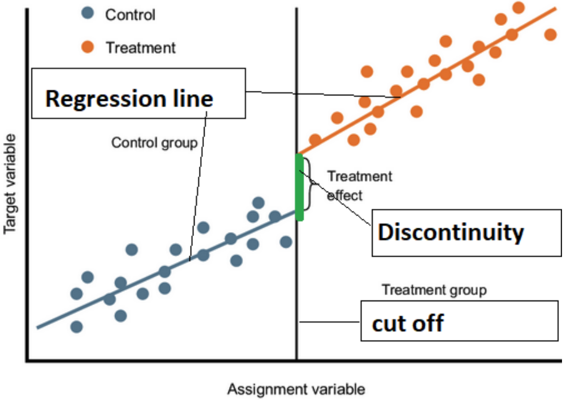
$$Y_{it} = \alpha + \beta \text{Treat}_i + \gamma \text{Post}_t + \delta(\text{Treat}_i \times \text{Post}_t) + \epsilon_{it}$$

- $\delta$ is the DiD estimator (causal effect)
- Controls for time-invariant differences between groups and common time trends
- Increasingly used with multiple time periods and staggered adoption
- Recent advances: Sun & Abraham (2021), Callaway & Sant'Anna (2021), de Chaisemartin & D'Haultfœuille (2020) Estimators

# Regression Discontinuity Design (RDD)

$$Y_i = \alpha + \beta T_i + f(X_i - c) + \epsilon_i$$

- $X_i$: "Running variable" or "score"
- $c$: Cutoff value
- $T_i$: Treatment indicator (1 if $X_i \geq c$)
- $f()$: Flexible function




*Visualization of a sharp RDD*

Key assumption: Units cannot precisely manipulate their position around the threshold

# Instrumental Variables (IV)

**The Problem: Endogeneity**

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

But $Cov(D_i, \epsilon_i) \neq 0$ due to:

- Omitted variables, Reverse causality & Measurement error

**The Solution: Two-stage least squares (2SLS)**

$$\text{First stage:} D_i = \alpha_0 + \alpha_1 Z_i + \eta_i$$

$$\text{Second stage:} Y_i = \beta_0 + \beta_1 \hat{D}_i + \varepsilon_i$$

**Valid instrument must satisfy:**

- Relevance: $Cov(Z_i, D_i) \neq 0$
- Exclusion: $Z_i$ affects $Y_i$ only through $D_i$
- Independence: $Z_i$ is as good as randomly assigned

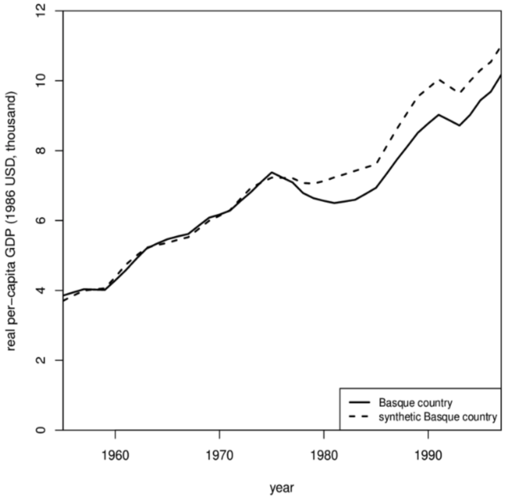
**Classic examples in development:**

- Distance to school as IV for education
- Rainfall shocks as IV for income
- Policy eligibility as IV for program participation
- Historical institutions as IV for current institutions

**Key insight:**

- Use variation in $Z_i$ that affects $D_i$ but not $Y_i$ directly
- This "purges" the endogenous part of $D_i$

# Synthetic Control Method




*Basque Country vs. Synthetic Control*

**Key Features:**

- Developed by Abadie & Gardeazabal (2003) and Abadie et al. (2010)
- Creates artificial control unit as weighted combination of donor units
- Particularly useful for comparative case studies with single treated unit
- Transparent and data-driven approach

**Applications:**

- Economic effects of terrorism (Basque Country)
- Policy evaluation at aggregate level
- Increasingly popular in development economics

# Statistical Inference in Causal Analysis

**The Challenge: Sampling Variance**

- Our sample is just one of many possible samples
- Treatment effect estimate $\hat{\beta}$ varies across samples
- Need to quantify uncertainty around our estimate

**Standard Errors:**

- Measure precision of our estimate
- Smaller SE = more precise estimate
- Affected by sample size, variance in data

**Decision Rule for Significance:**

$$\text{Significant if: } SE(\hat{\beta}) < (\frac{1}{2}) \times |\hat{\beta}|$$

- Rule of thumb: Standard error should be less than half the coefficient
- Equivalent to $|t| > 2$ (approximately p $<$ 0.05)
- $t$-statistic: $t = \frac{\hat{\beta}}{SE(\hat{\beta})}$

Now let's see these methods in action...

# EMPIRICAL APPLICATION

**Free Maternal & Child Healthcare Program in Nigeria**

*A Difference-in-Differences Analysis*

Singh & Yusuff (2025)

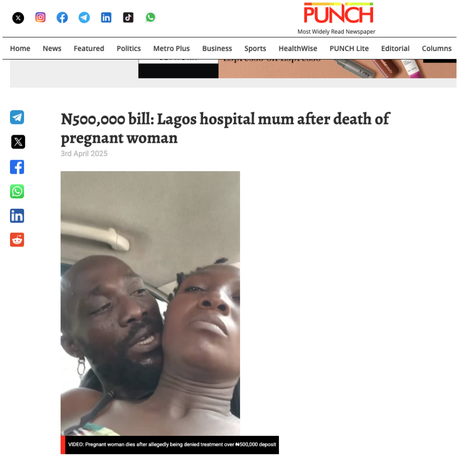
# Unlocking Health Potential: Health Effects of Free Maternal and Child Health Care Program (Singh & Yusuff 2025)

## Motivation & Background

- **Death related to pregnancy or childbirth** $\Rightarrow$ every two minutes; 70% in Sub-Saharan Africa.
- User fee elimination programs gaining attention
- Low vaccination rates major contributor

## Research Questions

- Does eliminating user fee $\Delta$ child mortality?
- Policy effects across subpopulations?
- Mechanisms?
- Cost per death averted?


PUNCH
Most Widely Read Newspaper
Home News Featured Politics Metro Plus Business Sports HealthWise PUNCH Lite Editorial Columns

N500,000 bill: Lagos hospital mum after death of pregnant woman
3rd April 2025

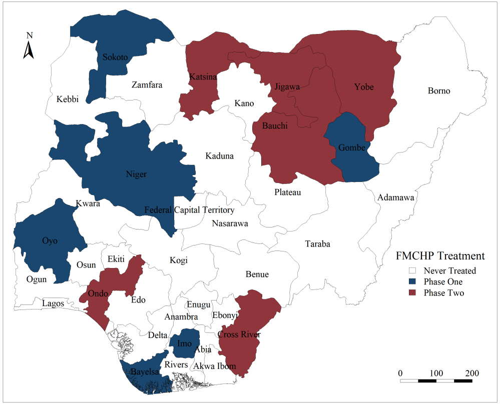
VIDEO: Pregnant woman dies after allegedly being denied treatment over ₦500,000 deposit

*Source: Punch (April 2025)*

# Free Maternal and Child Health Program (FMCHP)

- Implemented under WB's HIPC initiative
- Staggered roll out: Phase One (10/2008) and Two (12/2009)
- 1.2 million births covered
- All primary healthcare services covered
- Mothers: pregnancy to six weeks post-delivery
- Children: birth to five years

**Figure:** Treatment vs Control



# Empirical Strategy and Data

## Data Sources

- DHS: 2008, 2013, 2018
- Treatment info: USAID reports
- Health Facility Registry

## Identification Strategy

- Mother fixed-effects
- Month-year of birth fixed-effects
- Controls: birth order, multiple birth, sex

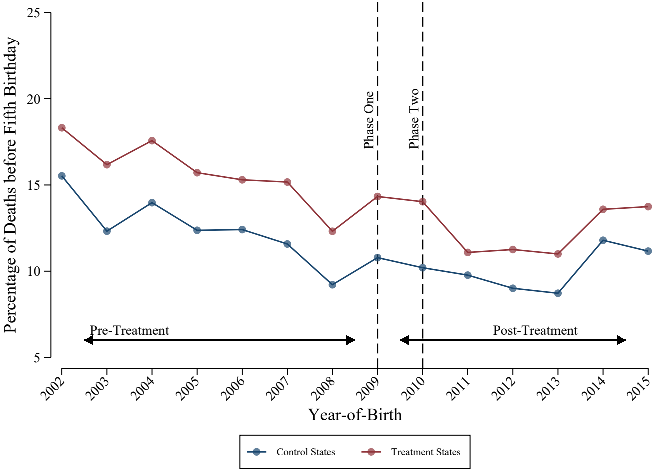
$$y_b = \alpha_{m(b)} + \alpha_{MoB(b),YoB(b)} + \beta \left[ \mathbb{1}\left\{ Treat_{s(b)} \right\} \times \mathbb{1}\left\{ Post_{MoB(b),YoB(b)} \right\} \right] + \mathbf{X}_b \gamma + \epsilon_b$$

## Identification Checks

- No fertility anticipation
- No composition changes
- Parallel pre-trends
- No SUTVA violations
- No supply side changes

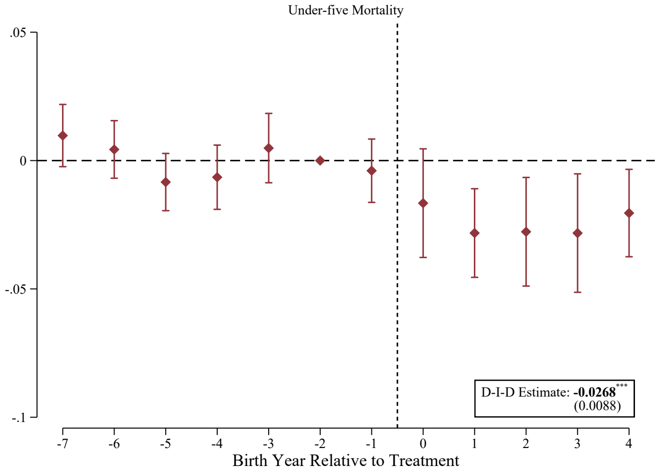
# Paper in One Figure: The Parallel Trends Assuption Check

**Figure:** Temporal Variation in Deaths before Fifth Birthday



# Main Result: Event-study Estimates

**Figure:** Event-study Estimates: Under-five Mortality



# Heterogeneity and Mechanisms

**Heterogeneity:** Stronger effects for:

- Northern region households
- Islamic households
- Poorer households
- Areas with better healthcare access

**Mechanisms**

- Increased vaccination rates
- More institutional deliveries
- More prenatal care visits
- Reduced financial barriers
- Improved medical trust

# Quantifying the Costs and Benefits

- FMCHP reached $\sim$ 1.2 million children
- Reduction in under-five mortality: 0.02683 (26.8%)
- Total deaths averted: 32, 196 children
- Program cost: 13.2 billion Naira ($82.6 million)
- Cost per death averted: $2, 565.23
- This equals 53.7% of average annual household expenditure

# Conclusion

- User fee elimination significantly reduced under-five mortality
- Reduced socioeconomic inequities in health outcomes
- Key mechanism: increased preventive care (vaccinations)
- Cost-effective: $2,565.23 per child life saved
- Policy implications: healthcare access crucial for effectiveness of financial barrier removal

**Link to the paper** ⇒ [https://osf.io/preprints/osf/y6wzt_v2]

# Challenges and Opportunities in Development Context

**Data Challenges:**

- Limited administrative data
- Sample selection and attrition
- Measurement error
- Lack of baseline data
- Difficulty tracking migrants

**Implementation Challenges:**

- Political constraints
- Ethical considerations
- Compliance and take-up issues
- Spillover effects

**Opportunities:**

- Natural policy variation across regions
- Increasing digitization of government services
- Mobile phone and satellite data
- Growing interest in evidence-based policy
- Partnerships with governments and NGOs
- Innovative data collection methods

# External Validity and Policy Relevance

- **External validity concerns:**
  - Will findings generalize to other contexts?
  - Tension between internal and external validity
  - Importance of understanding mechanisms
- **Making research policy-relevant:**
  - Engaging with policymakers early
  - Addressing questions of policy interest
  - Considering cost-effectiveness
  - Communicating findings clearly
  - Building local research capacity
- **Scaling up successful interventions:**
  - From efficacy to effectiveness
  - Accounting for general equilibrium effects
  - Considering political economy constraints

# Future Directions in Causal Inference for Development

- **Methodological innovations:**
  - Machine learning for heterogeneous treatment effects
  - Improved methods for external validity
  - Better approaches for interference and spillovers
  - Combining structural and reduced-form approaches
- **Data innovations:**
  - Remote sensing and satellite data
  - Administrative data linkages
  - Mobile phone and digital traces
  - High-frequency monitoring data
- **Capacity building:**
  - Training local researchers
  - Building data infrastructure
  - Creating collaborative networks
  - Democratizing access to methods and tools

# Key Takeaways

1. Causal inference is critical for effective development policy
2. Multiple identification strategies are available, each with strengths and limitations
3. Careful attention to context matters for both internal and external validity
4. Developing country settings present unique challenges and opportunities
5. The Free Maternal and Child Health Program in Nigeria demonstrates how causal methods can inform policy
6. Building local research capacity is essential for sustainable evidence-based policy

*"The goal of causal inference in development economics is not just methodological purity, but to improve the lives of the poor through better policy."*

# Recommended Resources

**Introductory Resources:**

- Angrist & Pischke (2009). *Mostly Harmless Econometrics*
- Cunningham (2021). *Causal Inference: The Mixtape*
- World Bank's Impact Evaluation Series

**Online Resources:**

- J-PAL's Evaluation Resources
- World Bank's Development Impact Blog
- GitHub repositories of code for causal inference
- Causal Inference Mixtape website

# Thank You!

Thank you for your attention!

Questions and comments are welcome
Please reach out at

oyusuff4@student.gsu.edu

olanrewajuyusuffecon.github.io/website

# Appendix: Balance of Covariates Example

**TABLE 2 Summary statistics and covariate balance.**

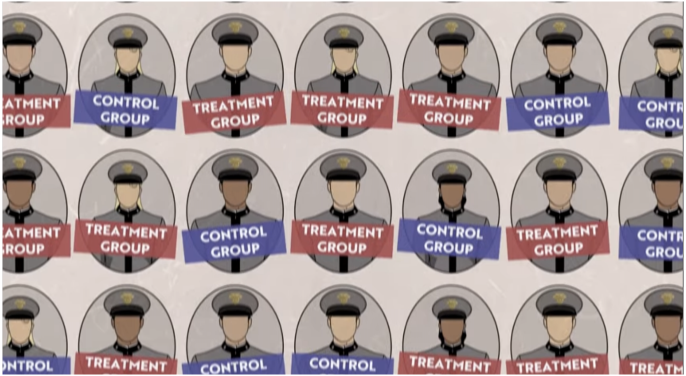
| | Control (1) | Treatment (laptops/tablets) (2) |
|---|---|---|
| A. BASELINE CHARACTERISTICS | | |
| Female | 0.17 | 0.20 |
| White | 0.64 | 0.67 |
| Black | 0.11 | 0.10 |
| Hispanic | 0.13 | 0.13 |
| Age | 20.12 [1.06] | 20.15 [1.00] |
| Prior military service | 0.19 | 0.19 |
| Division I athlete | 0.29 | 0.40 |
| GPA at baseline | 2.87 [0.52] | 2.82 [0.54] |
| Composite ACT | 28.78 [3.21] | 28.30 [3.46] |

*Source: Marginal Revolution University*

*Carter, S. P., Greenberg, K., & Walker, M. S. (2017). The impact of computer usage on academic performance*

- Balance table shows baseline characteristics are similar across groups
- No statistically significant differences in pre-treatment variables
- Confirms successful randomization eliminated selection bias

# Appendix: Randomization Process




*Source: Source: Marginal Revolution University*

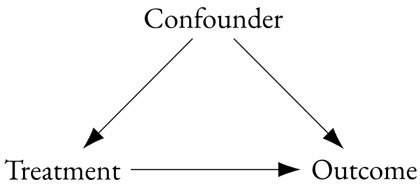
- Random assignment ensures each unit has equal probability of treatment
- Eliminates systematic differences between treatment and control groups
- Creates comparable groups that differ only in treatment status

Back to Main Slides

# Appendix: Common Causal Inference Pitfalls

1. **Controlling for post-treatment variables**
   - Creates “bad controls” problem
   - Can introduce selection bias
2. **Weak instruments**
   - Leads to biased estimates and poor inference
   - Rule of thumb: F-statistic $> 10$
3. **Violated parallel trends assumption**
   - DiD will be biased if trends differ pre-treatment
   - Test with placebo treatments and event studies
4. **Extrapolating beyond data support**
   - RDD only identifies effects at the cutoff
   - IV estimates LATE for compliers only
5. **Ignoring clustering of standard errors**
   - Leads to overstated precision
   - Cluster at the level of treatment assignment

# Appendix: Causal Graphs (DAGs)




Back to Main Slides

- Directed Acyclic Graphs (DAGs) visualize causal relationships
- Help identify:
  - Confounders to control for
  - Colliders to avoid controlling for
  - Valid instrumental variables
  - Mediating pathways