Review of Empirical Methods

Impact Evaluation

Stanislao Maldonado
Universidad del Rosario

Answer these questions

1. What is impact evaluation?
2. Why is impact evaluation valuable?
3. What makes a good impact evaluation?

Answer these questions

1. What is impact evaluation?
2. Why is impact evaluation valuable?
3. What makes a good impact evaluation?

Impact Evaluation

"

An assessment of the causal effect of a project , program or policy on beneficiary outcomes.

Estimates the change in outcomes attributable to the intervention.

Impact Evaluation Answers

What is effect of information on risky sexual behavior and HIV prevalence?

Does contracting out primary health care lead to an increase in access and quality?

Do bonuses to sales people generate more revenue than consumer price discounts?

Do micro loans increase the productivity of small entrepreneurs?

Impact Evaluation Answers

What was the effect of the program on outcomes?

How much better off are the beneficiaries because of the program/policy?

How would outcomes change if changed program design?

Is the program cost-effective?

Traditional M&E cannot answer these.

Answer these questions

1. What is impact evaluation?

2. Why is impact evaluation valuable?

3. What makes a good impact evaluation?

Why Evaluate?

1. Need evidence on what works

   Limited budget and bad policies could hurt

2. Improve program/policy implementation

   - Design (eligibility, benefits)
   - Operations (efficiency & targeting)

3. Information key to sustainability

   - Budget negotiations
   - Informing beliefs and the press
   - Results agenda and Aid effectiveness

Answer these questions

1. Why is evaluation valuable?
2. What makes a good impact evaluation?
3. How to implement an impact evaluation?

How to assess impact

- e.g. How much does an safe water intervention reduce diarrhea?
- What is beneficiary's diarrhea incidence of diarrhea in last 3-days with program compared to without program?
- Formally, program impact is:

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

- Compare same individual with & without programs at same point in time

Solving the evaluation problem

- Counterfactual: what would have happened without the program.
- Estimated impact is difference between treated observation and counterfactual.
- Never observe same individual with and without program at same point in time.
- Need to estimate counterfactual.
- Counterfactual is key to impact evaluation.

Counterfactual Criteria

- Treated & Counterfactual
  (1) Have identical characteristics,
  (2) Except for benefiting from the intervention.
- No other reason for differences in outcomes of treated and counterfactual.
- Only reason for the difference in outcomes is due to the intervention.

2 *Counterfeit* Counterfactuals

Before and After

Same individual before the treatment

Those not enrolled

- Those who choose not to enroll in the program
- Those who were not offered the program

Problem: Cannot completely know why the treated are treated and the others not.

1. Before and After: Example

- Agricultural assistance program
  - Financial assistance to purchase inputs.
  - Compare rice yields before and after.
  - Before is normal rainfall, but after is drought.
  - Find fall in rice yield.
  - Did the program fail?
  - Could not separate (identify) effect of financial assistance program from effect of rainfall.

2. Those not enrolled: Example

- Health insurance offered
- Compare health care utilization of those who got insurance to those who did not
  - Who buys insurance?: those that expect large medical expenditures
  - Who does not?: those who are healthy
- With no insurance: Those that did not buy, have lower medical costs than that did
- Poor estimate of counterfactual

Program placement: example

- Government offers a family planning program to villages with high fertility
- Compare fertility in villages offered program to fertility in other villages
- Program targeted based on fertility, so
  (1) Treatments have high fertility and
  (2) counterfactuals have low fertility.
- Estimated program impact confounded with targeting criteria

What's wrong?

1. Selection bias: People choose to participate for specific reasons

2. Many times reasons are related to the outcome of interest
   - Job Training: ability and earning
   - Health Insurance: health status and medical expenditures

3. Cannot separately identify impact of the program from these other factors/reasons

Need to know…

All the reasons why someone gets the program and others not.

All the reasons why individuals are in the treatment versus control group.

If reasons correlated w/ outcome cannot identify/separate program impact from other explanations of differences in outcomes.

Possible Solutions

- Need to guarantee comparability of treatment and control groups.
- ONLY remaining difference is intervention.
- In this seminar we will consider:
  - Experimental design/randomization
  - Quasi-experiments (Regression Discontinuity, Double differences)
  - Instrumental Variables.

These solutions all involve…

- Knowing how the data are generated.

- Randomization
  - Give all equal chance of being in control or treatment groups
  - Guarantees that all factors/characteristics will be on average equal btw groups
  - Only difference is the intervention

- If not, need transparent & observable criteria for who is offered program.

1

Causal Inference

Counterfactuals

False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled (Apples & Oranges)

Our Objective

Estimate the causal effect (impact) of intervention (P) on outcome (Y).

(P) = Program or Treatment

(Y) = Indicator, Measure of Success

Example: What is the effect of a Cash Transfer Program **(P)** on Household Consumption **(Y)**?

Causal Inference

What is the impact of (P) on (Y)?

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

Can we all go home?

Problem of Missing Data

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

For a program beneficiary:

- we observe
  $(Y \mid P=1)$: Household Food Consumption (Y) with a cash transfer program ($P=1$)

- but we do not observe
  $(Y \mid P=0)$: Food Consumption (Y) without a cash transfer program ($P=0$)

Solution

Estimate what ***would*** have happened to *Y* in the absence of *P*.

We call this the **Counterfactual.**

The key to a good impact evaluation is a valid counterfactual!

Estimating impact of *P* on *Y*

$$\alpha = (Y \mid P=1) - (Y \mid P=0)$$

OBSERVE $(Y \mid P=1)$
Outcome with treatment

ESTIMATE $(Y \mid P=0)$
The Counterfactual

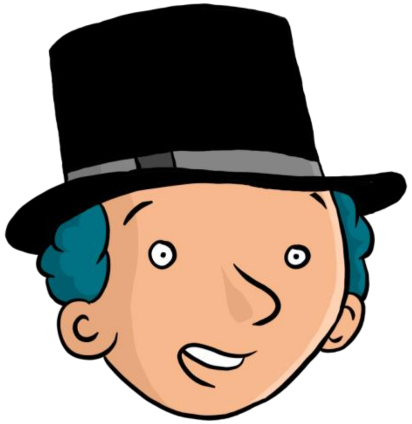
IMPACT = Outcome with treatment **-** counterfactual

- Intention to Treat (**ITT**) – *Those to whom we wanted to give treatment*
- Treatment on the Treated (**TOT**) – *Those actually receiving treatment*

- Use **comparison** or **control** group

Example: What is the Impact of…

giving Fulanito



additional pocket money

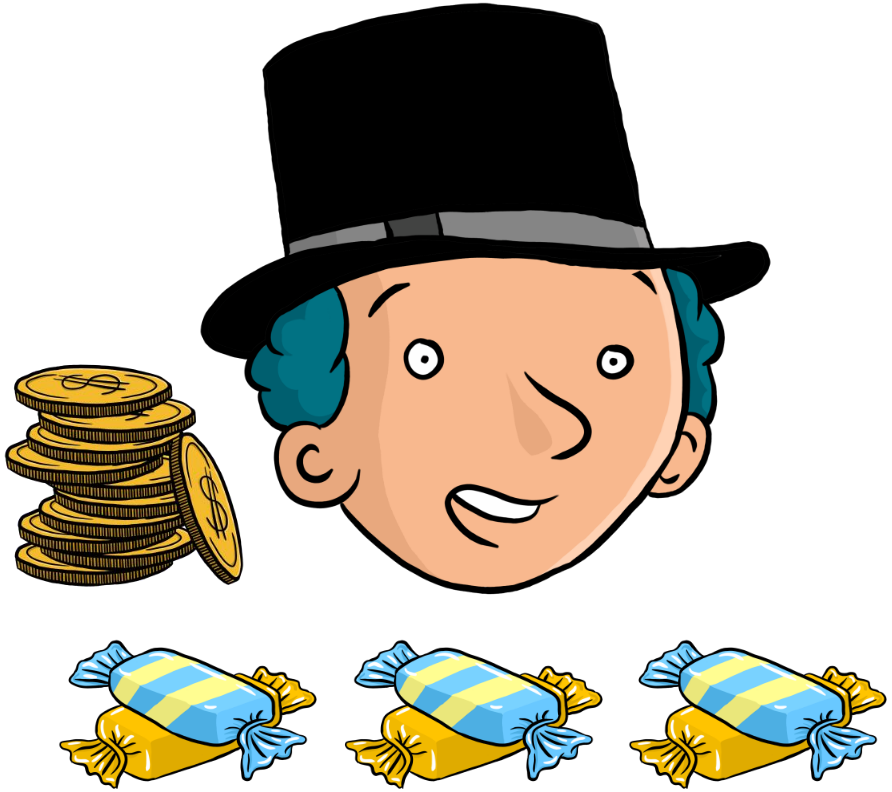
(P)

on Fulanito's consumption of candies

(Y)?

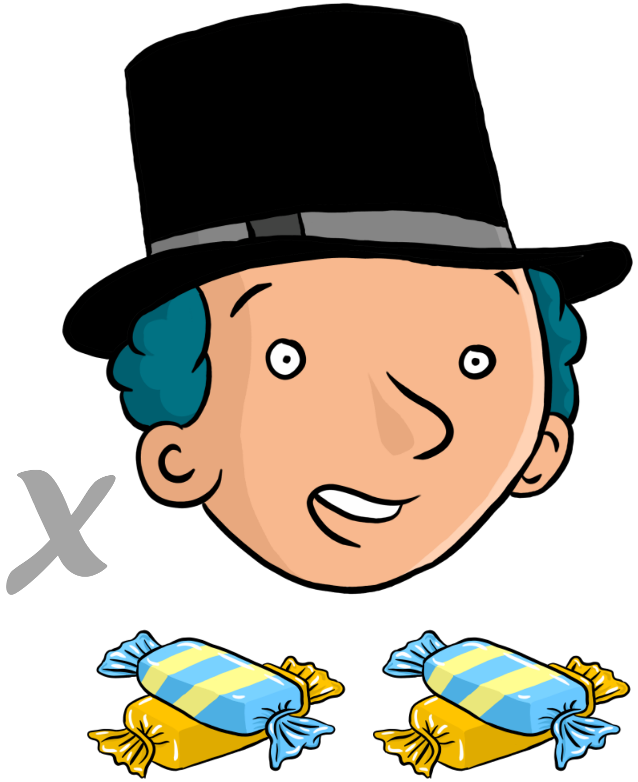
The Perfect *Clone*

Fulanito



6 candies

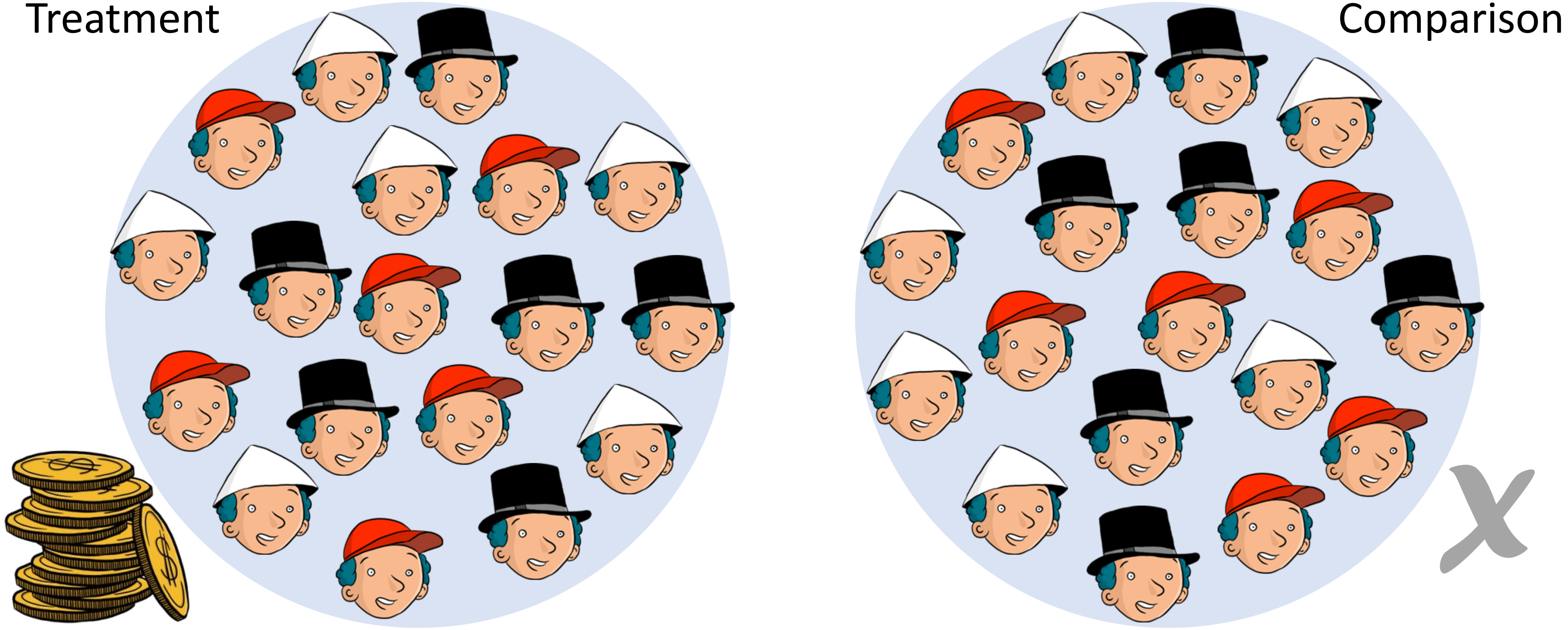
Fulanito's Clone



4 candies

IMPACT=6-4=2 Candies

In reality, use statistics



Average Y=6 candies

Average Y=4 Candies

IMPACT=6-4=2 Candies

Finding good comparison groups

We want to find **clones** for the Fulanitos in our programs.

The treatment and comparison groups should

- have identical characteristics
- except for benefiting from the intervention.

m eligibility & assignment rules
valid counterfactuals

With a good comparison group, the **only reason** for different outcomes between treatments and comparisons is the **intervention** (P)

Case Study: Progresa

- National anti-poverty program in Mexico
  - Started 1997
  - 5 million beneficiaries by 2004
  - Eligibility – based on poverty index

- Cash Transfers
  - Conditional on school and health care attendance.

Case Study: Progresa

- Rigorous impact evaluation with rich data
  - 506 communities, 24,000 households
  - Baseline 1997, follow-up 1998

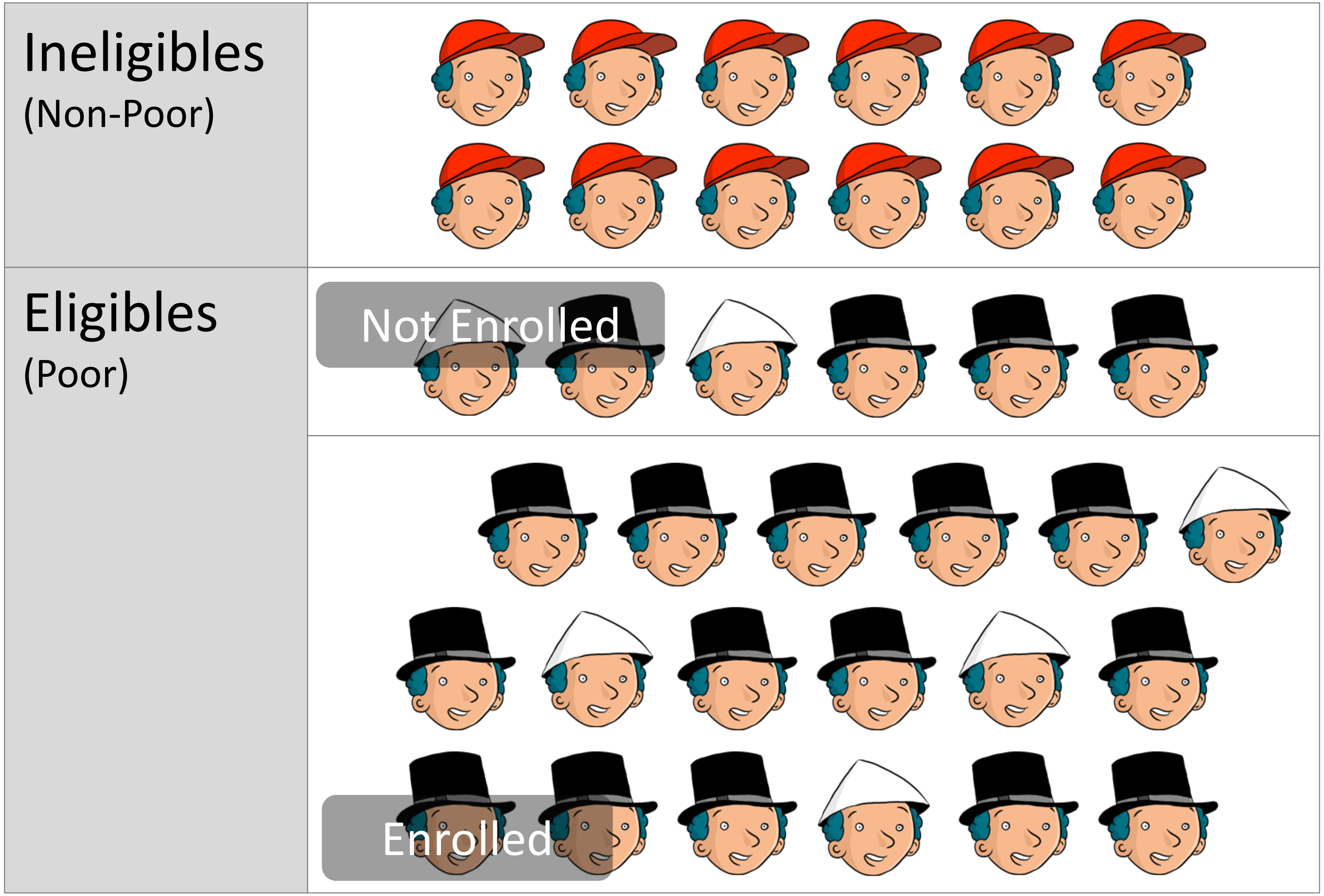
- Many outcomes of interest

  Here: Consumption per capita

- What is the effect of Progresa (P) on Food Consumption Per Capita (Y)?

  If impact is an increase of **$20** or more, then scale up nationally

Eligibility and Enrollment



1

Causal Inference

Counterfactuals

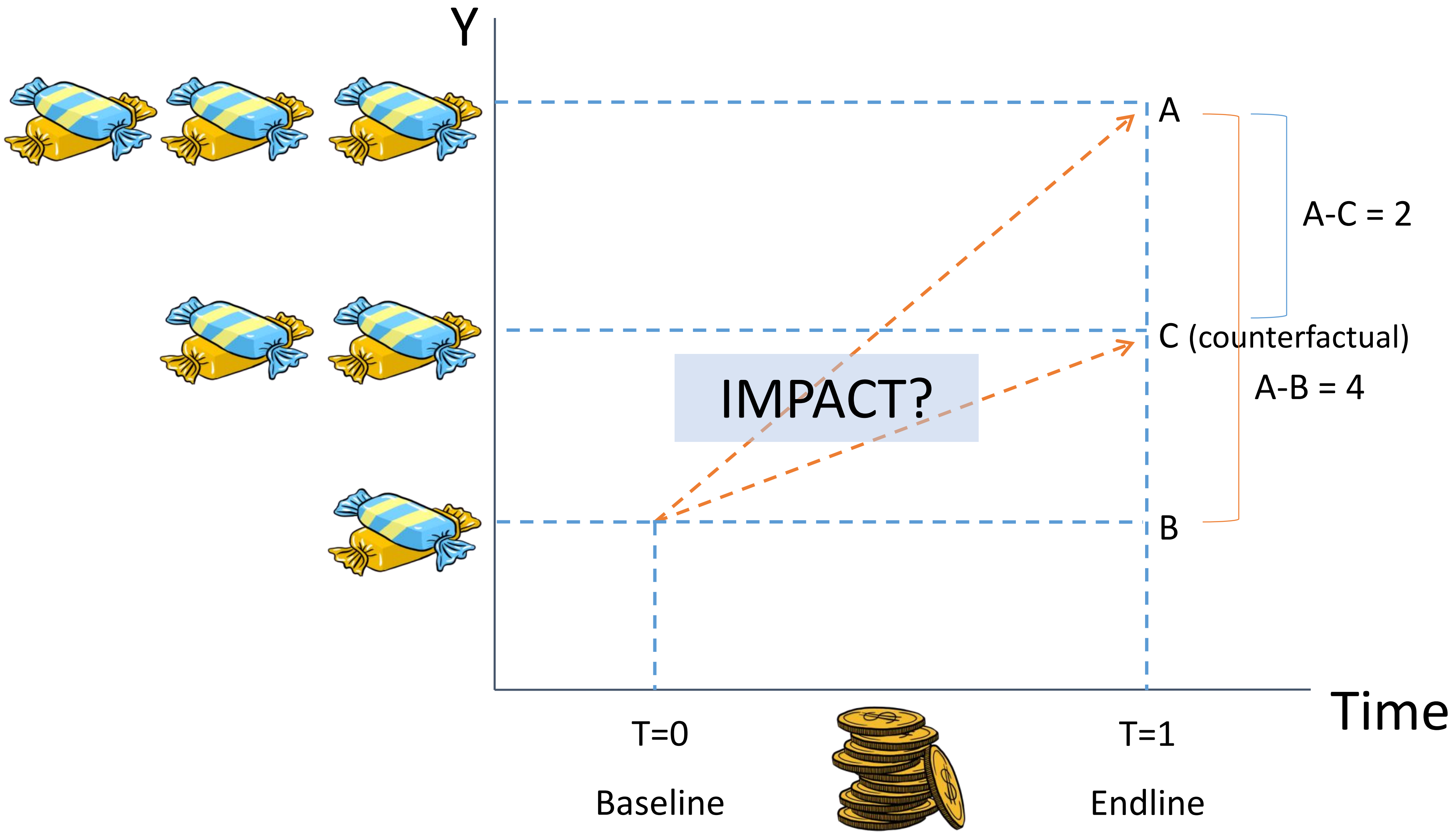
False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled (Apples & Oranges)

False Counterfactual #1

Before & After

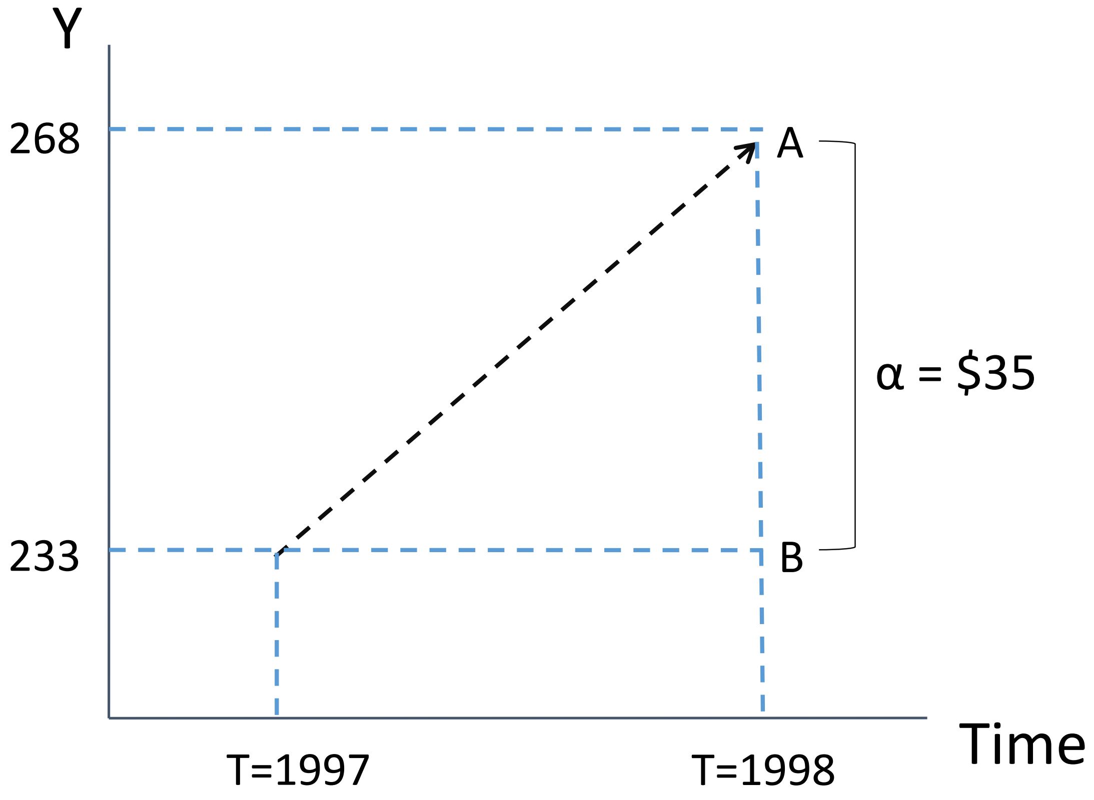


Case 1: Before & After

What is the effect of Progresa (P) on consumption (Y)?

(1) Observe only beneficiaries (P=1)

(2) Two observations in time:
Consumption at T=0 and consumption at T=1.



IMPACT=A-B= $35

Case 1: Before & After

Consumption (Y)	
Outcome with Treatment *(After)*	268.7
Counterfactual *(Before)*	233.4
Impact *(Y | P=1) - (Y | P=0)*	35.3***

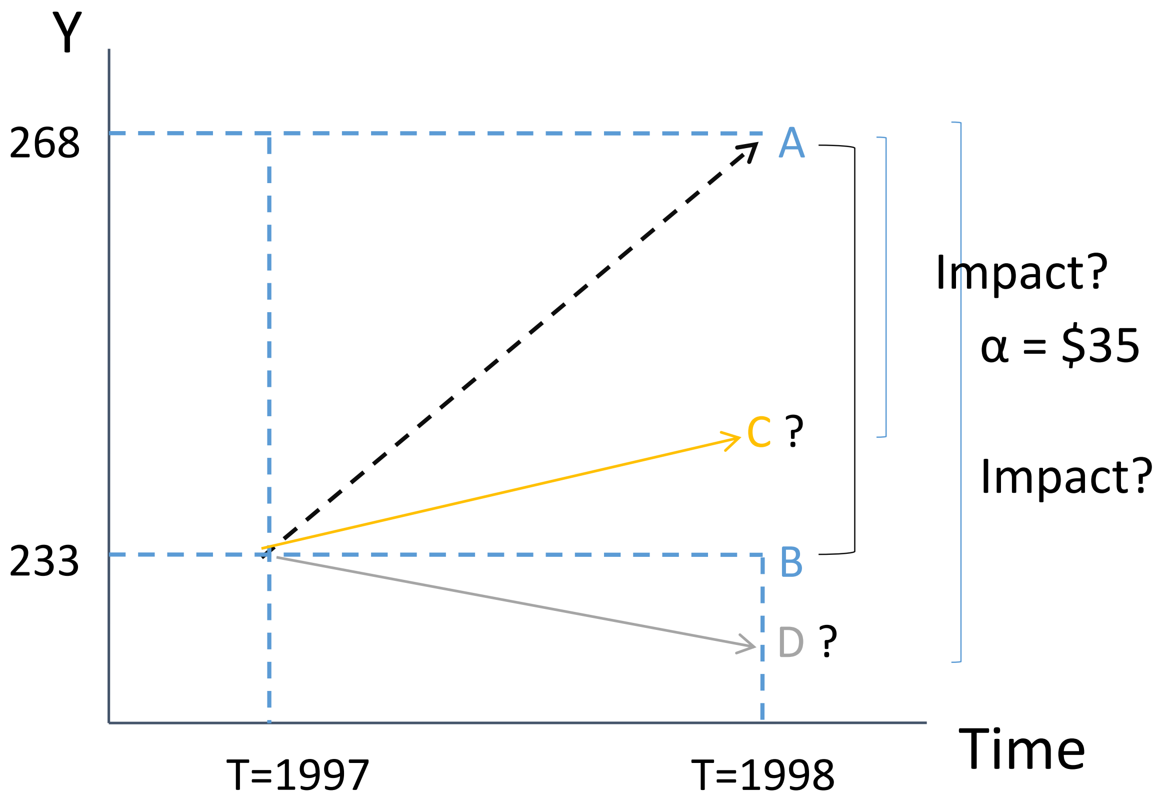
Estimated Impact on Consumption (Y)	
Linear Regression	35.27**
Multivariate Linear Regression	34.28**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Case 1: What's the problem?

Economic Boom:
- Real Impact=A-C
- A-B is an *overestimate*

Economic Recession:
- Real Impact=A-D
- A-B is an *underestimate*



Before & After doesn't control for other time-varying factors!

1

Causal Inference

Counterfactuals

False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled (Apples & Oranges)

False Counterfactual #2

Enrolled & Not Enrolled

- If we have post-treatment data on
  - Enrolled: treatment group
  - Not-enrolled: "comparison" group (counterfactual)
    - *Those ineligible to participate.*
    - *Those that choose NOT to participate.*
- Selection Bias
  - Reason for not enrolling may be correlated with outcome (Y)
    - *Control for observables.*
    - *But not un-observables!*
  - Estimated impact is confounded with other things.

Case 2: Enrolled & Not Enrolled

Measure outcomes in post-treatment (T=1)

In what ways might **E&NE** be different, other than their enrollment in the program?

Case 2: Enrolled & Not Enrolled

Consumption (Y)	
Outcome with Treatment *(Enrolled)*	268
Counterfactual *(Not Enrolled)*	290
Impact *(Y | P=1) - (Y | P=0)*	-22**

Estimated Impact on Consumption (Y)	
Linear Regression	-22**
Multivariate Linear Regression	-4.15

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Progresa Policy Recommendation?

Impact on Consumption (Y)		
Case 1: Before & After	Linear Regression	35.27**
	Multivariate Linear Regression	34.28**
Case 2: Enrolled & Not Enrolled	Linear Regression	-22**
	Multivariate Linear Regression	-4.15

- Will you recommend scaling up Progresa?
- B&A: Are there other time-varying factors that also influence consumption?
- E&BNE:
  - Are reasons for enrolling correlated with consumption?
  - Selection Bias.

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Keep in Mind

B&A

Compare: Same individuals Before and After they receive P.

Problem: Other things may have happened over time.

E&NE

Compare: Group of individuals Enrolled in a program with group that chooses not to enroll.

Problem: Selection Bias. We don't know why they are not enrolled.

Both counterfactuals may lead to biased estimates of the impact.

Choosing your IE method(s)

Choose the **best possible design** given context:

Proof of concept vs scaling up?	Efficacy vs Effectiveness? Who controls intervention? Operational rules/constraint Least operational risk
Have we controlled for all alternative explanations?	Internal validity Causal inference Valid counterfactual estimate Good comparison group
Is the result valid for *everyone*?	External validity Local vs global treatment effect Evaluation results apply to population we're interested in

Choosing your IE method(s)

Key information you will need for identification in an effectiveness study implemented by someone esle:

Prospective/Retrospective Evaluation?

Eligibility rules and criteria?

- Poverty targeting?
- Geographic targeting?
- Lottery?

Roll-out plan (pipeline)?

Is the number of eligible units larger than available resources at a given point in time?

- Budget and capacity constraints?
- Excess demand for program?

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

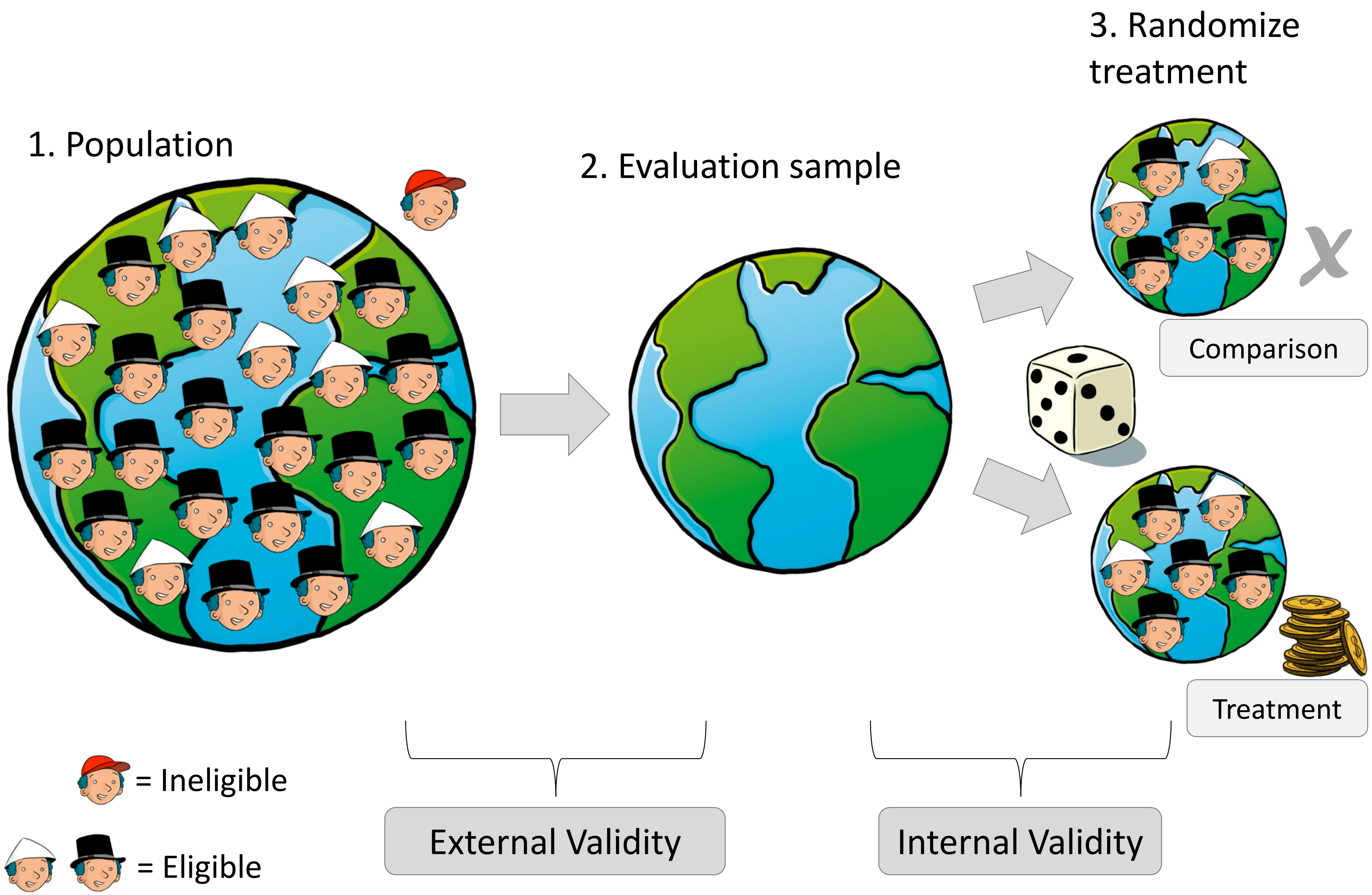
P-Score matching

2

IE Methods

Toolbox

Randomized treatments and comparisons



Unit of Randomization

- Choose according to type of program
  - Individual/Household
  - School/Health Clinic/catchment area
  - Block/Village/Community/Region

As a rule of thumb, randomize at the smallest viable unit of implementation.

- Keep in mind
  - Need "sufficiently large" number of units to detect minimum desired impact: Power.
  - Clustering reduces effective sample size
  - Standard Errors need to be "clustered"
  - Spillovers/contamination
  - Operational and survey costs

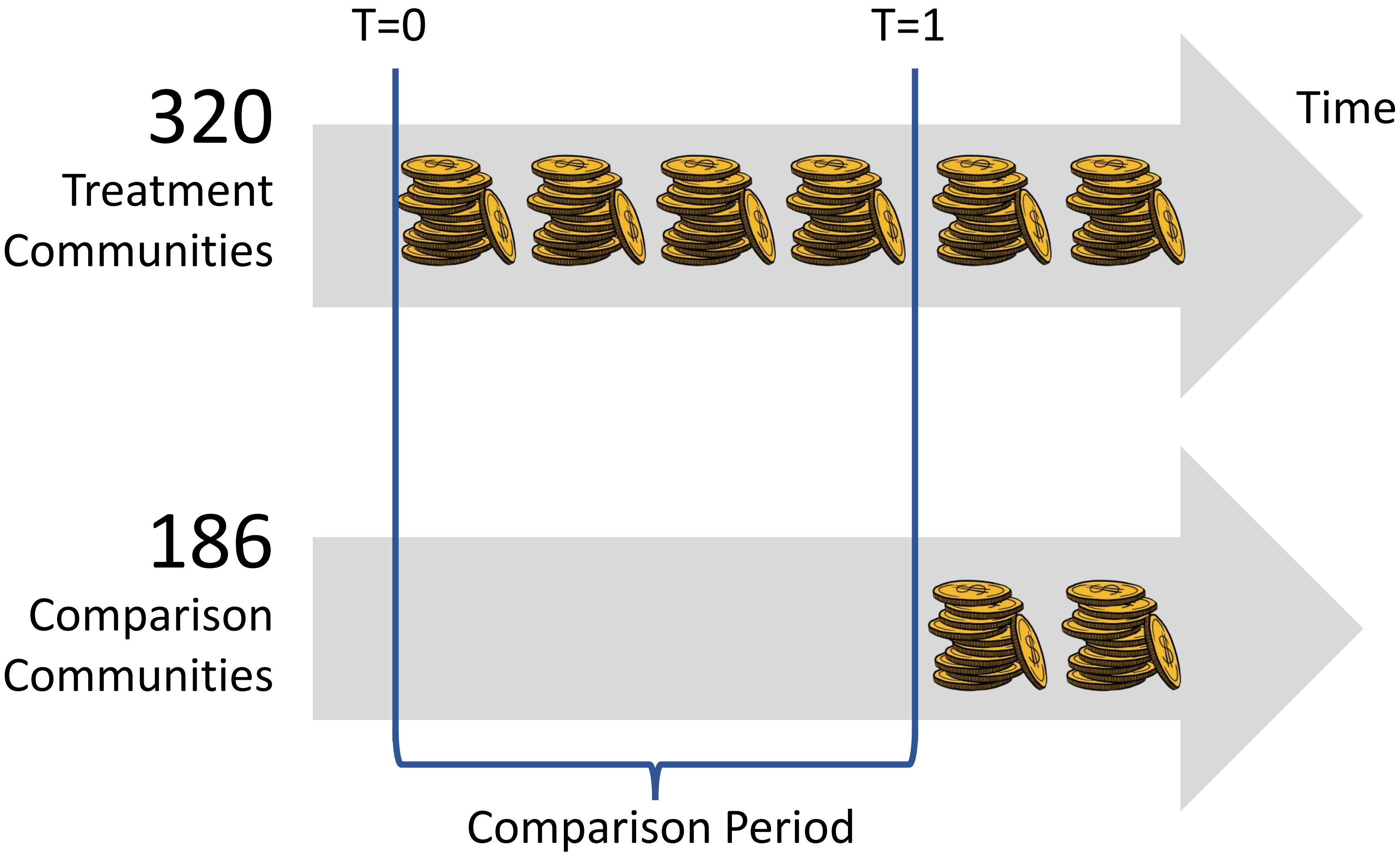
Case 3: Randomized Assignment

- Progresa CCT program
- Unit of randomization: Community
- 506 communities in the evaluation sample
- Randomized phase-in
  - 320 treatment communities (14446 households):

    First transfers in April 1998.
  - 186 comparison communities (9630 households):

    First transfers November 1999

Case 3: Randomized Assignment



Case 3: Randomized Assignment

How do we know we have good clones?

In the absence of Progresa, treatment and comparisons should be identical

Let's compare their characteristics at baseline (T=0)

Case 3: Balance at Baseline

Case 3: Randomized Assignment			
	Treatment	Comparison	*T-stat*
Consumption ($ monthly per capita)	233.4	233.47	-0.39
Head's age (years)	41.6	42.3	-1.2
Spouse's age (years)	36.8	36.8	-0.38
Head's education (years)	2.9	2.8	**2.16** **
Spouse's education (years)	2.7	2.6	0.006

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Case 3: Balance at Baseline

Case 3: Randomized Assignment			
	Treatment	Comparison	*T-stat*
Head is female=1	0.07	0.07	-0.66
Indigenous=1	0.42	0.42	-0.21
Number of household members	5.7	5.7	1.21
Bathroom=1	0.57	0.56	1.04
Hectares of Land	1.67	1.71	-1.35
Distance to Hospital *(km)*	109	106	1.02

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Case 3: Randomized Assignment

	Treatment Group *(Randomized to treatment)*	**Counterfactual** *(Randomized to Comparison)*	**Impact** *(Y | P=1) - (Y | P=0)*
Baseline ***(T=0)*** Consumption (Y)	233.47	233.40	0.07
Follow-up ***(T=1)*** Consumption (Y)	268.75	239.5	29.25**

Estimated Impact on Consumption (Y)	
Linear Regression	29.25**
Multivariate Linear Regression	29.75**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Progresa Policy Recommendation?

Impact of Progresa on Consumption (Y)		
Case 1: Before & After	Multivariate Linear Regression	34.28**
Case 2: Enrolled & Not Enrolled	Linear Regression	-22**
	Multivariate Linear Regression	-4.15
Case 3: Randomized Assignment	Multivariate Linear Regression	29.75**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Real Opportunities for Randomization

Eligibles > Number of Benefits

- Randomize!
- Lottery for who is offered benefits
- Fair, transparent and ethical way to assign benefits to equally deserving populations.

Oversubscription

- Give each eligible unit the same chance of receiving treatment
- Compare those offered treatment with those not offered treatment *(comparisons).*

Randomized Phase In

- Give each eligible unit the same chance of receiving treatment first, second, third…
- Compare those offered treatment first, with those offered later *(comparisons).*

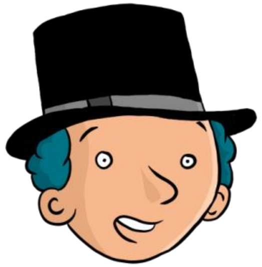
Keep in Mind

Randomized Assignment

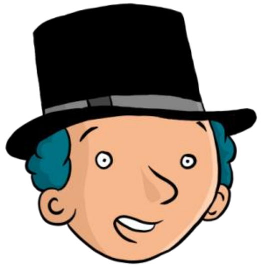
In Randomized Assignment, large enough samples, produces 2 statistically equivalent groups.

We have identified the perfect clone.

Randomized beneficiary



Randomized comparison



Feasible for prospective evaluations with over-subscription/excess demand.

Most pilots and new programs fall into this category.

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

What if we can't *choose*?

- It's not always possible to choose a control group. What about:
  - National programs where everyone is eligible?
  - Programs where participation is voluntary?
  - Programs where you can't exclude anyone?

Can we compare Enrolled & Not Enrolled?

Selection Bias!

Randomly offering/promoting program

If you can exclude some units, but can't force anyone:

- Offer the program to a random sub-sample
- Many will accept
- Some will not accept

⇨ Randomized offering

- If you can't exclude anyone, and can't force anyone:

- Making the program available to everyone
- Provide additional encouragement/ incentives to a random sub-sample:

  Additional Information & Encouragement.

  Incentives (small gift or prize or bus fare).

⇨ Randomized promotion

Randomly offering or promoting program

Necessary conditions:

1. Offered/promoted and not-offered/ not-promoted groups are comparable:
   - Whether or not you offer or promote is not correlated with population characteristics
   - Guaranteed by randomization.

2. Offered/promoted group has higher enrollment in the program.

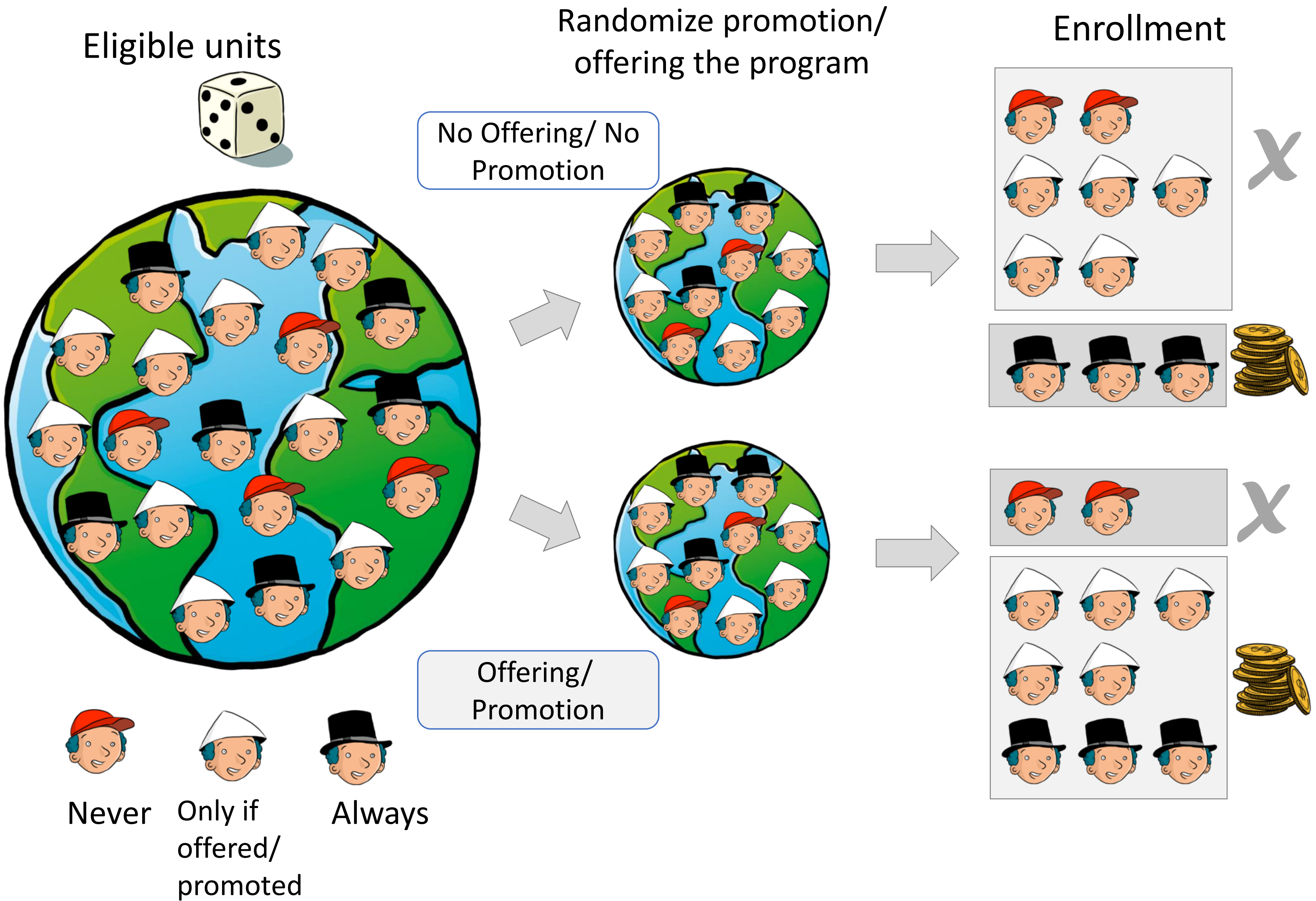
3. Offering/promotion of program does not affect outcomes directly.

Randomly offering or promoting program

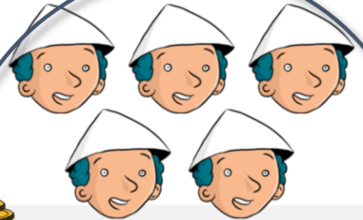
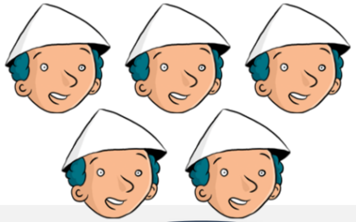
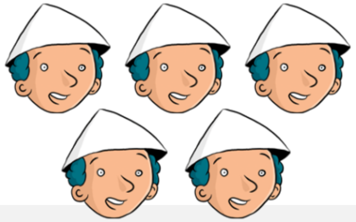
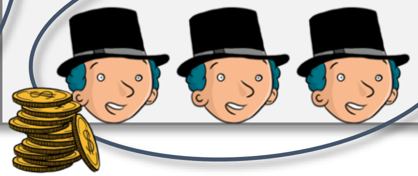
3 groups of units/individuals

		WITH offering/ promotion	WITHOUT offering/ promotion
	Never Enroll	X	X
	Only Enroll if offered/ promoted		X
	Always Enroll		

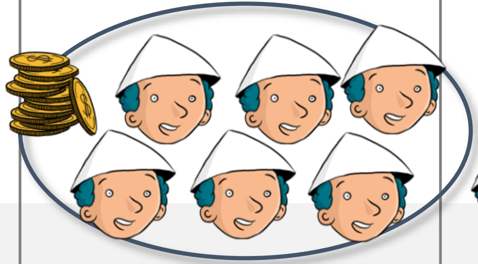
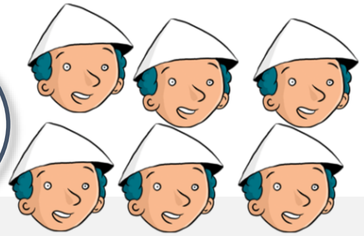
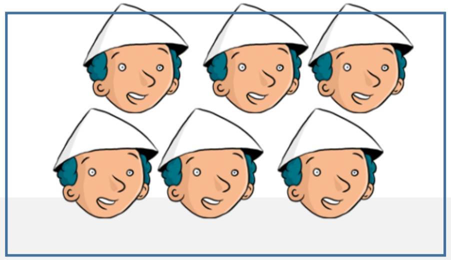
Randomly offering or promoting program



Randomly offering or promoting program

	Offered /Promoted Group	Not Offered/ Not Promoted Group	Impact
	%Enrolled=80% Average Y for entire group=100	%Enrolled=30% Average Y for entire group=80	ΔEnrolled=50% ΔY=20 Impact= 20/50%=40
Never Enroll			
Only Enroll if Offered/ Promoted			
Always Enroll			

Case 4: Progresa Randomized Offering

	Offered group	Not offered group	Impact
	%Enrolled=92% Average Y for entire group = 268	%Enrolled=0% Average Y for entire group = 239	ΔEnrolled=0.92 ΔY=29 Impact= 29/0.92 =31
Never Enroll			
Enroll if Offered			
Always Enroll			

Case 4: Randomized Offering

	Estimated Impact on Consumption (Y)
Instrumental Variables Regression	29.8**
Instrumental Variables with Controls	30.4**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Examples: Randomized Promotion

- Maternal Child Health Insurance in *Argentina*
  Intensive information campaigns

- Community Based School Management in *Nepal*
  NGO helps with enrollment paperwork

Community Based School Management in *Nepal*

- Context:
  - A centralized school system
  - 2003: Decision to allow local administration of schools
- The program:
  - Communities express interest to participate.
  - Receive monetary incentive ($1500)
- What is the impact of local school administration on:
  - School enrollment, teachers absenteeism, learning quality, financial management
- Randomized promotion:
  - NGO helps communities with enrollment paperwork.
  - 40 communities with randomized promotion *(15 participate)*
  - 40 communities without randomized promotion *(5 participate)*

Maternal Child Health Insurance in *Argentina*

- Context:
  - 2001 financial crisis
  - Health insurance coverage diminishes
- Pay for Performance (P4P) program:
  - Change in payment system for providers.
  - 40% payment upon meeting quality standards
- What is the impact of the new provider payment system on health of pregnant women and children?
- Randomized promotion:
  - Universal program throughout the country.
  - Randomized intensive information campaigns to inform women of the new payment system and increase the use of health services.

Case 5: Contaminated Designs

- Sources of Contamination
  - Not all intended treatments get treatment so less than full take up
  - Some intended comparisons get treatment

- Estimate effect of “Offering Program” on actual program takeup

- Intent to Treat: Use “Offering Program” as an instrument for actual program takeup

- Treatment on Treated: Intent to Treat divided by effect of “Offering Program” on actual takeup

Keep in Mind

Randomized Offering/Promotion

Randomized Promotion needs to be an effective promotion strategy
(Pilot test in advance!)

Promotion strategy will help understand how to increase enrollment in addition to impact of the program.

Don’t exclude anyone but…

Strategy depends on success and validity of offering/promotion.

Strategy estimates a local average treatment effect. Impact estimate valid only for the triangle hat type of beneficiaries.

Nightmares

- Imperfect Compliance with Randomization Protocol
  - Less than 100% takeup
  - Some controls takeup

- Spillover

- Attrition

What if we can't *assign*?

- It's not always possible to assign pure treatment control groups.
  - National programs where everyone is eligible?
  - Programs where participation is voluntary?
  - Programs where you can't exclude anyone?

Can we compare Enrolled & Not Enrolled?

Selection Bias!

Randomly offering/promoting program

- If you can exclude some units, but can't force anyone:
  - Offer the program to a random sub-sample
  - Many will accept
  - Some will not accept

→ Randomized offering

- If you can't exclude anyone, and can't force anyone:
  - Making the program available to everyone
  - Provide additional encouragement/ incentives to a random sub-sample:
    - Additional Information & Encouragement.
    - Incentives (small gift or prize or bus fare).

→ Randomized promotion

Randomly offering or promoting program

Necessary conditions:

1. Offered/promoted and not-offered/ not-promoted groups are comparable:
   - Whether or not you offer or promote is not correlated with population characteristics
   - Guaranteed by randomization.

2. Offered/promoted group has higher enrollment in the program.

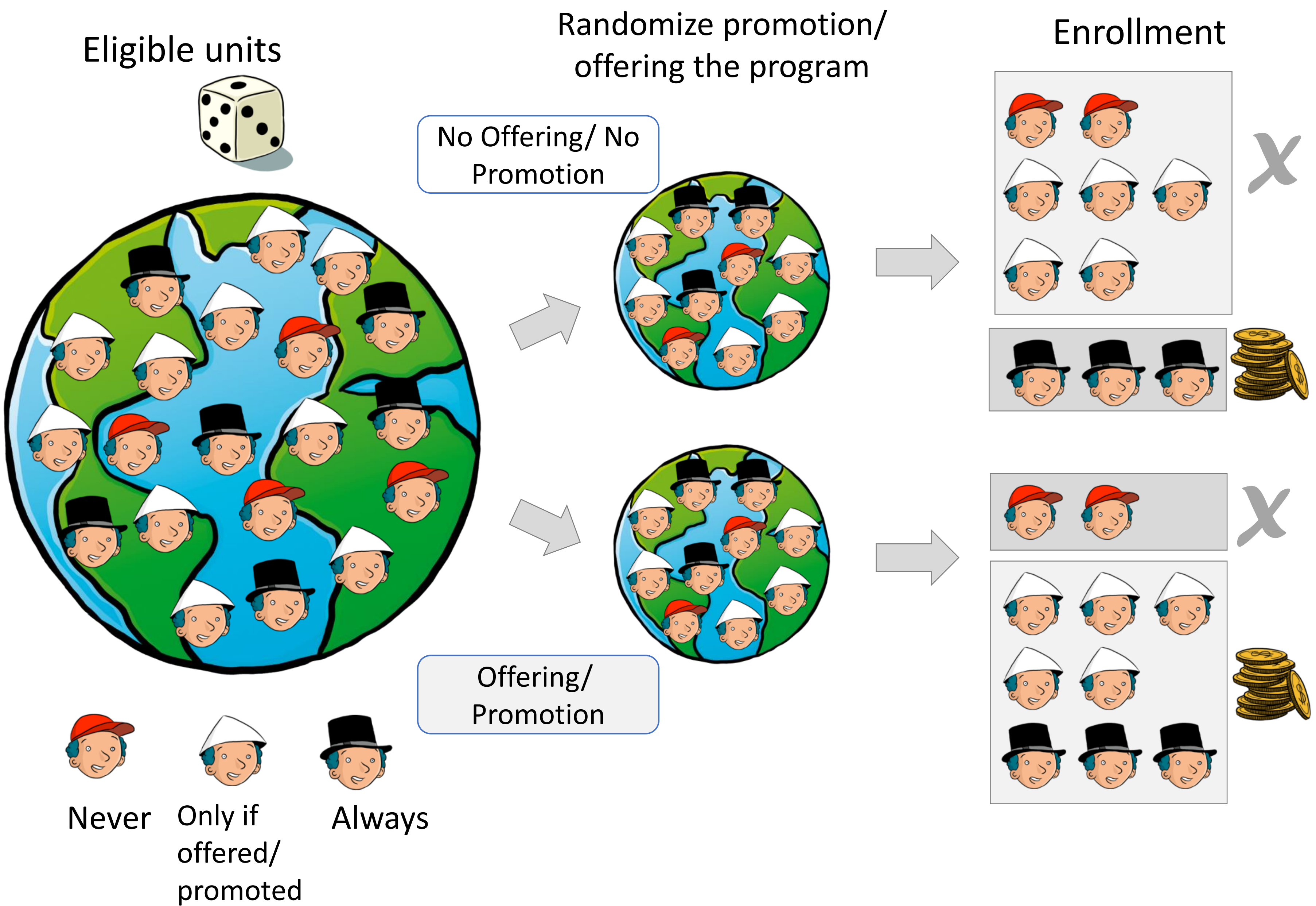
3. Offering/promotion of program does not affect outcomes directly.

Randomly offering or promoting program

3 groups of units/individuals

		WITH offering/ promotion	WITHOUT offering/ promotion
	Never Enroll	X	X
	Only Enroll if offered/ promoted		X
	Always Enroll		

Randomly offering or promoting program



Randomly offering or promoting program

	Offered /Promoted Group	Not Offered/ Not Promoted Group	Impact
	%Enrolled=80% Average Y for entire group=100	%Enrolled=30% Average Y for entire group=80	ΔEnrolled=50% ΔY=20 Impact= 20/50%=40
Never Enroll			
Only Enroll if Offered/ Promoted			
Always Enroll			

Case 4: Progresa Randomized Offering

	Offered group	Not offered group	Impact
	%Enrolled=92% Average Y for entire group = 268	%Enrolled=0% Average Y for entire group = 239	ΔEnrolled=0.92 ΔY=29 Impact= 29/0.92 =31
Never Enroll			
Enroll if Offered			
Always Enroll			

Case 4: Randomized Offering

Estimated Impact on Consumption (Y)	
Instrumental Variables Regression	29.8**
Instrumental Variables with Controls	30.4**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Case 5: Contaminated Designs

- Sources of Contamination
  - Not all intended treatments get treatment so less than full take up
  - Some intended comparisons get treatment

- Estimate effect of “Offering Program” on actual program takeup

- Intent to Treat: Use “Offering Program” as an instrument for actual program takeup

- Treatment on Treated: Intent to Treat divided by effect of “Offering Program” on actual takeup

Nightmares

- Imperfect Compliance with Randomization Protocol
  - Less than 100% takeup
  - Some controls takeup
- Spillover
- Attrition

Nightmares

- Imperfect Compliance with Randomization Protocol
  - Less than 100% takeup
  - Some controls takeup

- Spillover

- Attrition

Spillover -- Unit of Randomization

- Choose according to type of program
  - Individual/Household
  - School/Health Clinic/catchment area
  - Block/Village/Community/Region

- Keep in mind
  - Need "sufficiently large" number of units to detect minimum desired impact: Power.
  - Clustering reduces effective sample size
  - Standard Errors need to be "clustered"
  - Spillovers/contamination
  - Operational and survey costs

As a rule of thumb, randomize at the smallest viable unit of implementation.

Attrition -- Reweight observed data

- Want to look like original baseline sample: same distribution of observable characteristics

- Use “propensity score”,
  - Compute everyone’s probability attrition, based on their observable characteristics not balance.
  - Weight observed data with Inverse Probability of Attrition

Regression: The Selection Problem

- Example: Does Hospitalization make people healthier?
- Data:

Group	Sample Size	Mean Health	Std Error
Hospital	7,774	3.21	0.014
No Hospital	90,049	3.93	0.003

- Health Status: "Would you say your health is excellent, very good, good, fair or poor" coded 5,4,3,2, or 1
- Difference 39.3 – 3.21 = 0.72
- T-Stat for Null Hypothesis Diff = 0 is 58.9
- Does Hospitalization make people healthier?

More Formally

- People who are sicker to begin with go to hospital
- Let $D_i = 1$ if go to hospital, 0 otherwise
- Let $Y_i =$ Health with two potential outcomes:

$$Y_i = Y_{1i} \text{ when } D_i = 1$$
$$Y_i = Y_{0i} \text{ when } D_i = 0$$

- Observed outcome is

$$Y_i = Y_{0i} *(1 - D_i) + Y_{1i}*D_i$$
$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})*D_i$$

- This is just a simple linear equation

Can Be Written as a Regression

- Can write experiment in regression form

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})*D_i$$

$$Y_i = \alpha + \beta * D_i + \varepsilon_i$$

where

$$\alpha = E(Y_0) = \text{Control group mean}$$

$$\beta = E(Y_1 - Y_0) = \text{treatment effect}$$

$$\varepsilon_i = Y_i - E(Y_i) = \text{unobserved component of } \mathrm{Y_i}$$

Self Selection Revised

- Difference in health status can be written as:

$$E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0) = \left[E(Y_{1i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 1)\right] + \left[E(Y_{0i} \mid D_i = 1) - E(Y_{0i} \mid D_i = 0)\right]$$

Observed Difference = Treatment on Treated + Selection Bias

- Randomization solves selection problem by assuring that

$$E(Y_{0i} \mid D_i = 1) = E(Y_{0i} \mid D_i = 0)$$

Regression Analysis of Experiments

- Can write experiment in regression form

$$Y_i = \alpha + \beta * D_i + \varepsilon_i$$

- Where

$$\alpha = E(Y_0) = \text{Control group mean}$$

$$\beta = E(Y_1 - Y_0) = \text{treatment effect}$$

$$\varepsilon_i = Y_i - E(Y_i) = \text{unobserved component of } \mathrm{Y_i}$$

- Selection bias is a question of whether ε is correlated with D_i

Can Minimize Selection Bias with Multiple Regression

- Control for observable factors correlated with treatment status

$$Y_i = \alpha + \beta * D_i + \gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_k X_k + \varepsilon_i$$

- Each coefficient has its own ***t-statistic*** and ***p-value***
- β is change in Y when D=1 holding values of X's constant
- However, there still maybe factors cannot observe in the error term correlated w/ treatment status
- Called omitted variable bias
- This is reason we turn to randomization and other identification methods

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

Matching

Idea

For each treated unit pick up the best comparison unit (*match*) from another data source.

How?

Matches are selected on the basis of similarities in observed characteristics.

Issue?

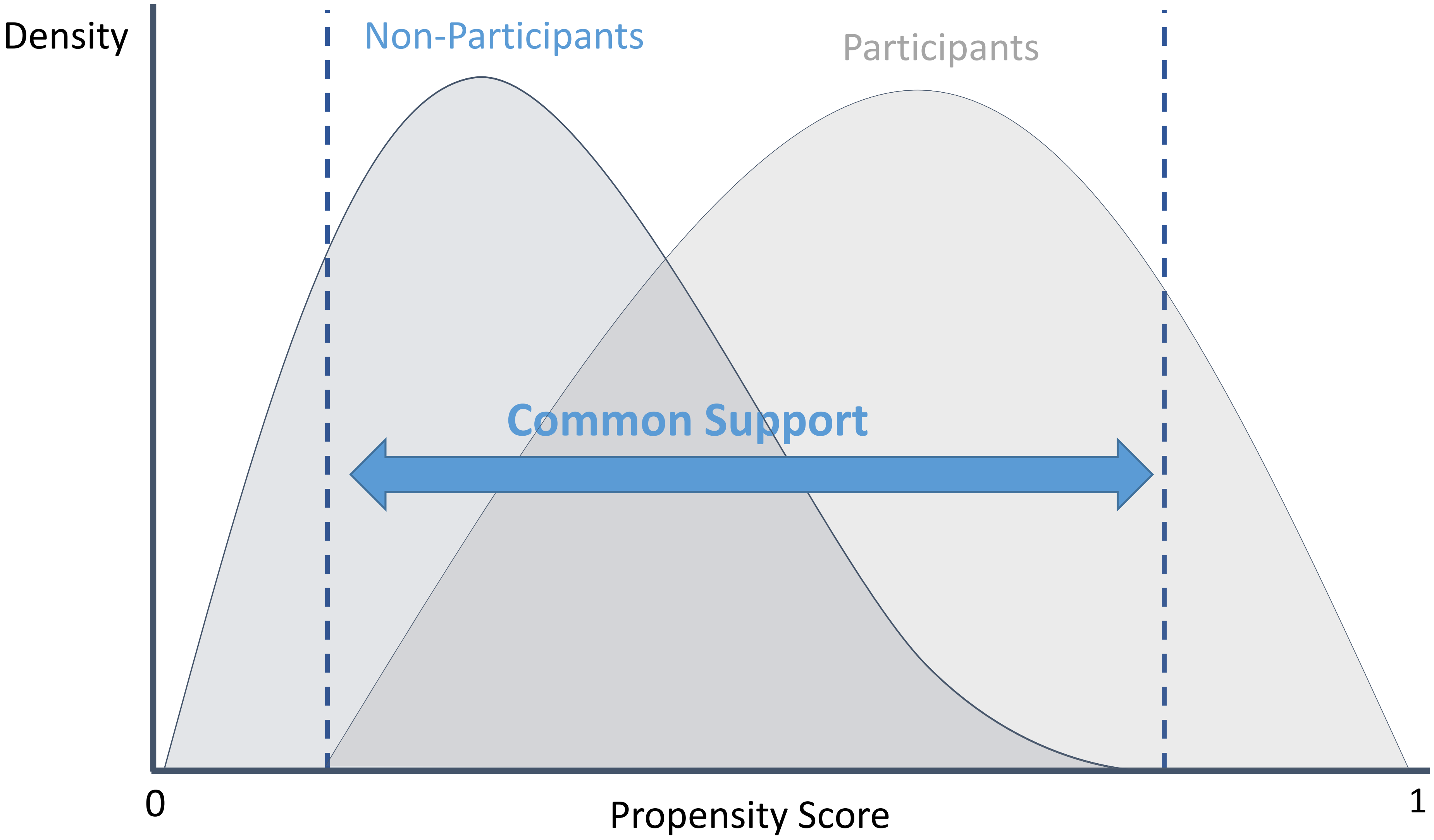
If there are *unobservable* characteristics and those unobservables influence participation: Selection bias!

Propensity-Score Matching *(PSM)*

- Comparison Group: non-participants with same observable characteristics as participants.
  - In practice, it is very hard.
  - There may be many important characteristics!

- Match on the basis of the "propensity score", *Solution proposed by Rosenbaum and Rubin:*
  - Compute everyone's probability of participating, based on their observable characteristics.
  - Choose matches that have the same probability of participation as the treatments.
  - See appendix 2.

Density of propensity scores

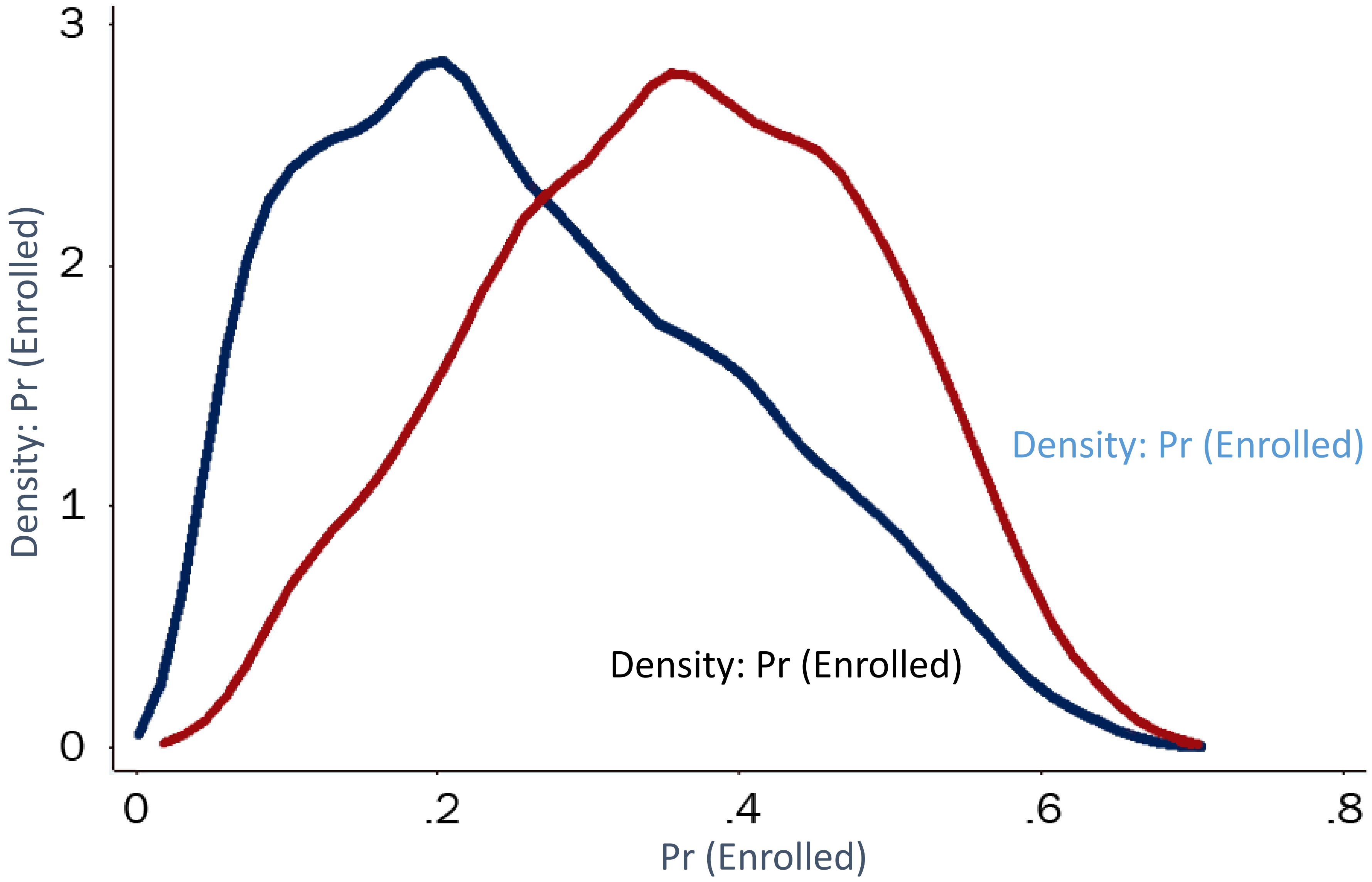


Case 7: Progresa Matching *(P-Score)*

Baseline Characteristics	**Estimated Coefficient** *Probit Regression, Prob Enrolled=1*
Head's age (years)	-0.022**
Spouse's age (years)	-0.017**
Head's education (years)	-0.059**
Spouse's education (years)	-0.03**
Head is female=1	-0.067
Indigenous=1	0.345**
Number of household members	0.216**
Dirt floor=1	0.676**
Bathroom=1	-0.197**
Hectares of Land	-0.042**
Distance to Hospital (km)	0.001*
Constant	0.664**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Case 7: Progresa Common Support



Case 7: Progresa Matching *(P-Score)*

Estimated Impact on Consumption (Y)	
Multivariate Linear Regression	7.06+

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**). If significant at 10% level, we label impact with +

Keep in Mind

Matching

Matching requires large samples and good quality data.

Matching at baseline can be very useful:

- Know the assignment rule and match based on it
- combine with other techniques (i.e. diff-in-diff)

Ex-post matching is risky:

- If there is no baseline, be careful!
- matching on endogenous ex-post variables gives **bad** results.

Progresa Policy Recommendation?

Impact of Progresa on Consumption (Y)	
Case 1: Before & After	34.28**
Case 2: Enrolled & Not Enrolled	-4.15
Case 3: Randomized Assignment	29.75**
Case 4: Randomized Offering	30.4**
Case 5: Discontinuity Design	30.58**
Case 6: Difference-in-Differences	25.53**
Case 7: Matching	7.06+

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**). If significant at 10% level, we label impact with +

Steps in Propensity Score Matching

1. Representative & highly comparables survey of non-participants and participants.
2. Pool the two samples and estimated a logit (or probit) model of program participation.
3. Restrict samples to assure **common support** (important source of bias in observational studies)
4. For each participant find a sample of non-participants that have similar propensity scores
5. Compare the outcome indicators. The difference is the **estimate of the gain** due to the program for that observation.
6. Calculate the mean of these individual gains to obtain the average overall gain.

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

Many social programs select beneficiaries using an index or score:

Discontinuity Design

Anti-poverty Programs	Targeted to households below a given poverty index/income
Pensions	Targeted to population above a certain age
Education	Scholarships targeted to students with high scores on standarized text
Agriculture	Fertilizer program targeted to small farms less than given number of hectares)

Example: Effect of fertilizer program on agriculture production

Goal

Improve agriculture production (rice yields) for small farmers

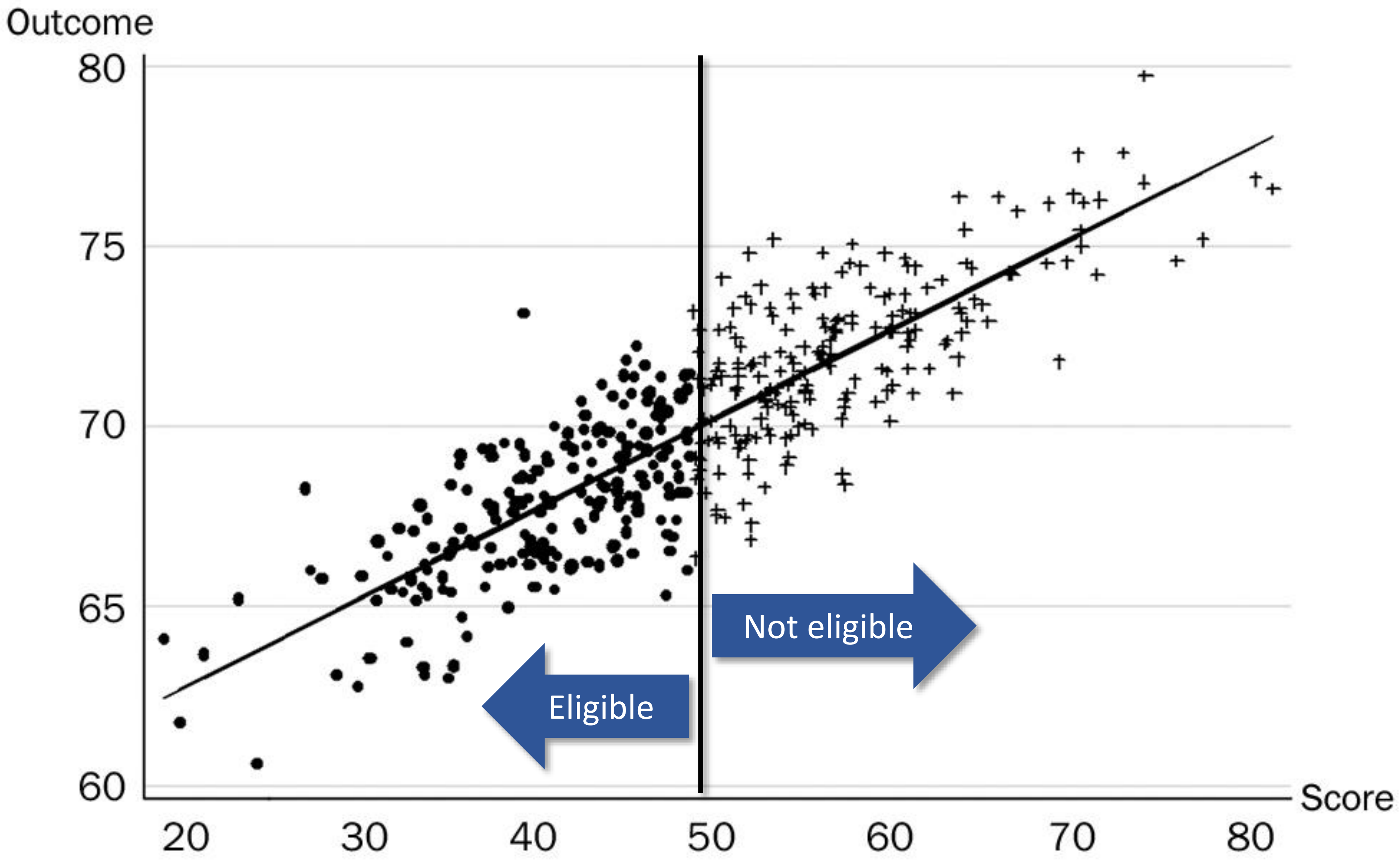
Method

- Farms with a score (Ha) of land ≤50 are small
- Farms with a score (Ha) of land >50 are not small

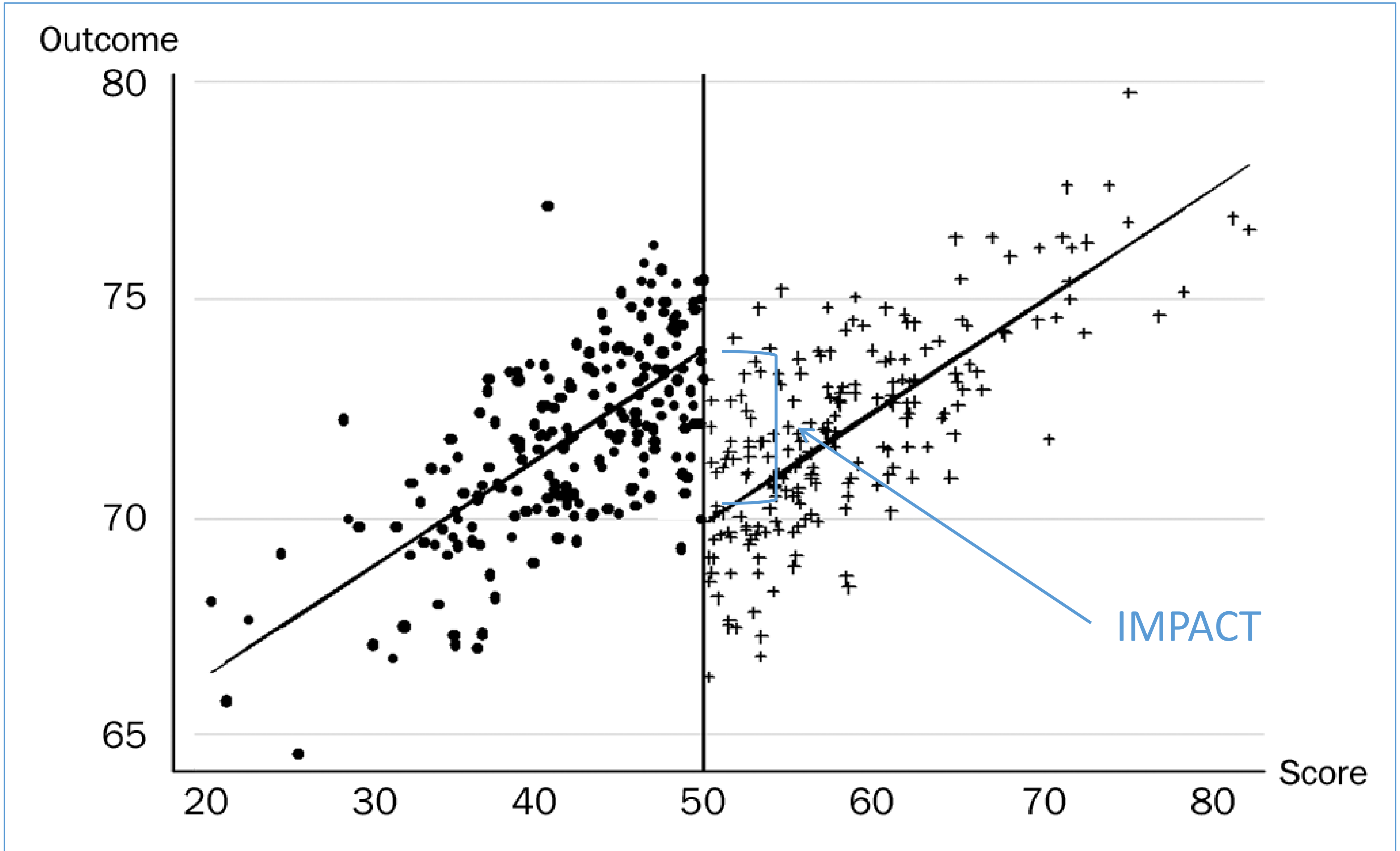
Intervention

Small farmers receive subsidies to purchase fertilizer

Regression Discontinuity Design-Baseline



Regression Discontinuity Design-Post Intervention



Case 5: Discontinuity Design

- We have a continuous eligibility index with a defined cut-off
  - Households with a score ≤ cutoff are eligible
  - Households with a score > cutoff are not eligible
  - Or vice-versa
- Intuitive explanation of the method:
  - Units just above the cut-off point are very similar to units just below it – *good comparison.*
  - Compare outcomes *Y* for units just *above and below* the cut-off point.

For a discontinuity design, you need:

1) Continuous eligibility index
2) Clearly defines eligibility cut-off.

Case 5: Discontinuity Design

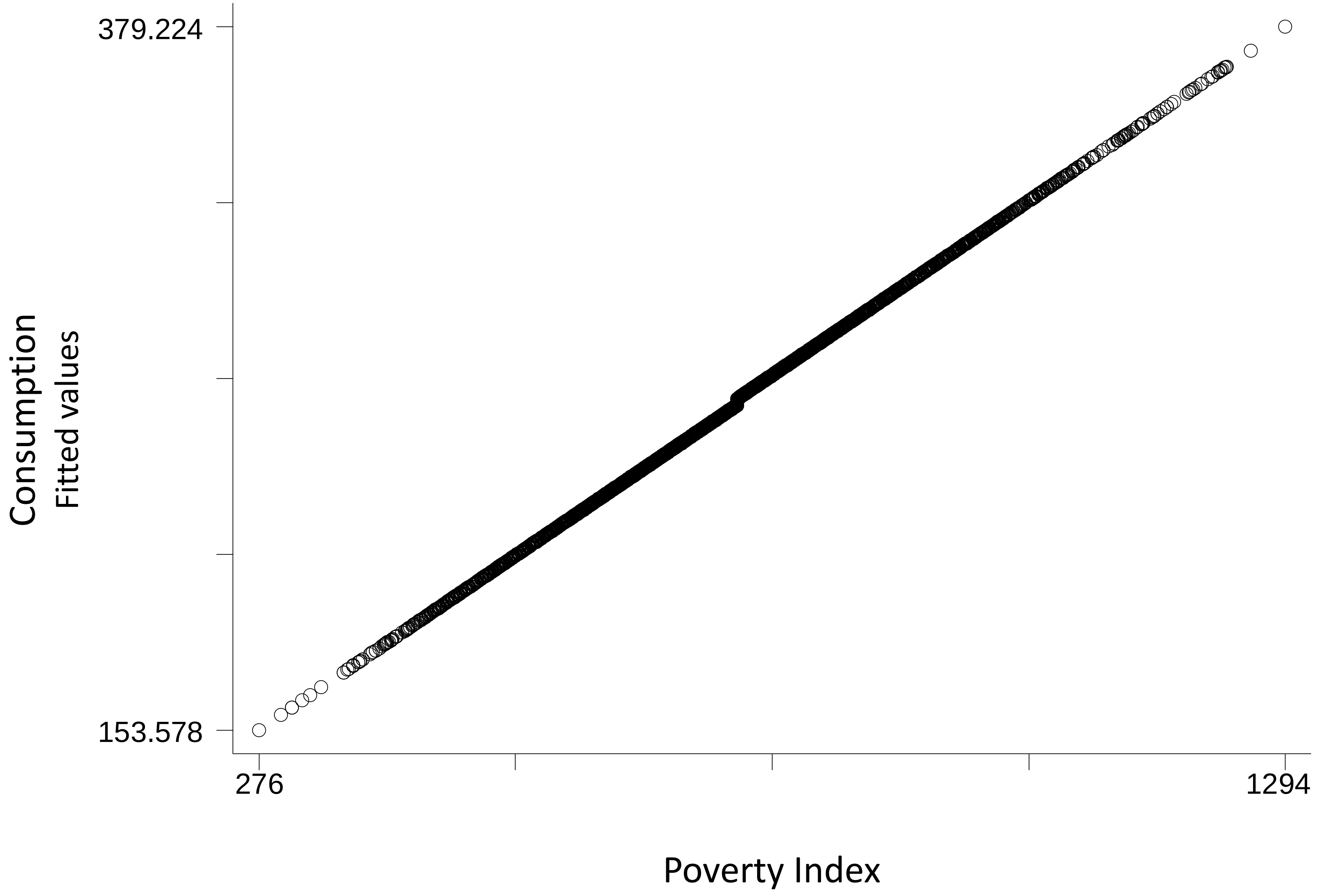
- Eligibility for Progresa is based on national poverty index

- Household is poor if score ≤ 750

- Eligibility for Progresa:
  - Eligible=1 if score ≤ 750
  - Eligible=0 if score > 750

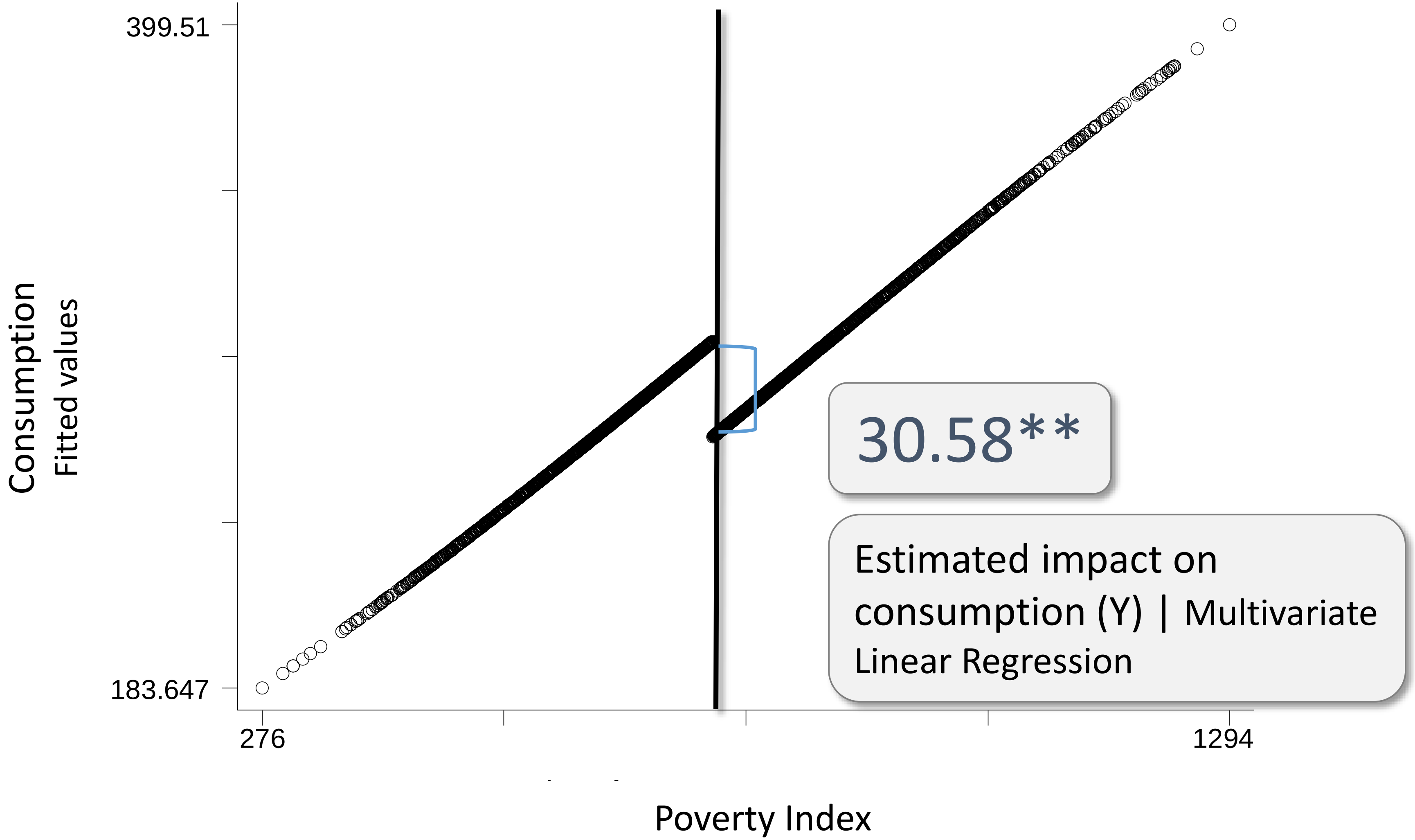
Case 5: Discontinuity Design

Score vs. consumption at Baseline-No treatment



Case 5: Discontinuity Design

Score vs. consumption post-intervention period-treatment



(**) Significant at 1%

Keep in Mind

Discontinuity Design

Discontinuity Design requires continuous eligibility criteria with clear cut-off.

Gives unbiased estimate of the treatment effect: *Observations* ***just across*** *the cut-off are good comparisons.*

No need to **exclude** a group of eligible households/ individuals from treatment.

Can sometimes use it for programs that already ongoing.

Keep in Mind

Discontinuity Design

Discontinuity Design produces a local estimate:

- *Effect of the program around the cut-off point/discontinuity.*
- *This is not always generalizable.*

Power:

- *Need many observations around the cut-off point.*

Avoid mistakes in the statistical model: *Sometimes what looks like a discontinuity in the graph, is something else.*

Randomized Assignment

Randomized Offering/Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

Difference-in-differences
(Diff-in-diff)

Y=Girls exam score (percentage of correct answers)
P= Tutoring

	Enrolled/ With tutoring		**Not Enrolled/ No tutoring**		
After	0.74		0.81		
	-		-		
Before	0.60		0.78		
Difference	**+0.14**	**-**	**+0.03**	**=**	**0.11**

Diff-in-Diff: Impact=$(Y_{t1}-Y_{t0})-(Y_{c1}-Y_{c0})$

Difference-in-differences
(Diff-in-diff)

Y=Girls exam score (percentage of correct answers)

P= Tutoring

	Enrolled/ With tutoring		Not Enrolled/ No tutoring	Difference
After	0.74	-	0.81	**-0.07**
Before	0.60	-	0.78	**-0.18**

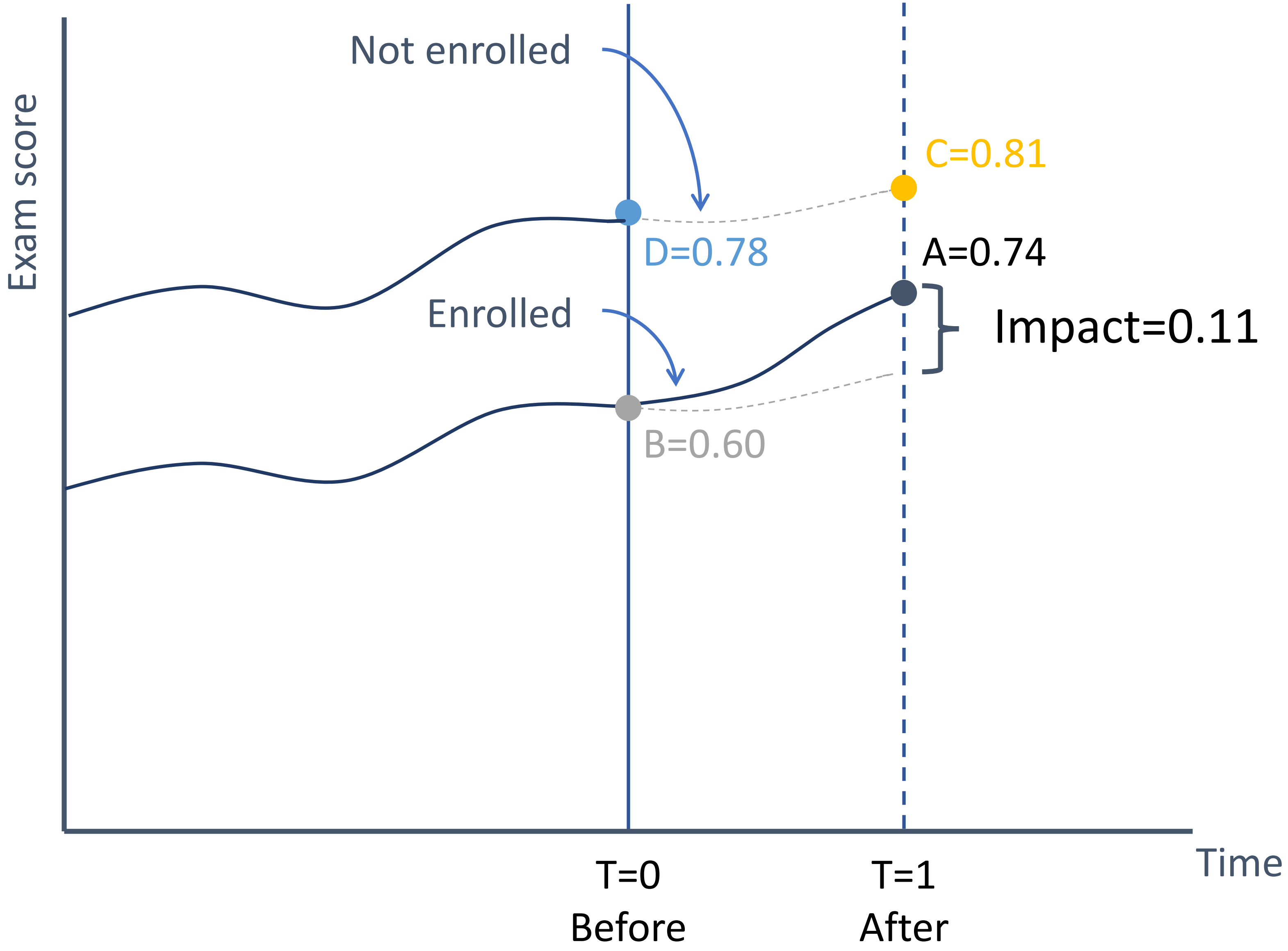
-

=

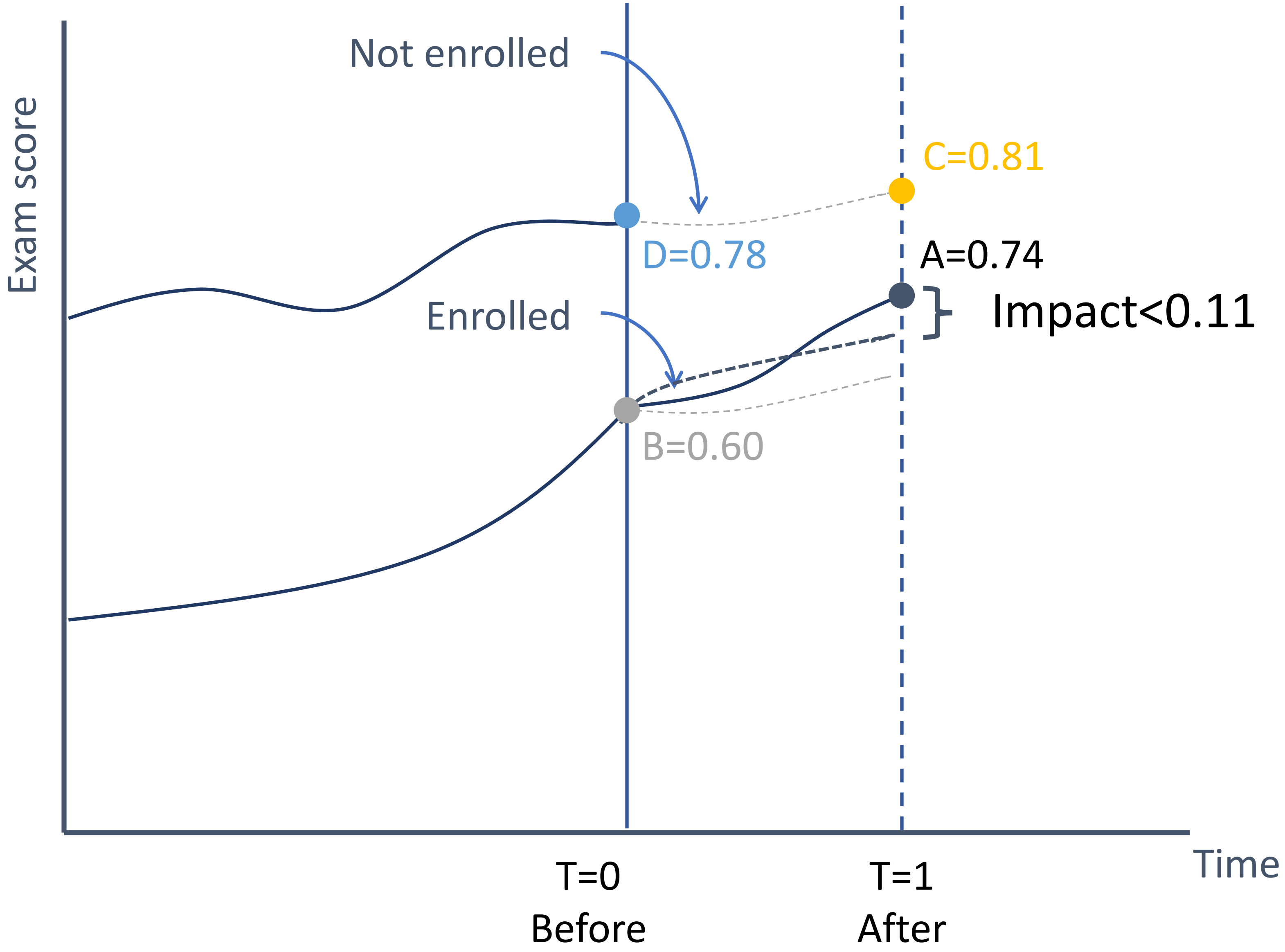
0.11

Diff-in-Diff: Impact=$(Y_{t1}-Y_{c1})-(Y_{t0}-Y_{c0})$

$$Impact = (A-B)-(C-D) = (A-C)-(B-D)$$



$$Impact = (A-B)-(C-D) = (A-C)-(B-D)$$



Dif in Dif in Regression form

- Estimating equation

$$y_{it} = \alpha_i + \beta_T + \gamma D_{it} + \sum_k \delta_k X_{kit} + \varepsilon_{it}$$

- Where:

$\alpha_i =$ individual fixed effect (i.e. dummy for each i)

$\beta_t =$ year or time fixed effect

$\gamma =$ treatment effect

Case 6: Difference-in-differences

	Enrolled	Not Enrolled	Difference
Follow-up ***(T=1)*** Consumption (Y)	268.75	290	-21.25
Baseline ***(T=0)*** Consumption (Y)	233.47	281.74	-48.27
Difference	35.28	8.26	27.02

Estimated Impact on Consumption (Y)	
Linear Regression	27.06**
Multivariate Linear Regression	25.53**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Progresa Policy Recommendation?

Impact of Progresa on Consumption (Y)	
Case 1: Before & After	34.28**
Case 2: Enrolled & Not Enrolled	-4.15
Case 3: Randomized Assignment	29.75**
Case 4: Randomized Offering	30.4**
Case 5: Discontinuity Design	30.58**
Case 6: Difference-in-Differences	25.53**

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).

Keep in Mind

Difference-in-Differences

Difference-in-Differences combines *Enrolled & Not Enrolled* with *Before & After.*

Slope: Generate counterfactual for change in outcome

Trends –slopes- are the same in treatments and comparisons *(Fundamental assumption).*

To test this, at least **3 observations** in time are needed:

- **2** observations before
- **1** observation after.