The
World Bank

HUMAN DEVELOPMENT
NETWORK

Human Development
Network

SIEF

Spanish Impact
Evaluation Fund

www.worldbank.org/hdchiefeconomist

MEASURING IMPACT

Impact Evaluation Methods for Policy Makers

Nazmul Chaudhury

The World Bank

These slides constitute supporting material to the *Impact Evaluation in Practice* Handbook : Gertler, P. J.; Martinez, S., Premand, P., Rawlings, L. B. and Christel M. J. Vermeersch, 2010, *Impact Evaluation in Practice*: Ancillary Material, The World Bank, Washington DC (www.worldbank.org/ieinpractice). The content of this presentation reflects the views of the authors and not necessarily those of the World Bank.

Reference

Spanish Version
& French Version
also available

www.worldbank.org/ieinpractice

HUMAN DEVELOPMENT
NETWORK

1

Causal Inference

Counterfactuals

False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled
(Apples & Oranges)

HUMAN DEVELOPMENT NETWORK

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

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)**?

HUMAN DEVELOPMENT NETWORK

Causal Inference

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

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

But difficult to estimate ...

HUMAN DEVELOPMENT NETWORK

Problem of Missing Data

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

For a program beneficiary:

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

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

HUMAN DEVELOPMENT NETWORK

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 estimate of the **counterfactual!**

HUMAN DEVELOPMENT NETWORK

Estimating impact of *P* on *Y*

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

OBSERVE (Y | P=1)
Outcome with treatment

ESTIMATE (Y | P=0)
The Counterfactual

IMPACT = Outcome with treatment - counterfactual

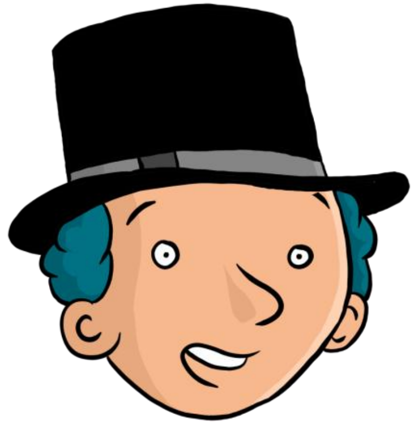
- Intention to Treat (**ITT**) – *Those offered treatment*
- Treatment on the Treated (**TOT**) – *Those receiving treatment*

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

HUMAN DEVELOPMENT NETWORK

Example: What is the Impact of…

giving Rana



additional pocket money

(P)

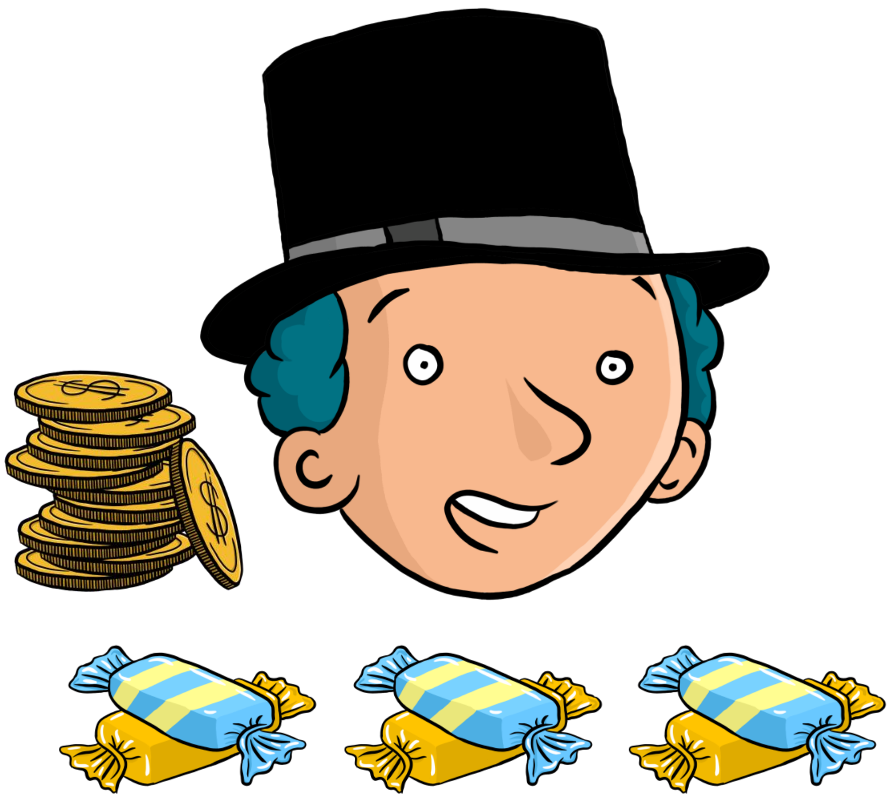
on Rana's consumption of candies

(Y)?

HUMAN DEVELOPMENT NETWORK

The Perfect *Clone*

Rana



6 candies

Rana's Clone

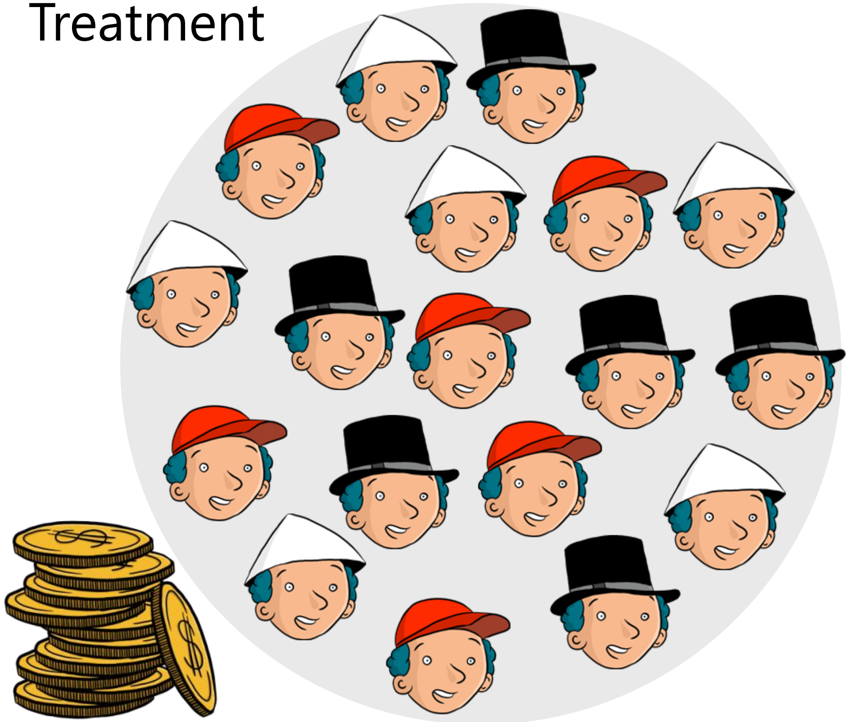
4 candies

IMPACT=6-4=2 Candies

HUMAN DEVELOPMENT NETWORK

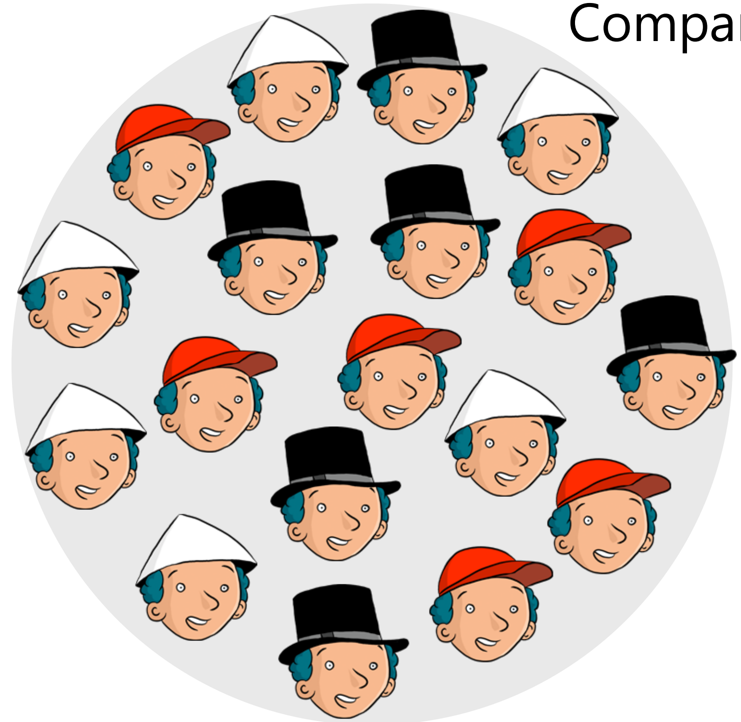
In reality, use statistics

Treatment



Average Y=6 candies

Comparison



Average Y=4 Candies

IMPACT=6-4=2 Candies

Case Study: Progresa

- National anti-poverty program in Mexico
  - Cash Transfers conditional on school and health care attendance
- Operational Rules:
  - Targeting:
    - Eligibility based on a proxy measure of poverty
  - Timing:
    - Started 1997
    - Phased Roll-out, 5 million beneficiaries by 2004

Case Study: Progresa

- Rigorous impact evaluation with rich data
  - 506 communities, 24,000 households
  - Baseline 1997, follow-up 2008

- Many outcomes of interest

  Here: Consumption per capita

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

Eligibility and Enrollment

1

Causal Inference

Counterfactuals

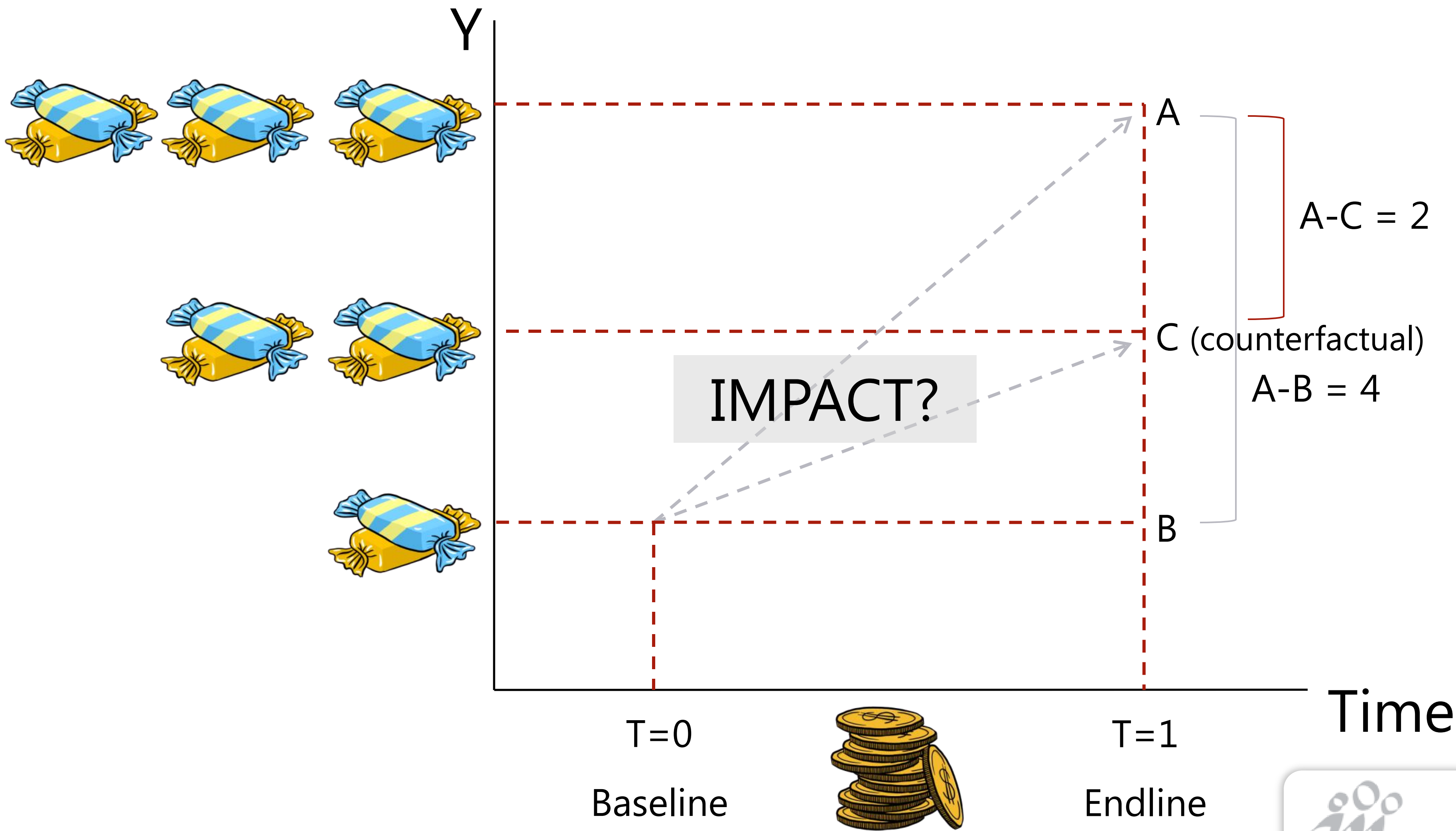
False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled (Apples & Oranges)

Counterfeit Counterfactual #1

Before & After



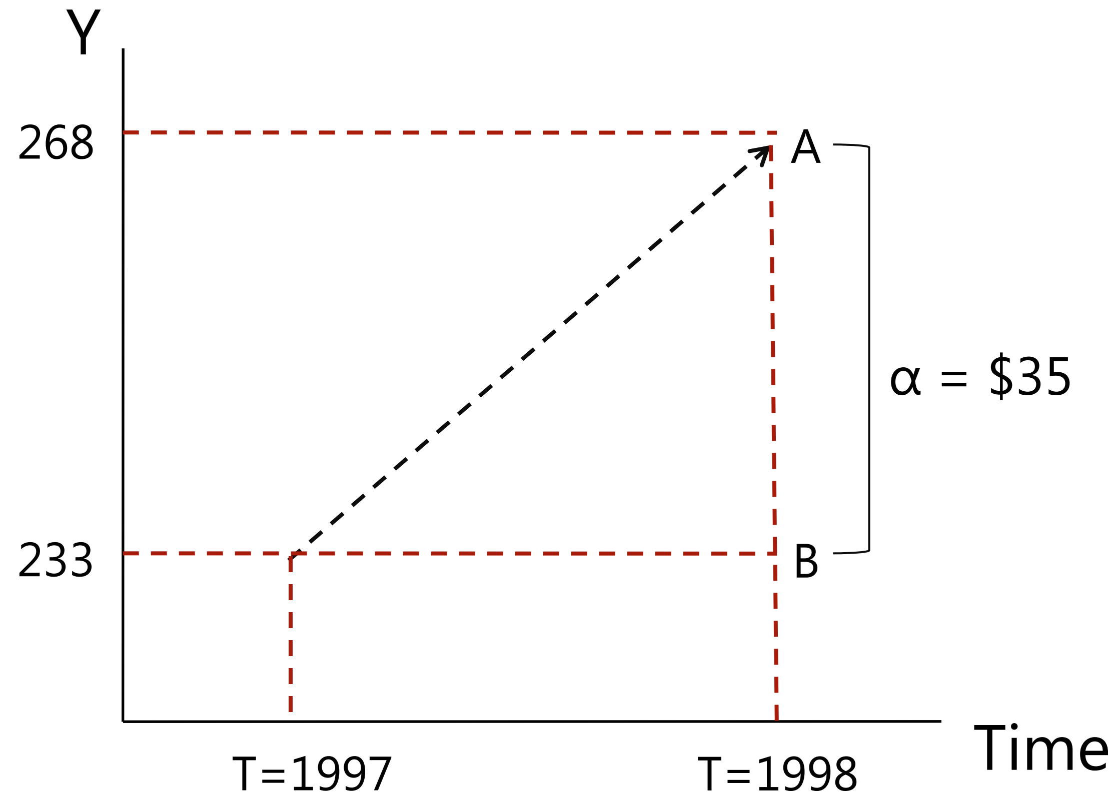
HUMAN DEVELOPMENT NETWORK

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

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

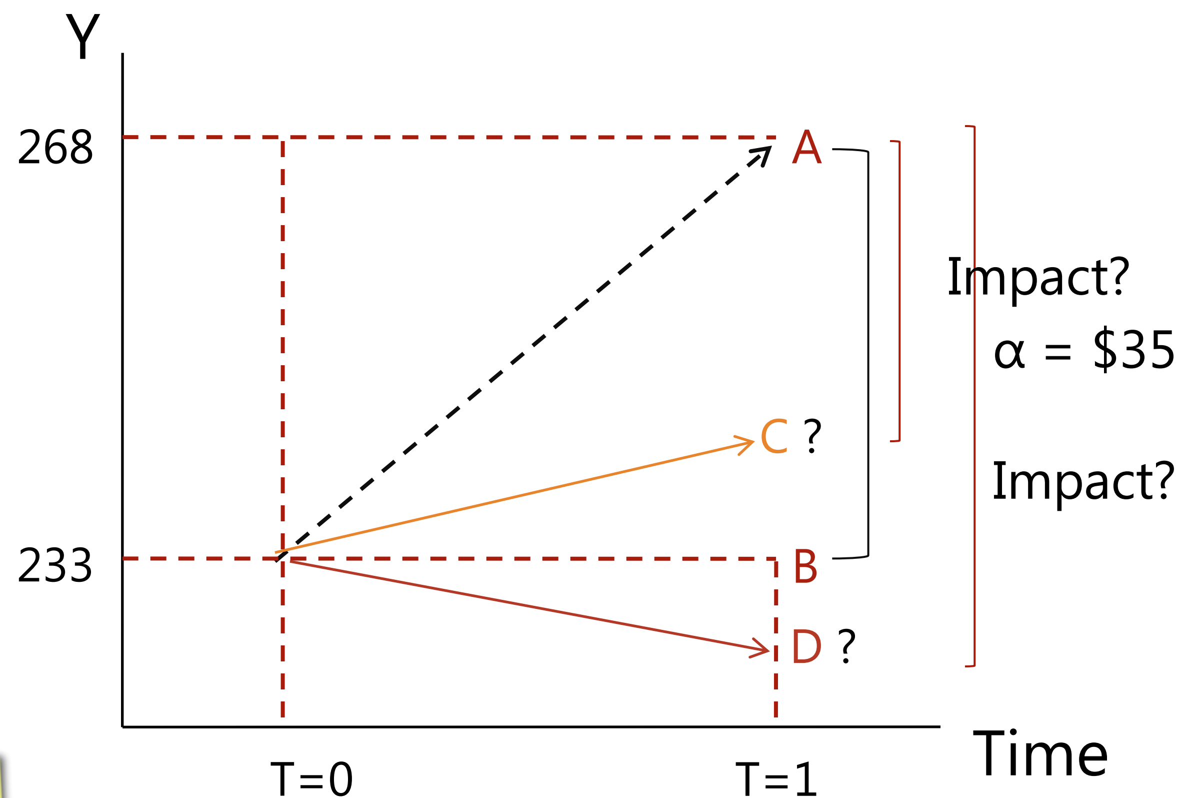
Case 1: What's the problem?

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!

HUMAN DEVELOPMENT NETWORK

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 *$(Y \mid P=1) - (Y \mid P=0)$*	-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 (**).

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

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 counterfactual and the impact.

HUMAN DEVELOPMENT NETWORK

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

Choosing your IE method(s)

Key information you will need for choosing the right method for your program:

Prospective/Retrospective Evaluation?

Targeting rules and criteria?

- Poverty targeting?
- Geographic targeting?

Timing: Roll-out plan (pipeline)?

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

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

HUMAN DEVELOPMENT NETWORK

Choosing your IE method(s)

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

Best Design

- Best comparison group you can find **+** least operational risk

Have we controlled for everything?

- Internal validity
- Good comparison group

Is the result valid for *everyone*?

- External validity
- Local versus global treatment effect
- Evaluation results apply to population we're interested in

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

Randomized Treatments & Comparison

Eligibles > Number of Benefits

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

Oversubscription

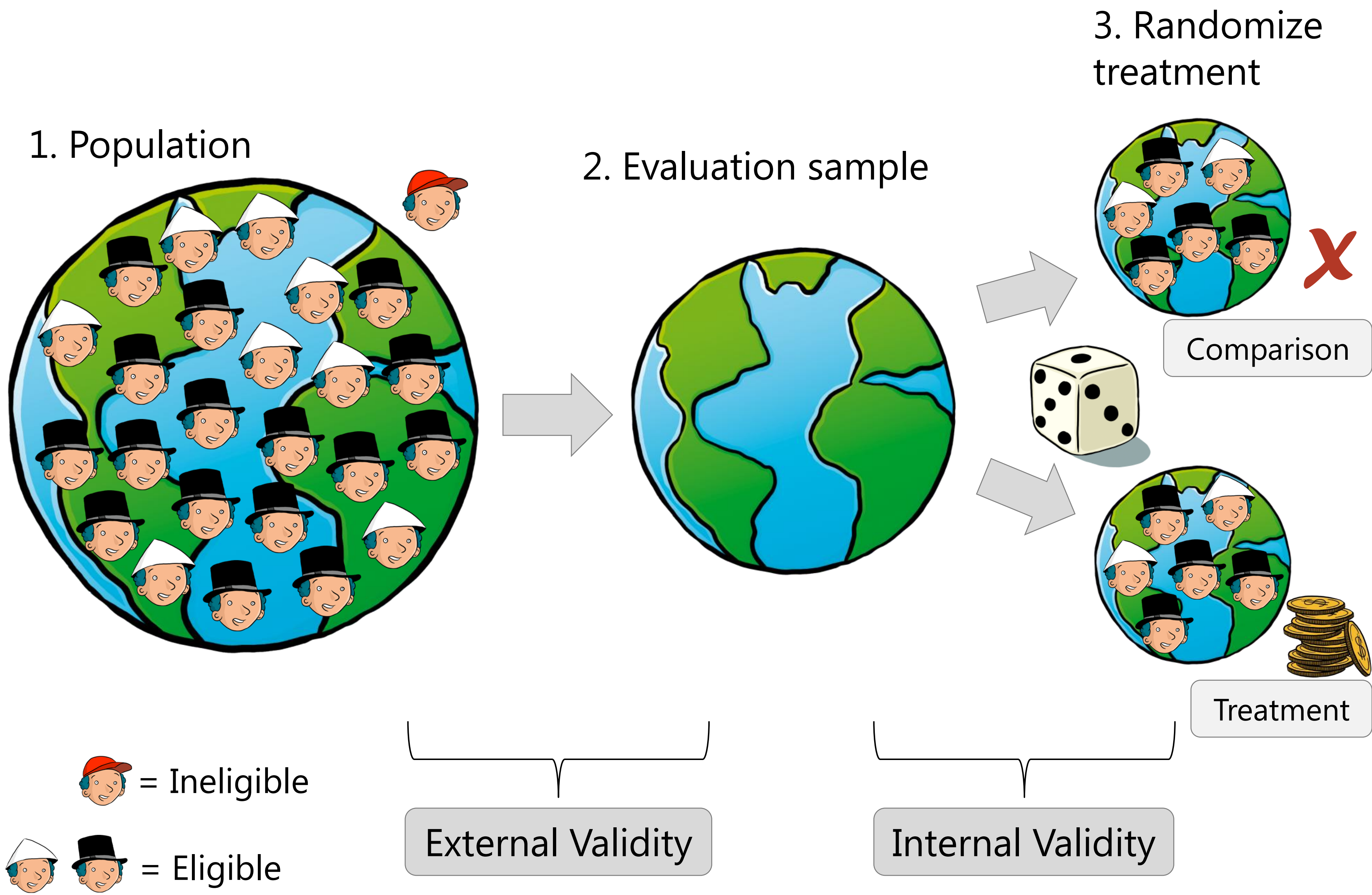
- Give each eligible unit the same chance of receiving treatment
- Compare those offered treatment with those not offered treatment *(comparisons)*.

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)*.

HUMAN DEVELOPMENT NETWORK

Randomized treatments and comparisons



Unit of Randomization

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

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

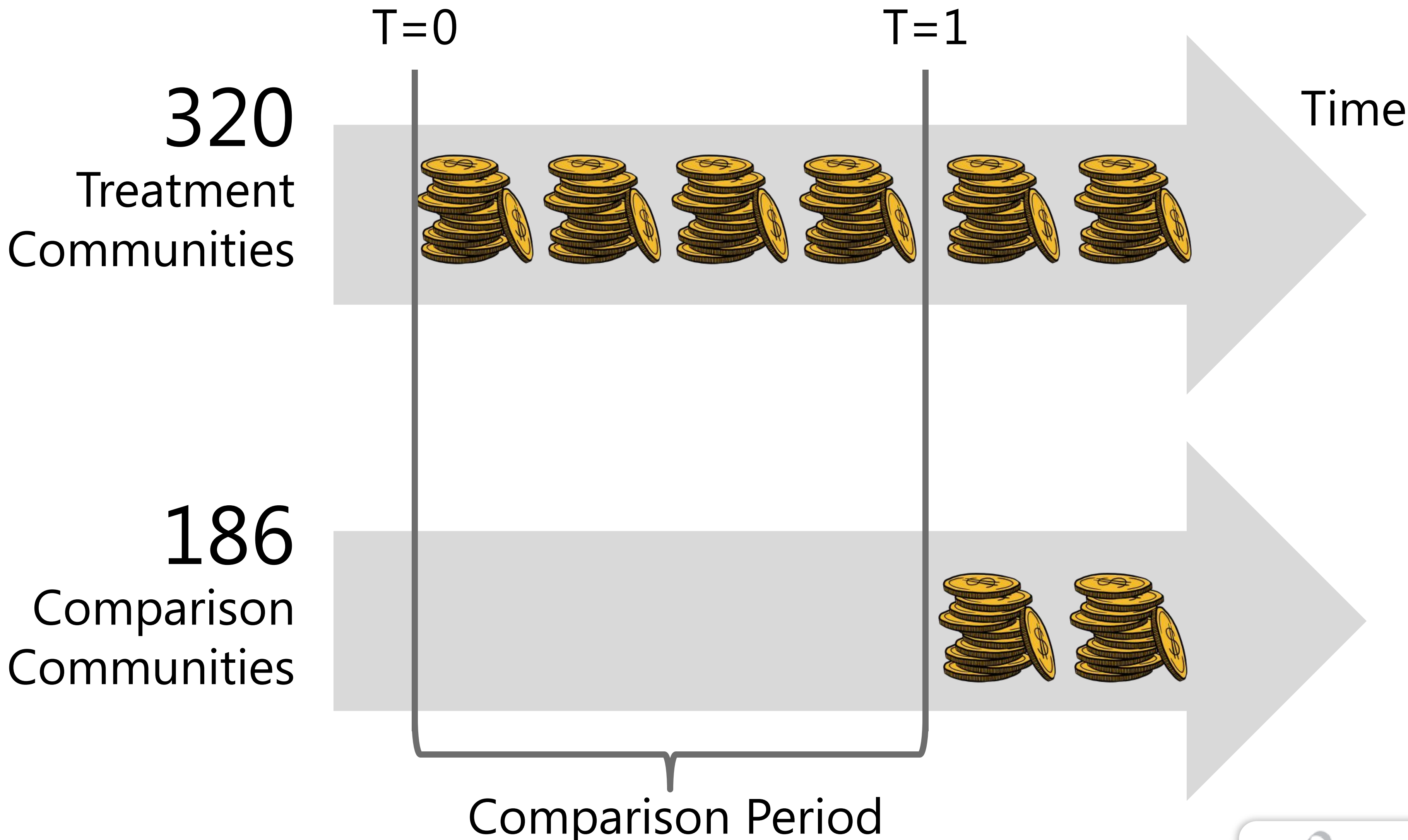
- Keep in mind
  - Need “sufficiently large” number of units to detect minimum desired impact: Power.
  - Spillovers/contamination
  - Operational and survey costs

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

HUMAN DEVELOPMENT NETWORK

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)

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

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 (**).

HUMAN DEVELOPMENT NETWORK

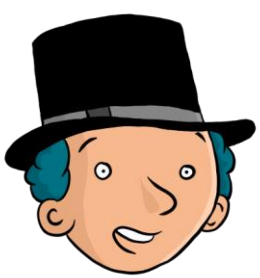
Keep in Mind

Randomized Assignment

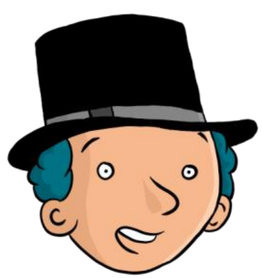
In Randomized Assignment, large enough samples, produces 2 statistically equivalent groups.

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.

HUMAN DEVELOPMENT NETWORK

Randomized assignment with different benefit levels

- Traditional impact evaluation question:
  - What is the impact of a program on an outcome?
- Other policy question of interest:
  - What is the optimal level for program benefits?
  - What is the impact of a "higher-intensity" treatment compared to a "lower-intensity" treatment?
- Randomized assignment with 2 levels of benefits:

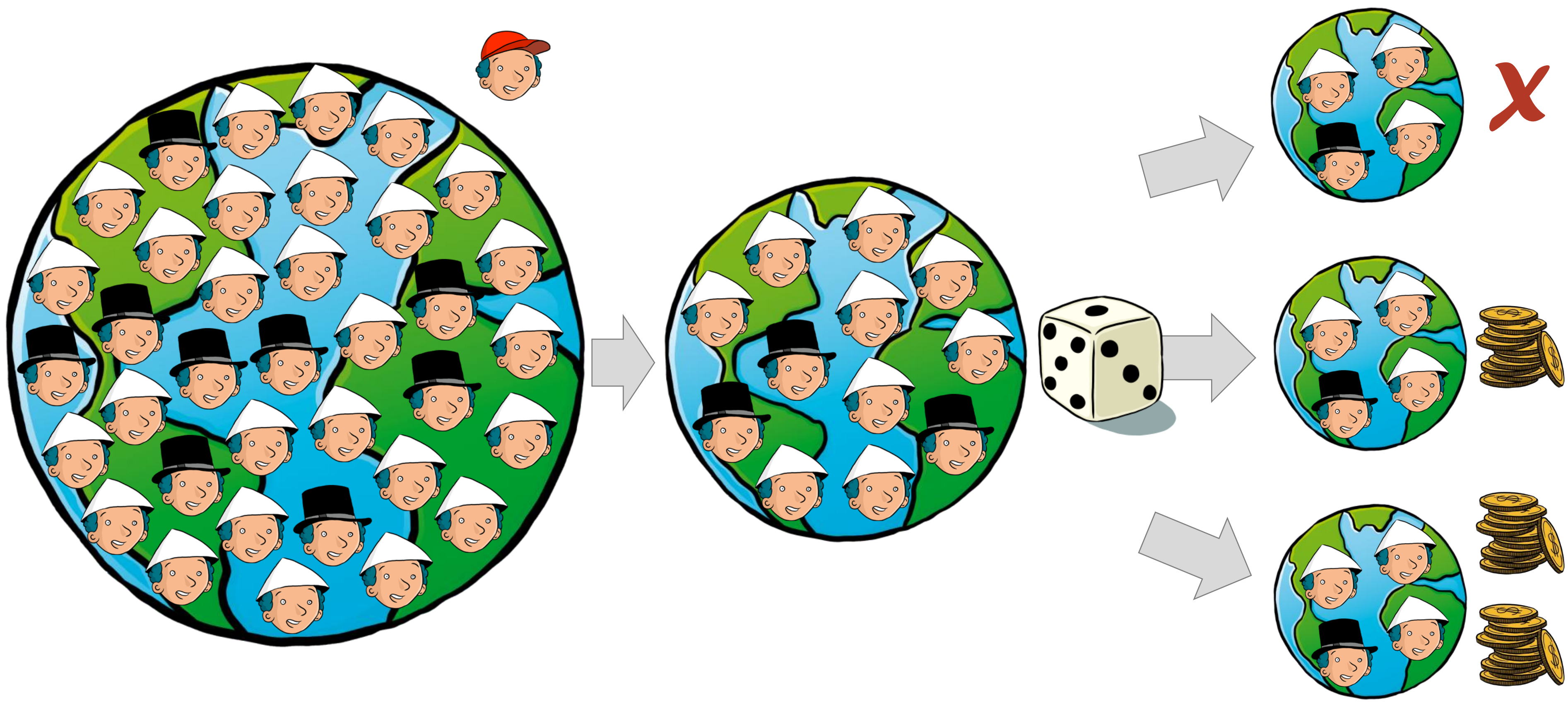
Comparison	Low Benefit	High Benefit
✗		

Randomized assignment with different benefit levels

1. Eligible Population

2. Evaluation sample

3. Randomize treatment
(2 benefit levels)



= Ineligible

= Eligible

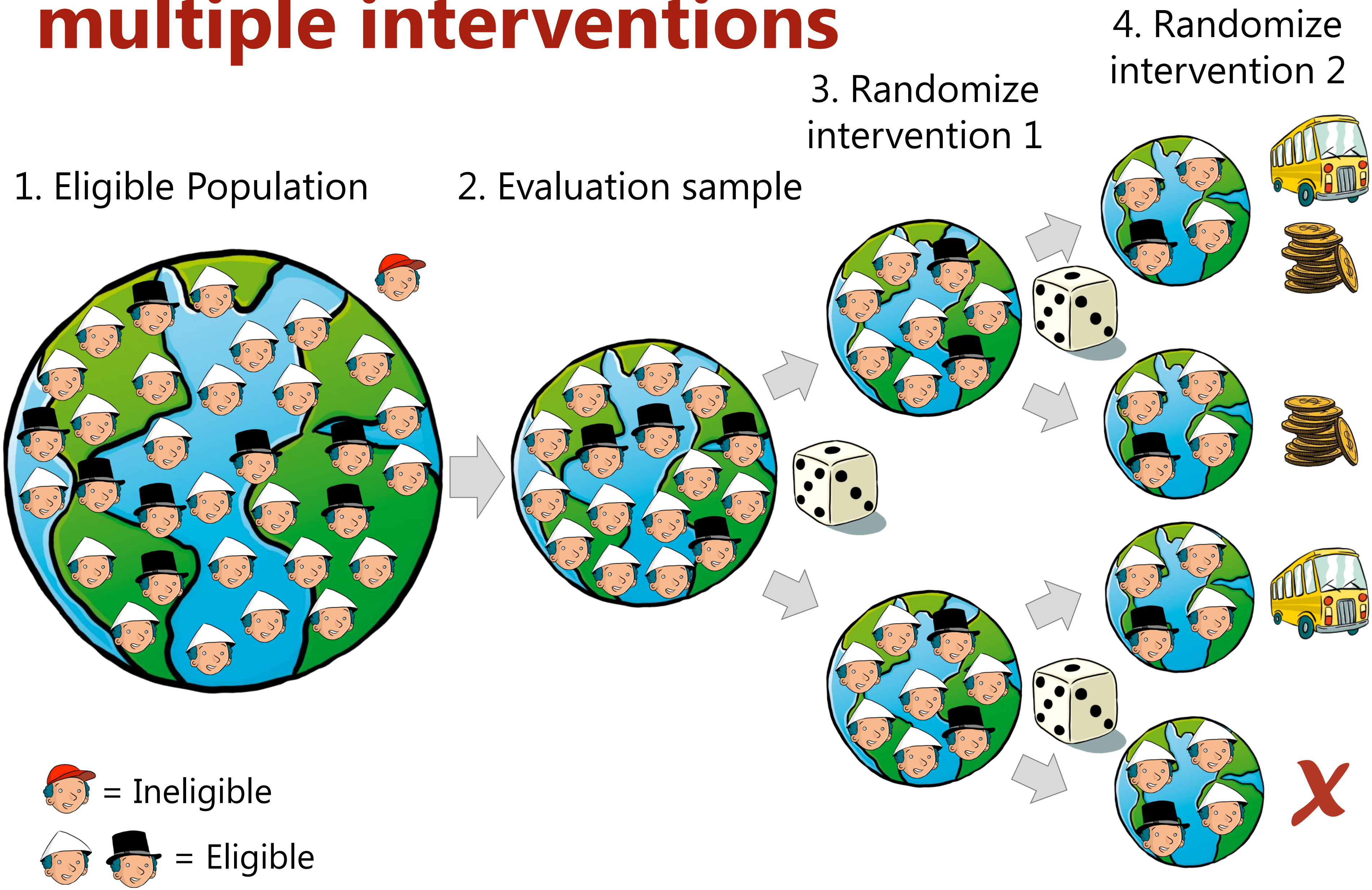
Randomized assignment with multiple interventions

- Other key policy question for a program with various benefits:
  - What is the impact of an intervention compared to another?
  - Are there complementarities between various interventions?

- Randomized assignment with 2 benefit packages:

		Intervention 1	
		Treatment	Comparison
Intervention 2	Treatment	Group A 	Group C 
	Comparison	Group B 	Group D X

Randomized assignment with multiple interventions



Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

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 or 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
  - But provide additional promotion, encouragement or incentives to a random sub-sample:
    - Additional Information.
    - Encouragement.
    - Incentives (small gift or prize).
    - Transport (bus fare).

  → Randomized promotion

HUMAN DEVELOPMENT NETWORK

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.

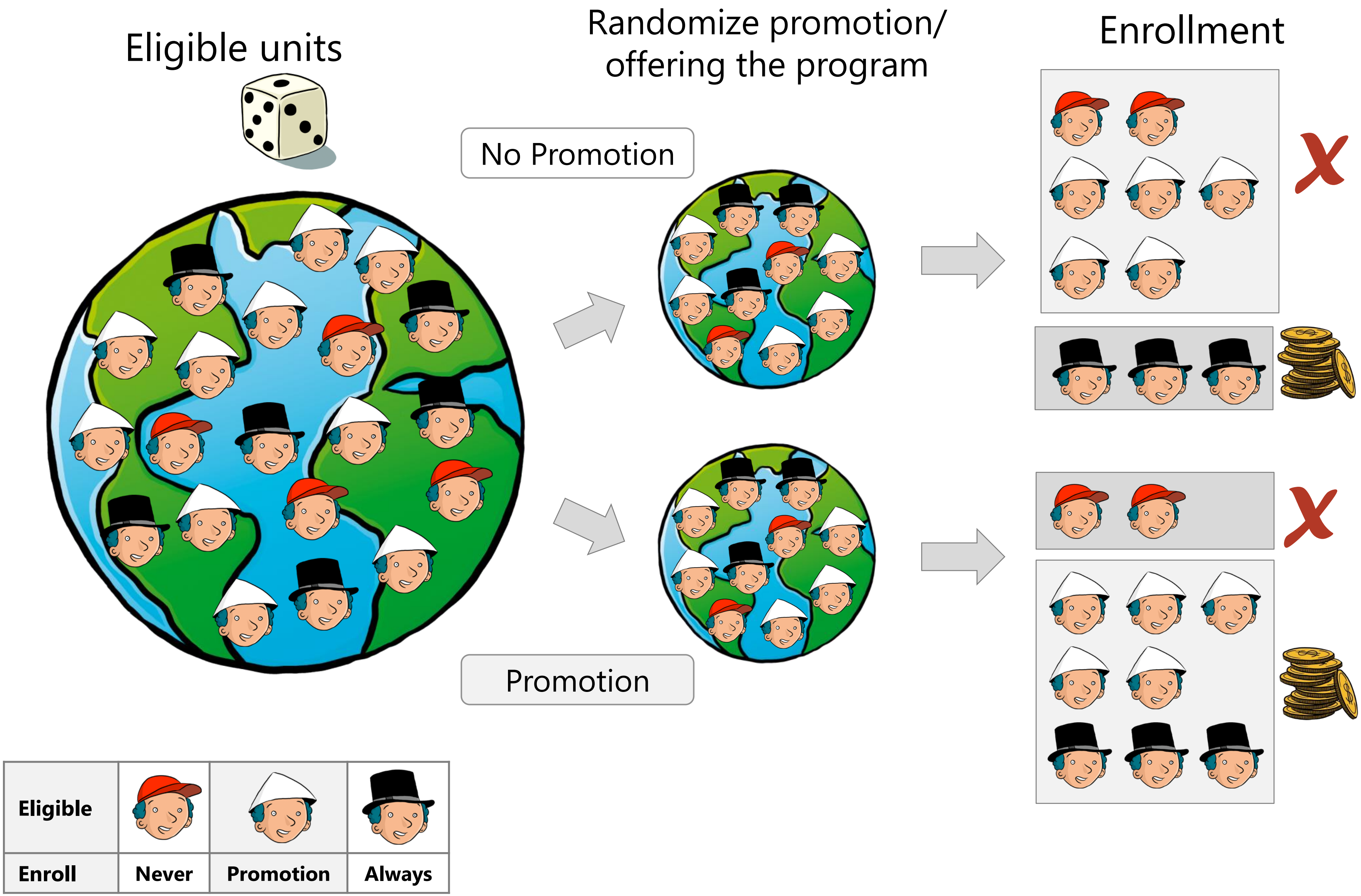
HUMAN DEVELOPMENT NETWORK

Randomly offering or promoting program

3 groups of units/individuals

	WITH promotion	WITHOUT promotion
Never Enroll		
Only Enroll if Encouraged		
Always Enroll		

Randomly offering or promoting program



Randomly offering or promoting program

	Promoted Group	**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 Encouraged			
Always Enroll			-

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)*

HUMAN DEVELOPMENT NETWORK

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.

HUMAN DEVELOPMENT NETWORK

Case 4: Randomized Promotion

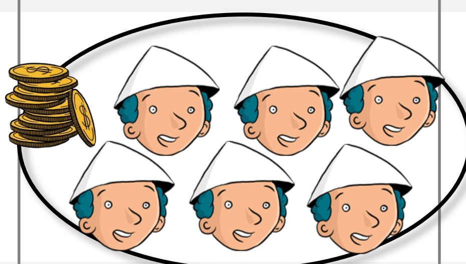
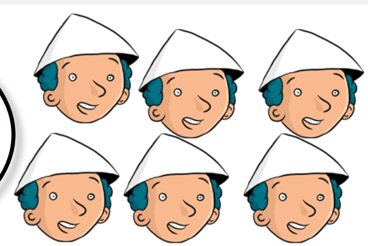
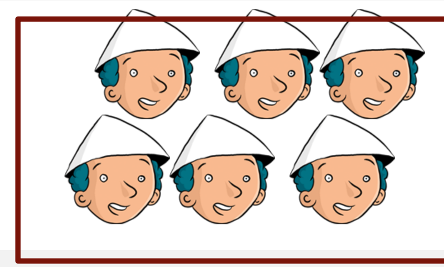
- Randomized Promotion is an "Instrumental Variable" (IV)
  - A variable correlated with treatment but nothing else (i.e. randomized promotion)
  - Use 2-stage least squares (see annex)
- When you randomly choose the units to which you offer the treatment but have less than 100% take-up
  - Using this method is equivalent to estimating the effect of "treatment on the treated"
  - How?

    "promoted" group = group offered treatment.

    "not promoted" group = group not offered treatment.

HUMAN DEVELOPMENT NETWORK

Case 4: Progresa Randomized Promotion

	Promoted Group	Not Promoted 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 Encouraged			
Always Enroll			

Case 4: Randomized Promotion

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 (**).

HUMAN DEVELOPMENT NETWORK

Keep in Mind

!

Randomized 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 promotion.

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

HUMAN DEVELOPMENT NETWORK

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

Discontinuity Design

Many social programs select beneficiaries using an index or score:

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

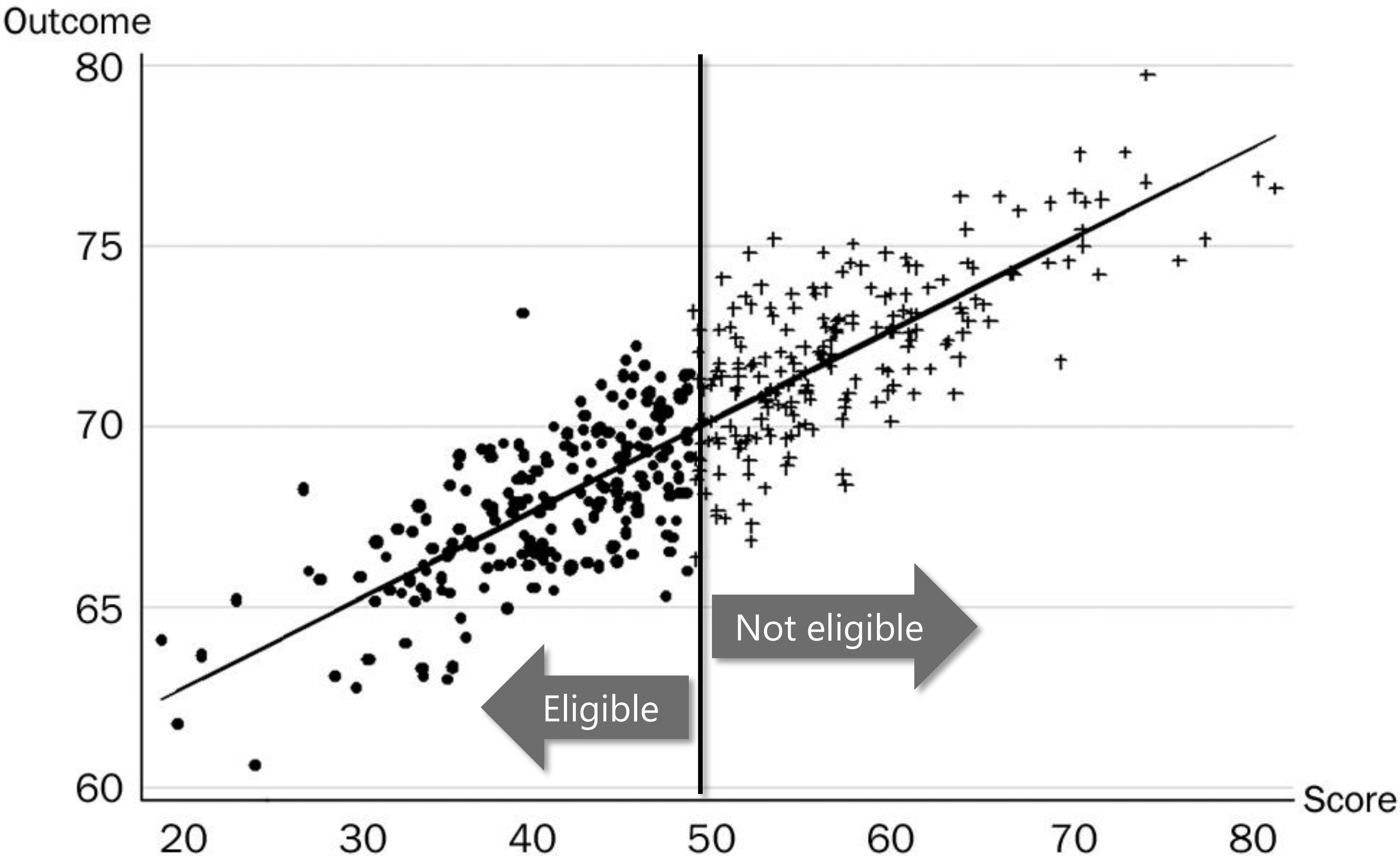
- Farms with a score (Ha) of land ≤50 are small
- Farms with a score (Ha) of land >50 are not small

Intervention

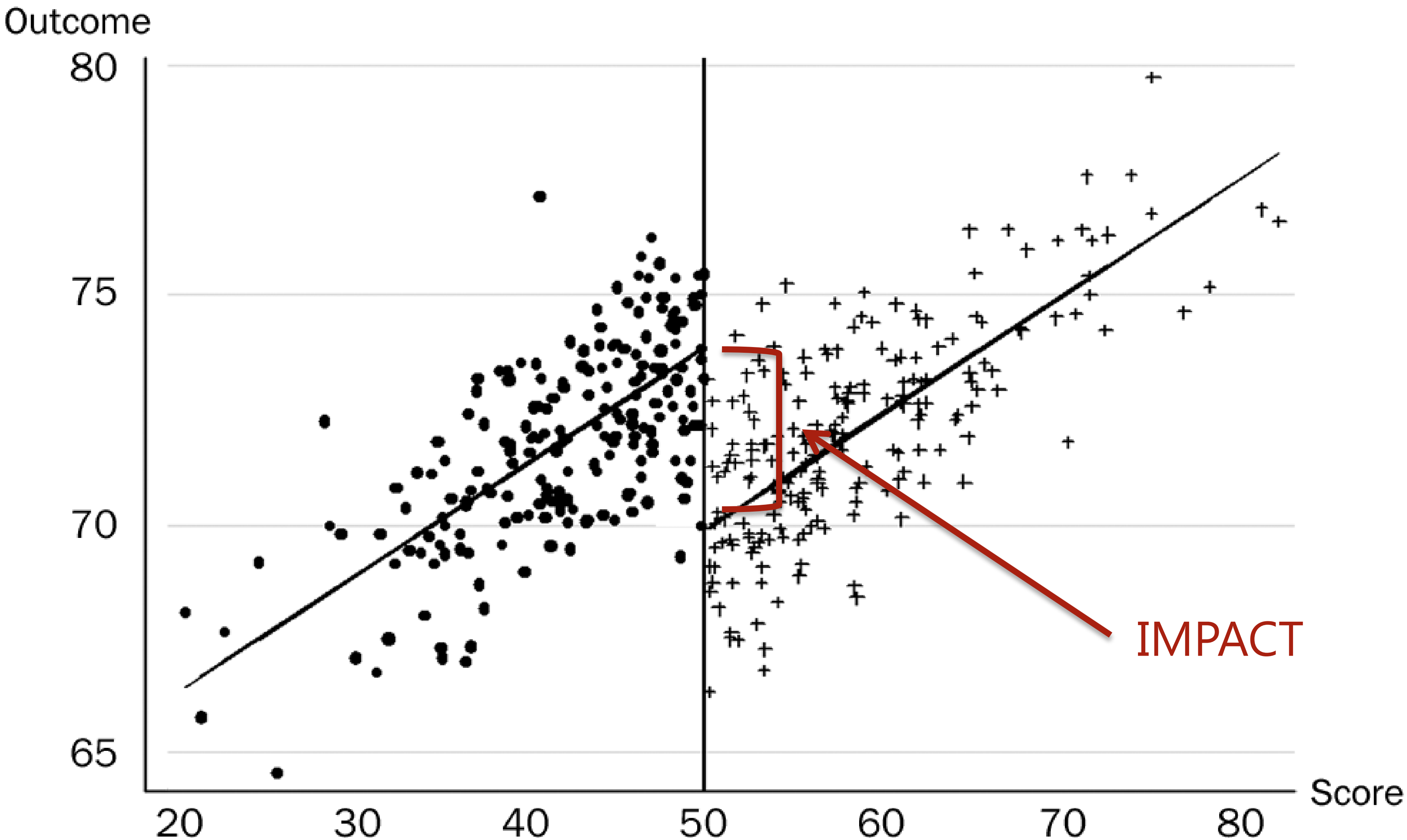
Small farmers receive subsidies to purchase fertilizer

HUMAN DEVELOPMENT NETWORK

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.

HUMAN DEVELOPMENT NETWORK

Case 5: Discontinuity Design

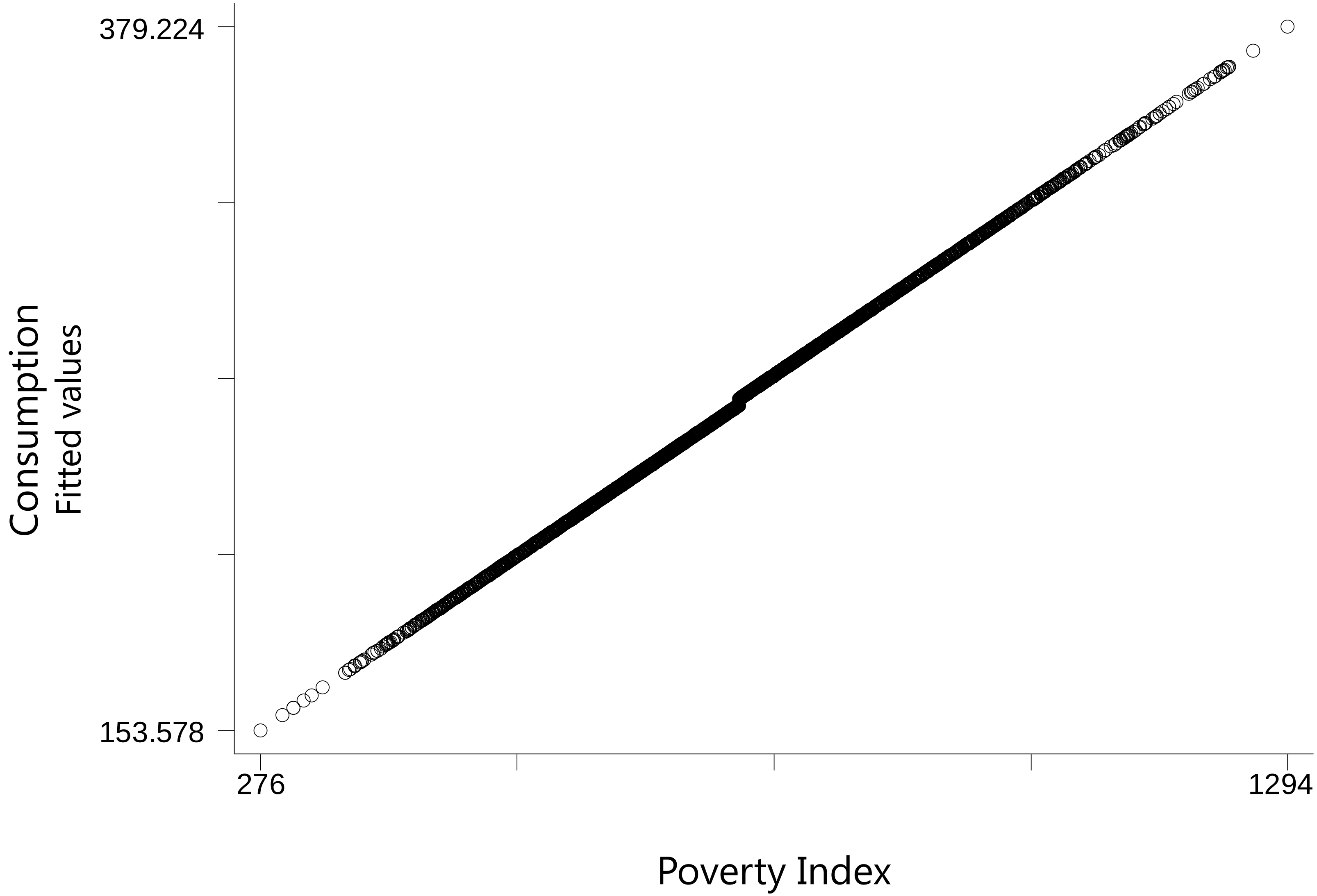
- Eligibility for Progresa is based on national poverty index

- Household is poor if score ≤ 750

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

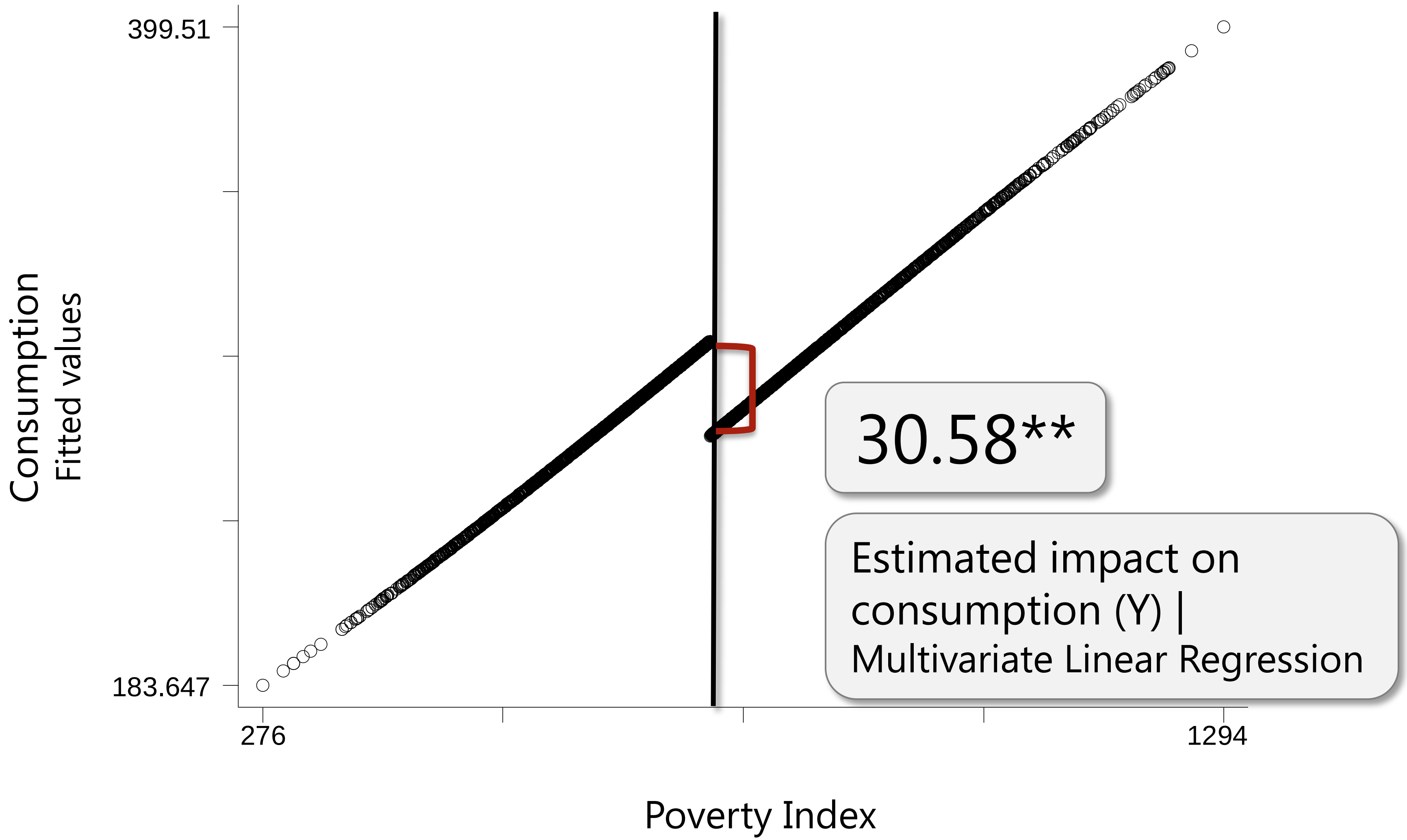
Case 5: Discontinuity Design

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.

HUMAN DEVELOPMENT NETWORK

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 Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

Difference-in-differences

(Diff-in-diff)

Y=Girl's school enrollment
P=Education enhancement program

	Enrolled	Not Enrolled	
After	0.74	0.81	
Before	0.60	0.78	
Difference	**+0.14** -	**+0.03**	**= 0.11**

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

HUMAN DEVELOPMENT NETWORK

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

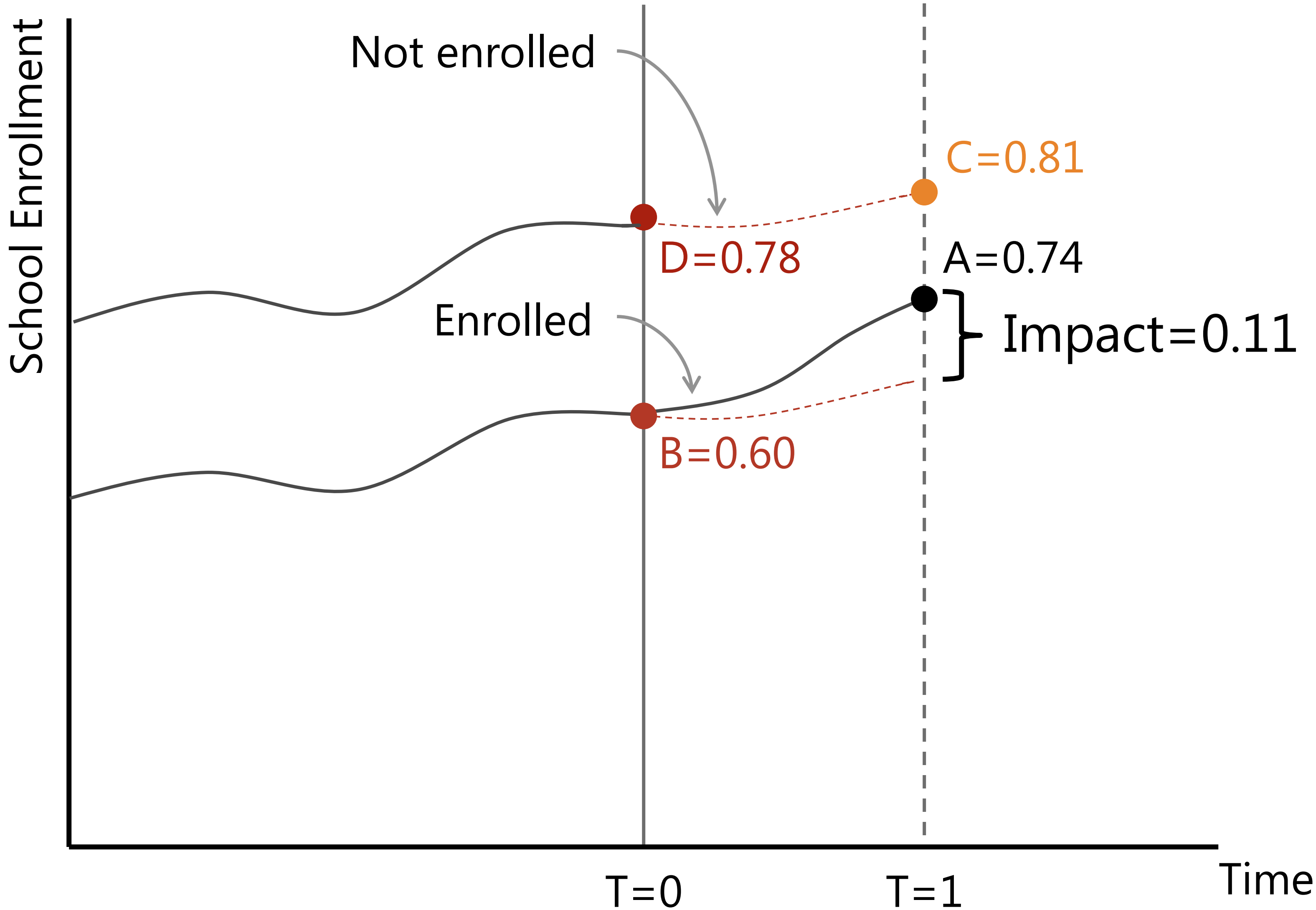
Y=Yield of soybeans, *tons per acre*
P=New type of inoculant

	Enrolled		Not Enrolled	Difference
After	0.74	-	0.81	**-0.07**
				-
Before	0.60	-	0.78	**-0.18**
				=
				0.11

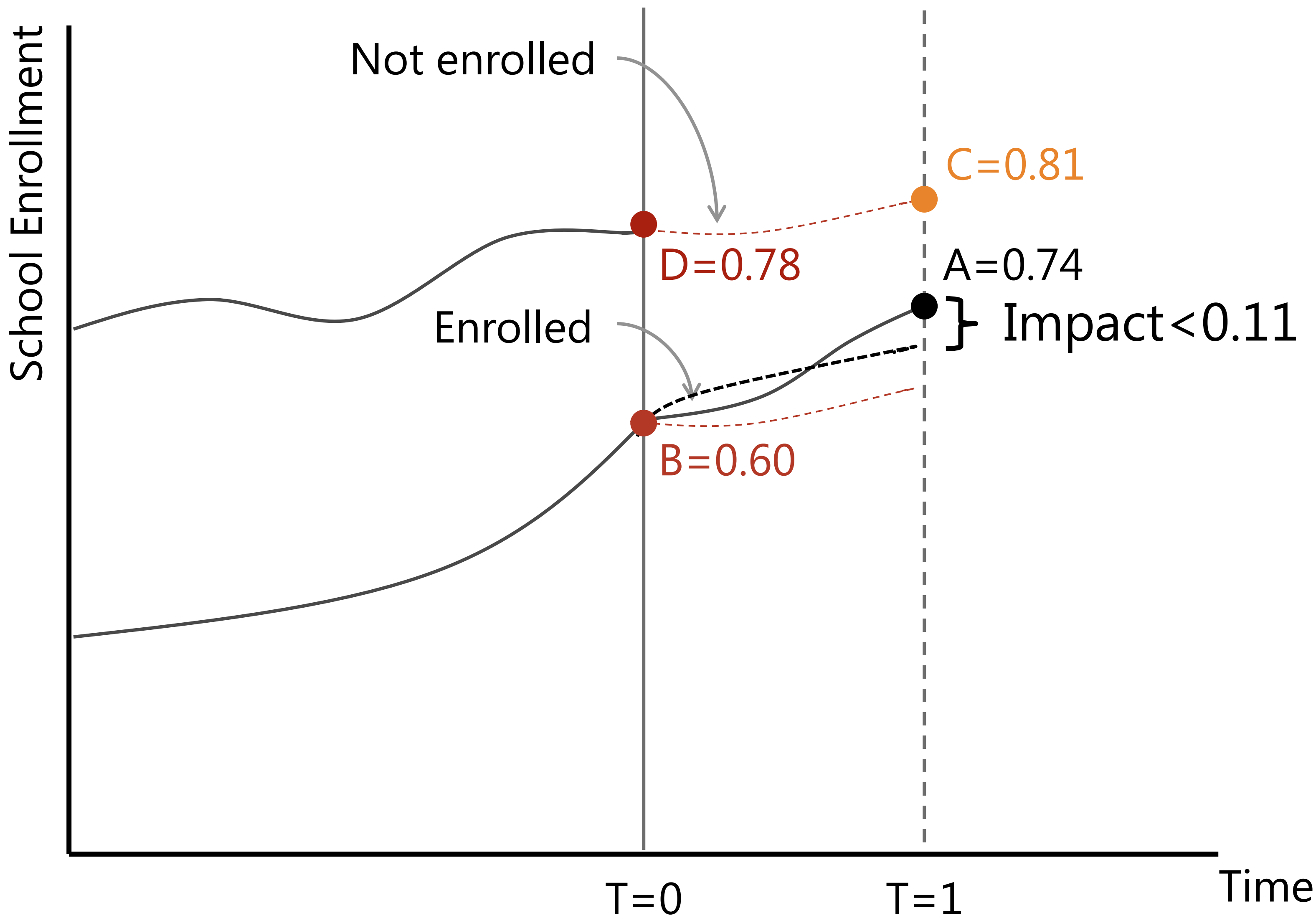
Diff-in-Diff: Impact=$(Y_{t1}-Y_{c1})-(Y_{t0}-Y_{c0})$

HUMAN DEVELOPMENT NETWORK

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



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



Case 6: Difference in difference

	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 (**).

HUMAN DEVELOPMENT NETWORK

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 Promotion	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

Differences 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**.

HUMAN DEVELOPMENT NETWORK

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

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!

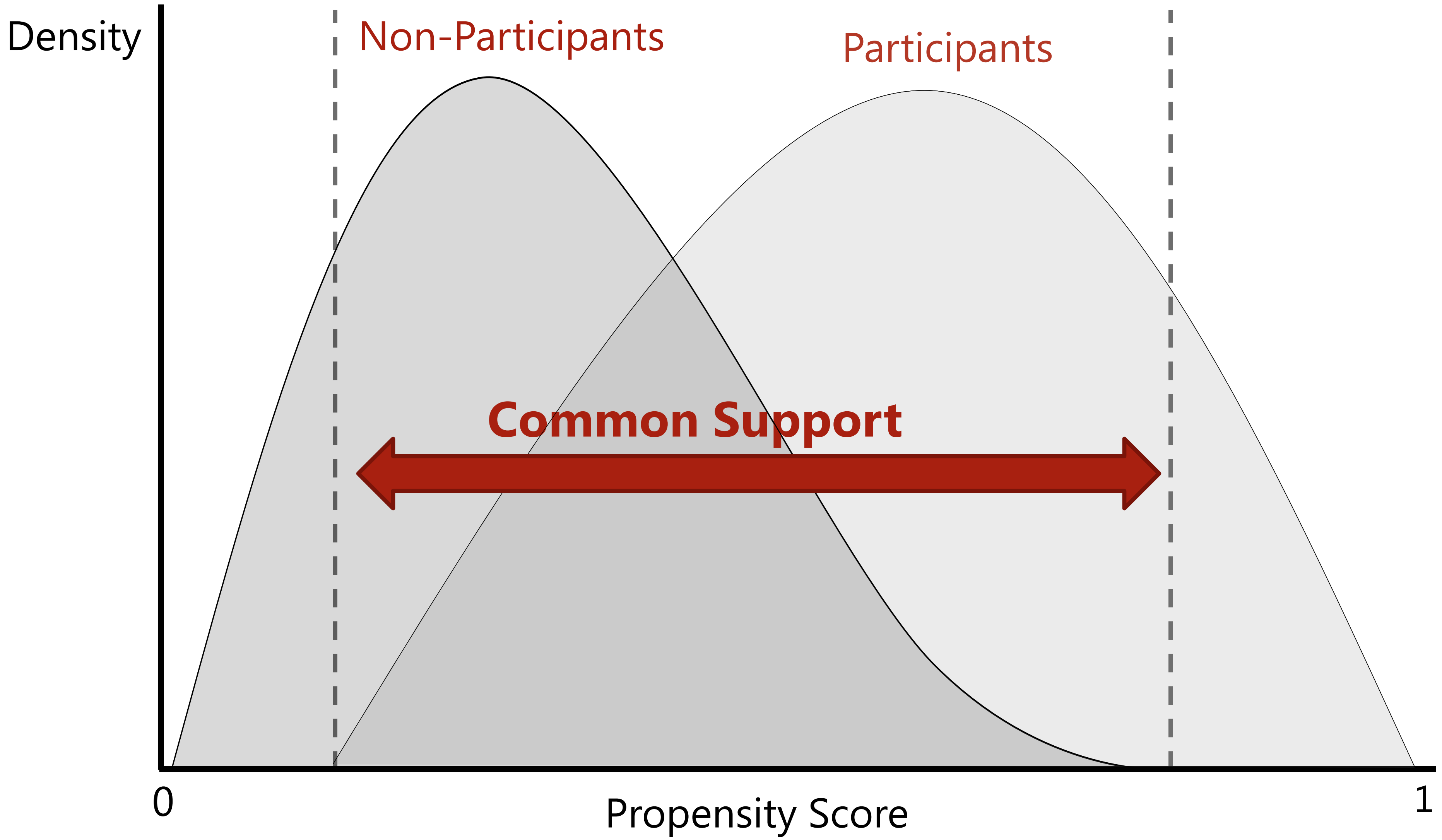
Propensity-Score Matching *(PSM)*

- 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.

HUMAN DEVELOPMENT NETWORK

Density of propensity scores

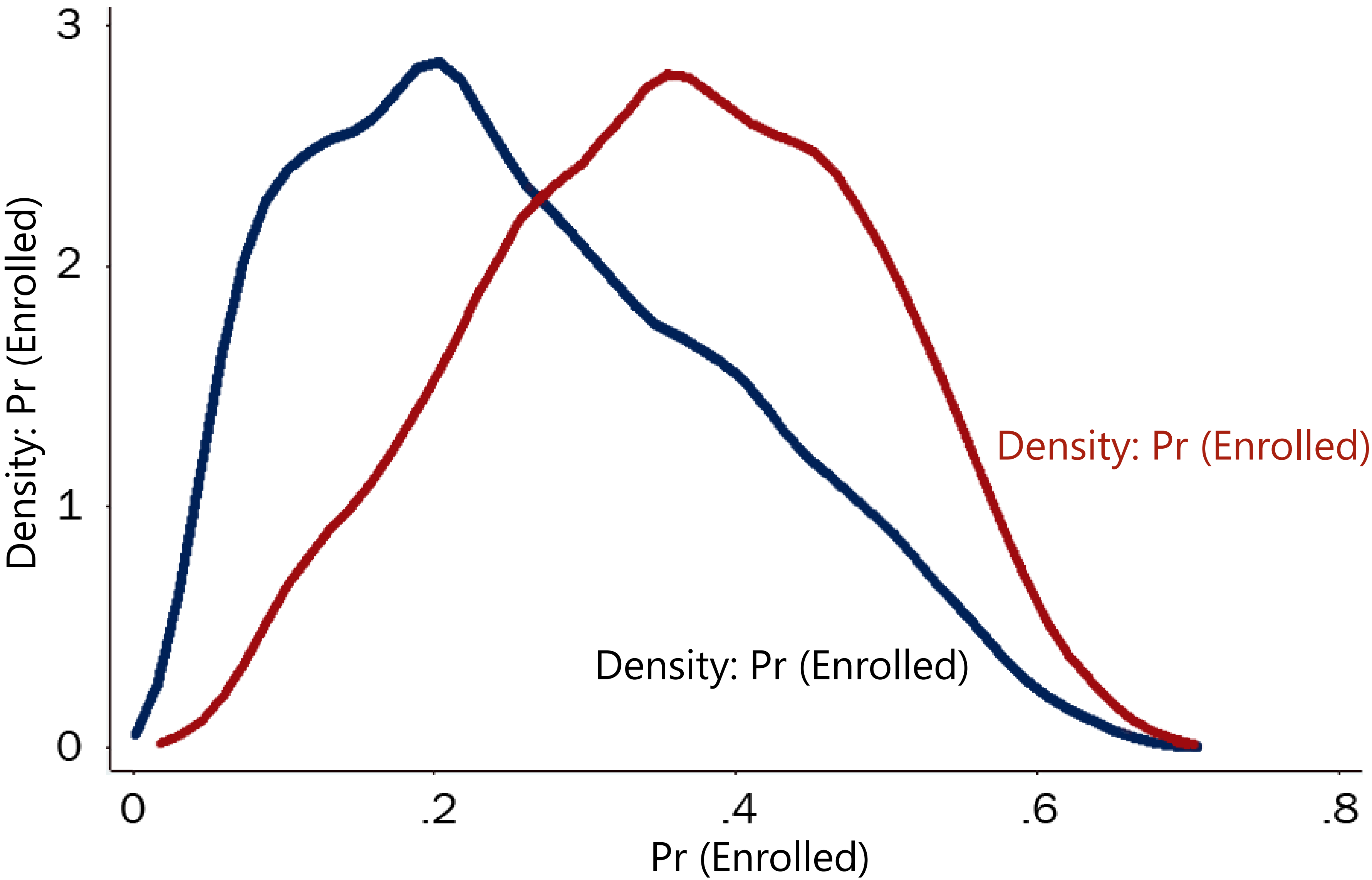


Case 7: Progresa Matching *(P-Score)*

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 +

HUMAN DEVELOPMENT NETWORK

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.

HUMAN DEVELOPMENT NETWORK

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 Promotion	30.4**
Case 5: Discontinuity Design	30.58**
Case 6: Differences 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 +

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 Promotion	30.4**
Case 5: Discontinuity Design	30.58**
Case 6: Differences 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 +

HUMAN DEVELOPMENT NETWORK

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

Combinations of methods

IE Methods Toolbox

HUMAN DEVELOPMENT NETWORK

Where Do Comparison Groups come from?

- The rules of program operation determine the evaluation strategy.

- We can almost always find a valid comparison group if:
  - ❑ the operational rules for selecting beneficiaries are equitable, transparent and accountable;
  - ❑ the evaluation is designed prospectively.

HUMAN DEVELOPMENT NETWORK

Operational rules and prospective designs

- Use opportunities to generate good comparison groups and ensure baseline data is collected.
- 3 questions to determine which method is appropriate for a given program
  - Money: Does the program have sufficient resources to achieve scale and reach full coverage of all eligible beneficiaries?
  - Targeting Rules: Who is eligible for program benefits? Is the program targeted based on an eligibility cut-off or is it available to everyone?
  - Timing: How are potential beneficiaries enrolled in the program – all at once or in phases over time?

HUMAN DEVELOPMENT NETWORK

Choosing your IE method(s)

Money →	***Excess demand***		***No Excess demand***	
Targeting → *Timing* ↓	***Targeted***	***Universal***	***Targeted***	***Universal***
Phased Roll-out	1 Randomized assignment 4 RDD	1 Randomized assignment 2 Randomized promotion 3 DD with 5 Matching	1 Randomized Assignment 4 RDD	1 Randomized assignment to phases 2 Randomized Promotion to early take-up 3 DD with 5 matching
Immediate Roll-out	1 Randomized Assignment 4 RDD	1 Randomized Assignment 2 Randomized Promotion 3 DD with 5 Matching	4 RDD	If less than full Take-up: 2 Randomized Promotion 3 DD with 5 Matching

HUMAN DEVELOPMENT NETWORK

Remember

"The objective of impact evaluation is to estimate the **causal** effect or **impact** of a program on outcomes of interest.

HUMAN DEVELOPMENT NETWORK

Remember

"To estimate impact, we need to estimate the **counterfactual.**

- what would have happened in the absence of the program and
- use comparison or control groups.

HUMAN DEVELOPMENT NETWORK

Remember

We have a **toolbox** with **5 methods** to identify good comparison groups.

Remember

“

Choose the best evaluation method that is feasible in the program’s operational context.

Reference

Spanish Version
& French Version
also available

www.worldbank.org/ieinpractice

HUMAN DEVELOPMENT
NETWORK

Thank You

Q & A

Appendix 1
Two Stage Least Squares *(2SLS)*

- Model with endogenous *Treatment (T)*:

$$y = \alpha + \beta_1 T + \beta_2 x + \varepsilon$$

- Stage 1: Regress endogenous variable on the IV (*Z*) and other exogenous regressors:

$$T = \delta_0 + \delta_1 x + \theta_1 Z + \tau$$

- Calculate predicted value for each observation: *T hat*

HUMAN DEVELOPMENT NETWORK

Appendix 1
Two Stage Least Squares *(2SLS)*

- Stage 2: Regress outcome y on predicted variable (and other exogenous variables):

$$y = \alpha + \beta_1(\hat{T}) + \beta_2 x + \varepsilon$$

- Need to correct Standard Errors (they are based on *T hat* rather than *T*)
- In practice just use STATA – ivreg.
- Intuition: *T* has been "cleaned" of its correlation with ε.

HUMAN DEVELOPMENT NETWORK

Appendix 2
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.

HUMAN DEVELOPMENT NETWORK