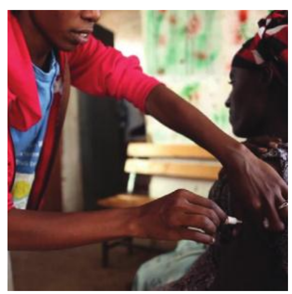
Introduction to Impact Evaluation of RBF Programs

Damien de Walque | Gil Shapira



THE WORLD BANK

HEALTH RESULTS INNOVATION TRUST FUND

RBF for Health Impact Evaluation

- Build evidence on what works, what doesn't and why

- RBF for Health impact evaluations characteristics
  - Built into program operations
  - Government ownership
  - Feedback loop for evidence-based decision making
  - Valid Treatment and Control Groups

Policy questions we are interested to answer

<u>Does RBF work?</u>

- What is the impact of RBF on:
  - Utilization of services?
  - Health outcomes?
- Does it impact differently different populations?
- Are there unintended consequences of RBF?
- Is RBF cost effective relative to other interventions?

Policy questions we are interested to answer

How can RBF work better?

- What components of an RBF "package" matter most:
  - Performance incentives? Increased financing? Autonomy? Improved supervision?
- What are the right incentives?
- Who should be incentivized? Providers? Households? Communities?
- How to reduce reporting errors and corruption?
- What are the optimal provider capabilities?
- What are the key organizational building blocks to make RBF work?

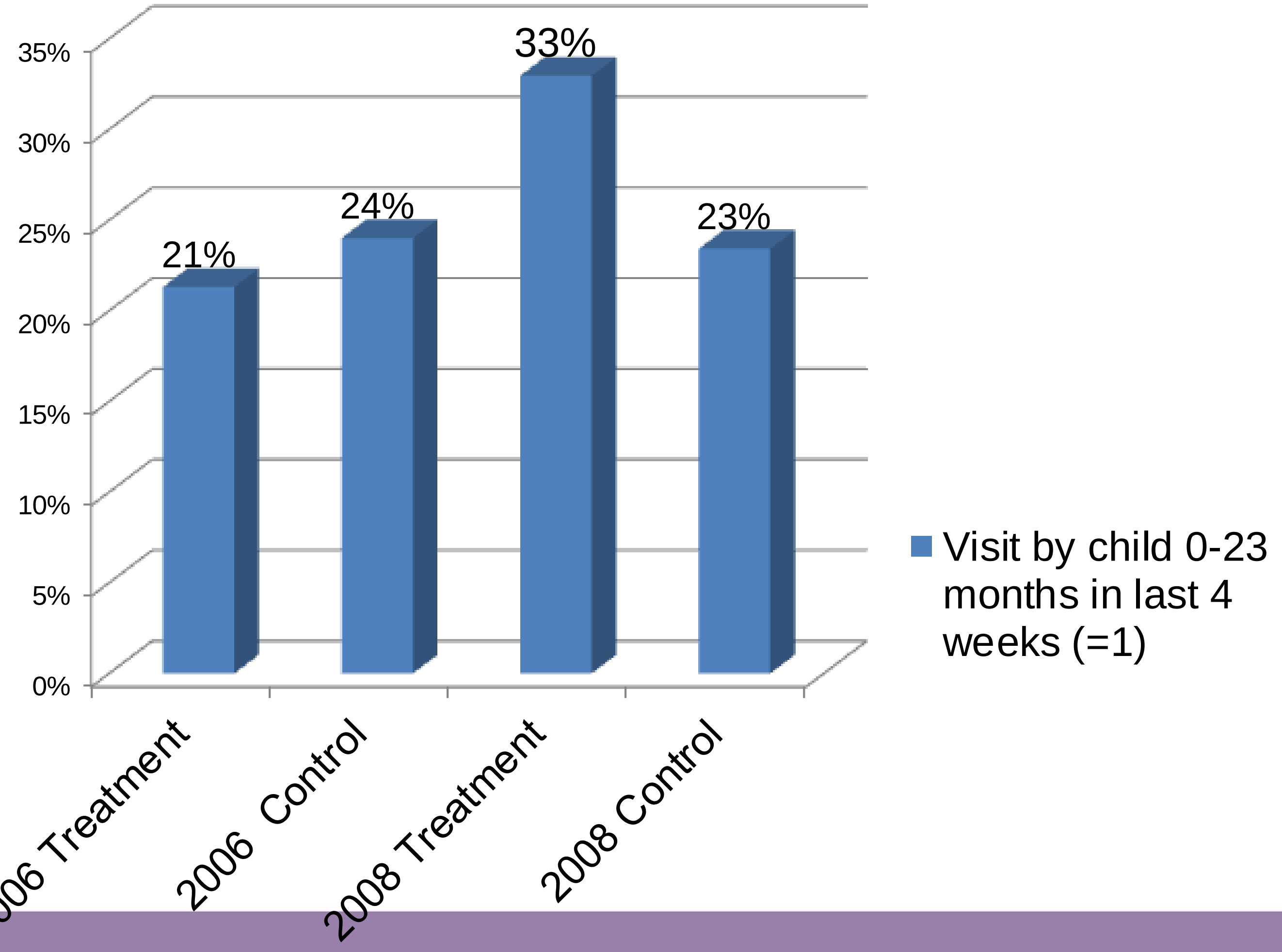
An Example:

The Impact Evaluation of the Rwanda Performance-Based Financing Project

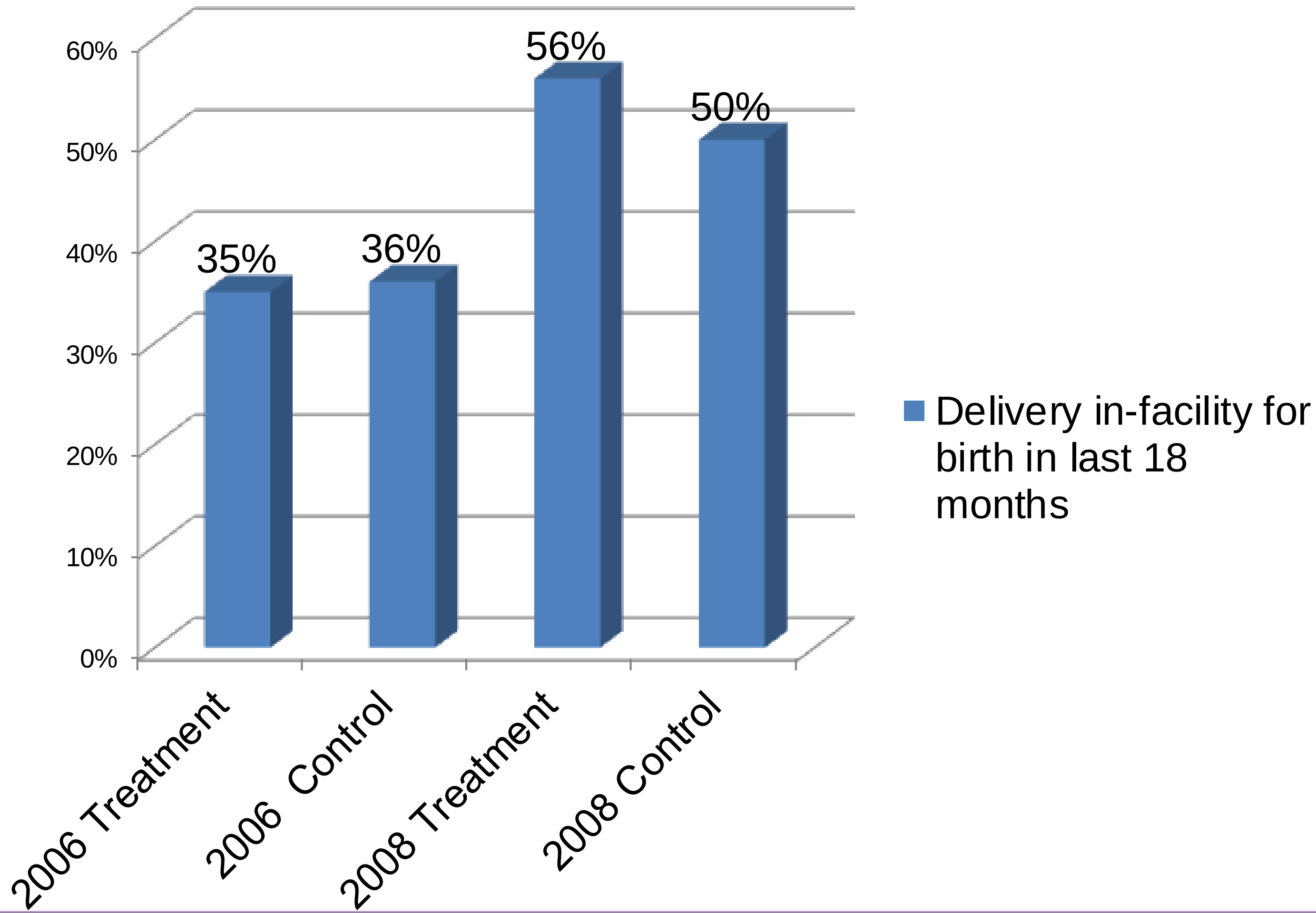
Rwanda Performance-Based Financing project (Basinga et al. 2011)

- Improved prenatal care quality (+0.16 std dev), increased utilization of skilled delivery (+8.1pp) and child preventive care services (+11 pp)
- No impact on timely prenatal care
- Greatest effect on services that are under the provider control and had the highest payment rates
- Financial performance incentives can improve both use of and quality of health services.
- An equal amount of financial resources without the incentives would not have achieved the same gain in outcomes.

Impact of Rwanda PBF on Child Preventive Care Utilization



Impact of Rwanda PBF on Institutional delivery



Rwanda Performance-Based Financing project (Gertler & Vermeersch forthcoming)

- No impact on family planning
- Large impacts on child health outcomes (weight 0-11 months, height 24-47 months)
- Impacts are larger for better skilled providers
- PBF worked through incentives, not so much through increased knowledge

Measuring Impact

Impact Evaluation

Methods for Policy Makers

Slides by Sebastian Martinez, Christel Vermeersch and Paul Gertler. We thank Patrick Premand and Martin Ruegenberg for contributions. The content of this presentation reflects the views of the authors and not necessarily those of the World Bank.

Impact Evaluation

- Logical Framework → How the program works *in theory*
- Measuring Impact → Identification Strategy
- Data
- Operational Plan
- Resources

1

Causal Inference

Counterfactuals

False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled
(Apples & Oranges)

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

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 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 ($P=0$)

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!**

Estimating impact of *P* on *Y*

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

OBSERVE (Y | P=1)
Outcome with treatment

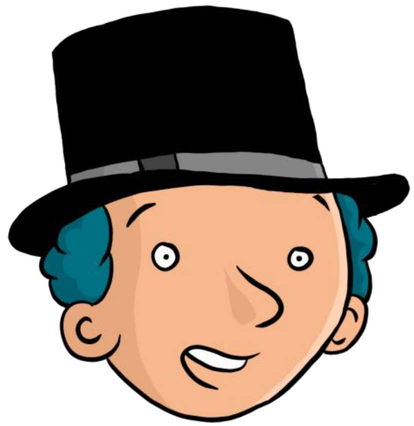
ESTIMATE (Y | P=0)
The Counterfactual

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

Example: What is the Impact of...

giving Fulanito 

additional pocket money  (P)

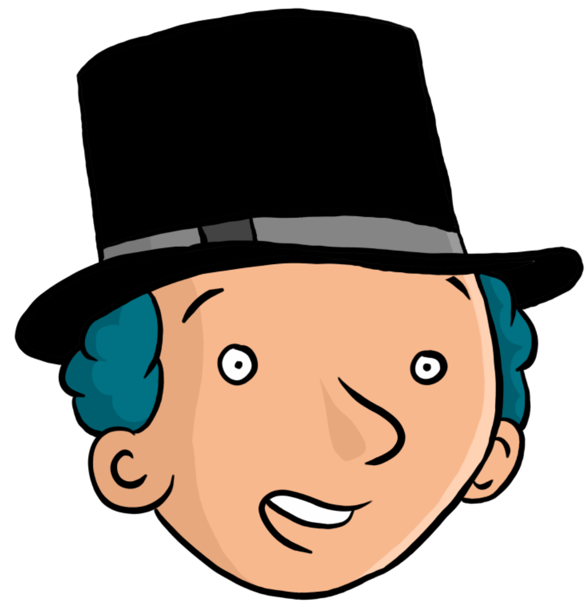
on Fulanito's consumption of candies  (Y)?

The Perfect *Clone*

Fulanito

6 candies

Fulanito's Clone

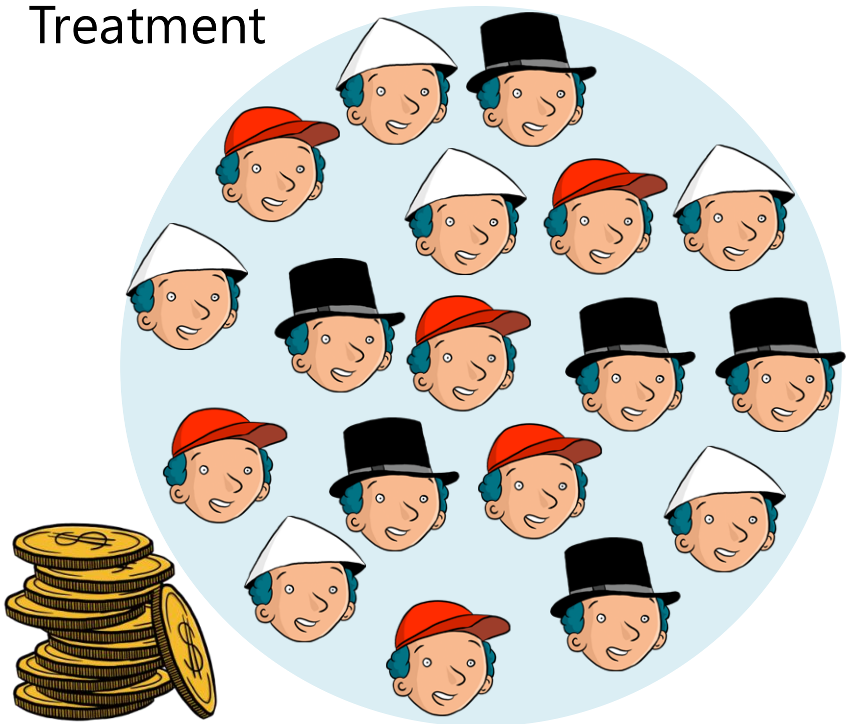


4 candies

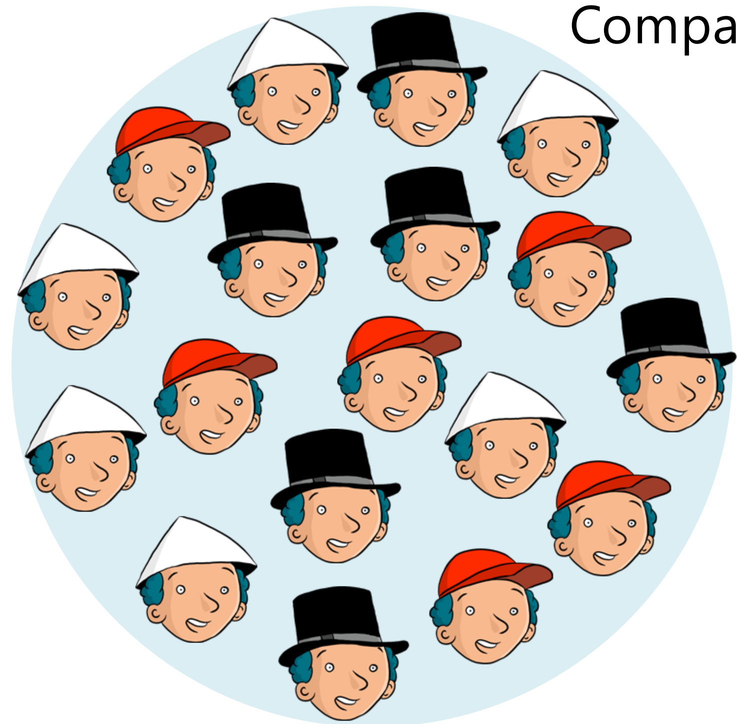
IMPACT=6-4=2 Candies

In reality, use statistics

Treatment



Comparison



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.

program eligibility & assignment

ct valid estimates of the

counterfactuals

With a good comparison group, the **only reason** for different outcomes between treatments and controls 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 2008

- Many outcomes of interest

  Here: Consumption per capita

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

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

Eligibility and Enrollment

1

Causal Inference

Counterfactuals

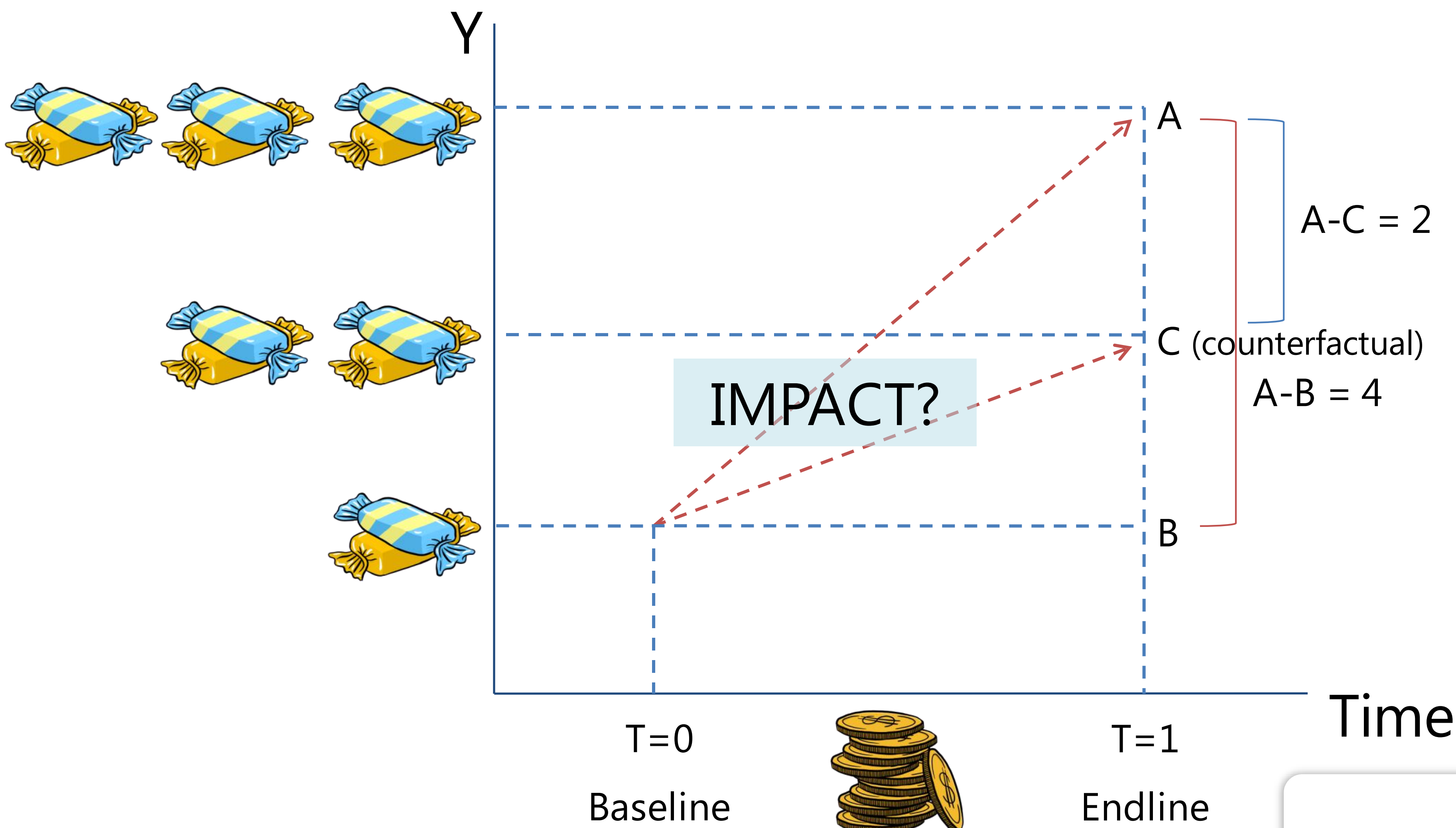
False Counterfactuals

Before & After (Pre & Post)

Enrolled & Not Enrolled (Apples & Oranges)

Counterfeit Counterfactual #1

Before & After

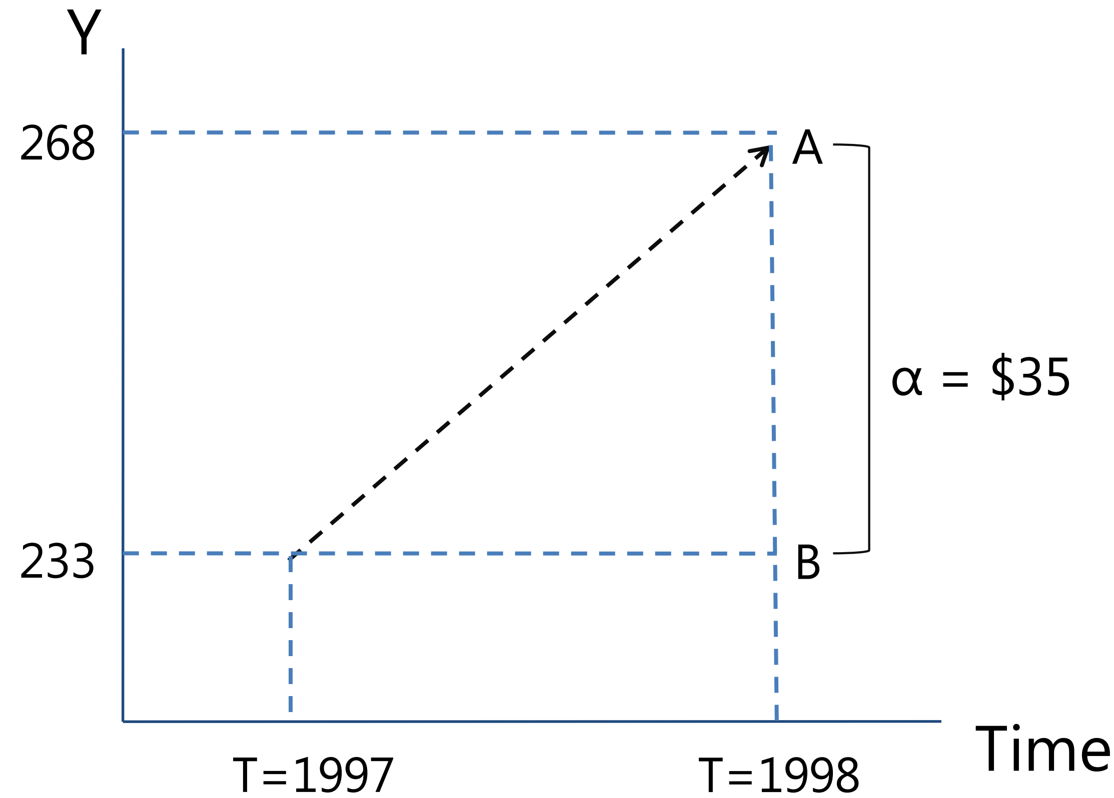


Case 1: Before & After

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 (**).

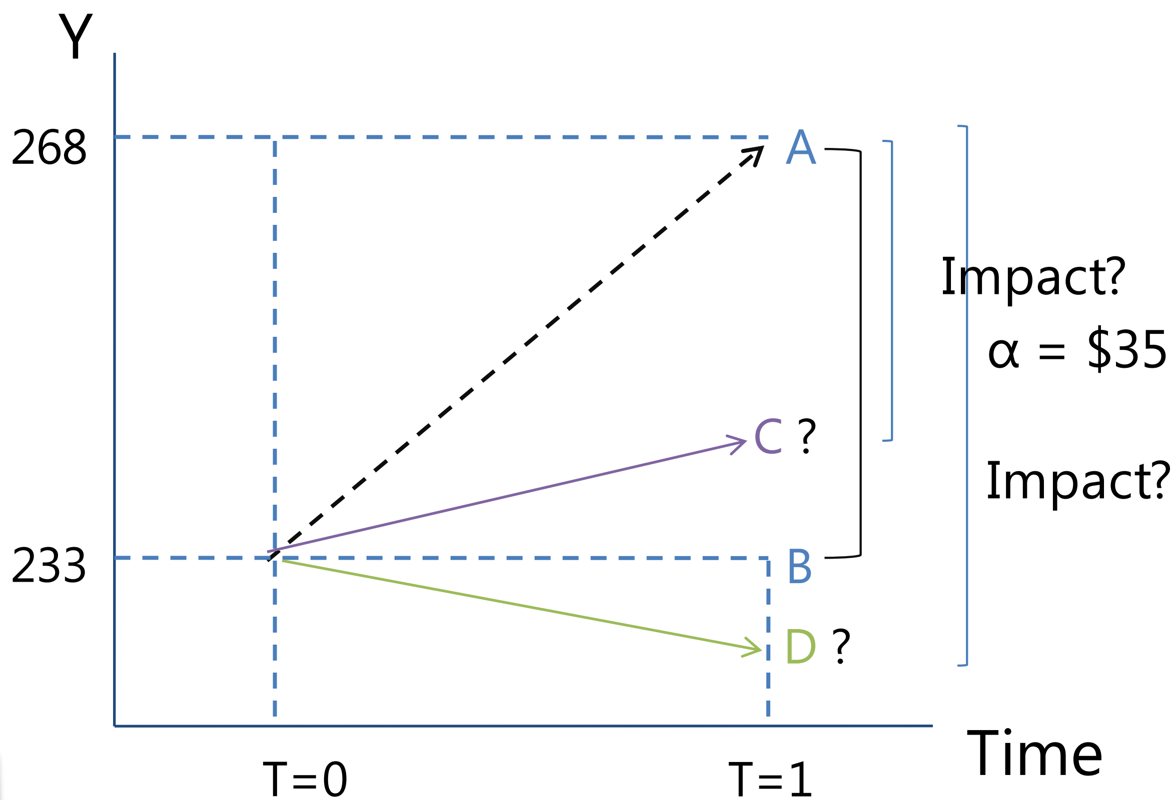
Case 1: What's the problem?

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: "control" group (counterfactual)

    Those ineligible to participate.

    Or 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 (**).

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&NE:
  - 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 counterfactual and the impact.

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

Randomized Treatments & Controls

Eligibles > Number of Benefits

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

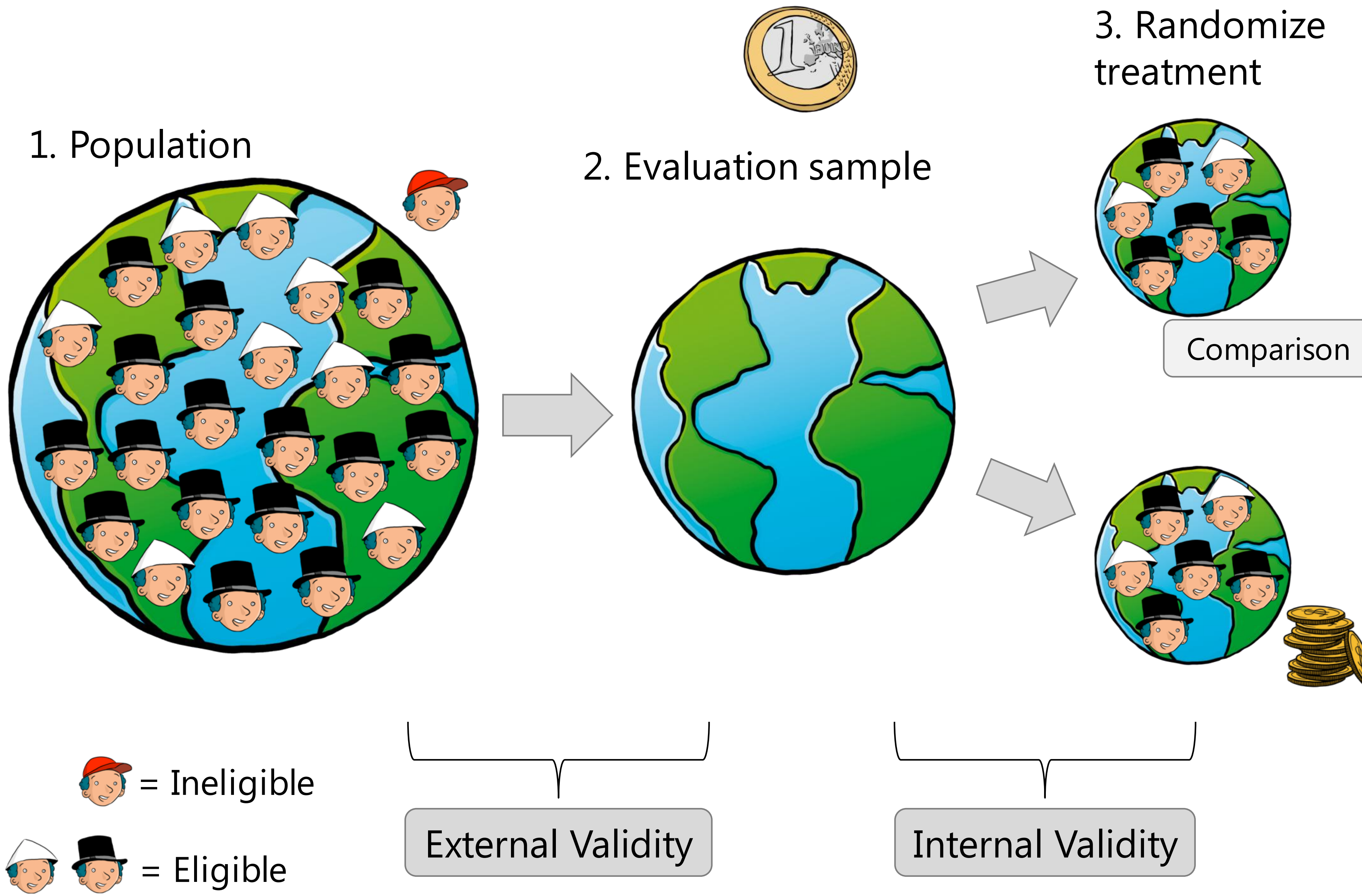
Oversubscription

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

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 *(controls)*.

Randomized treatments and controls



Unit of Randomization

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

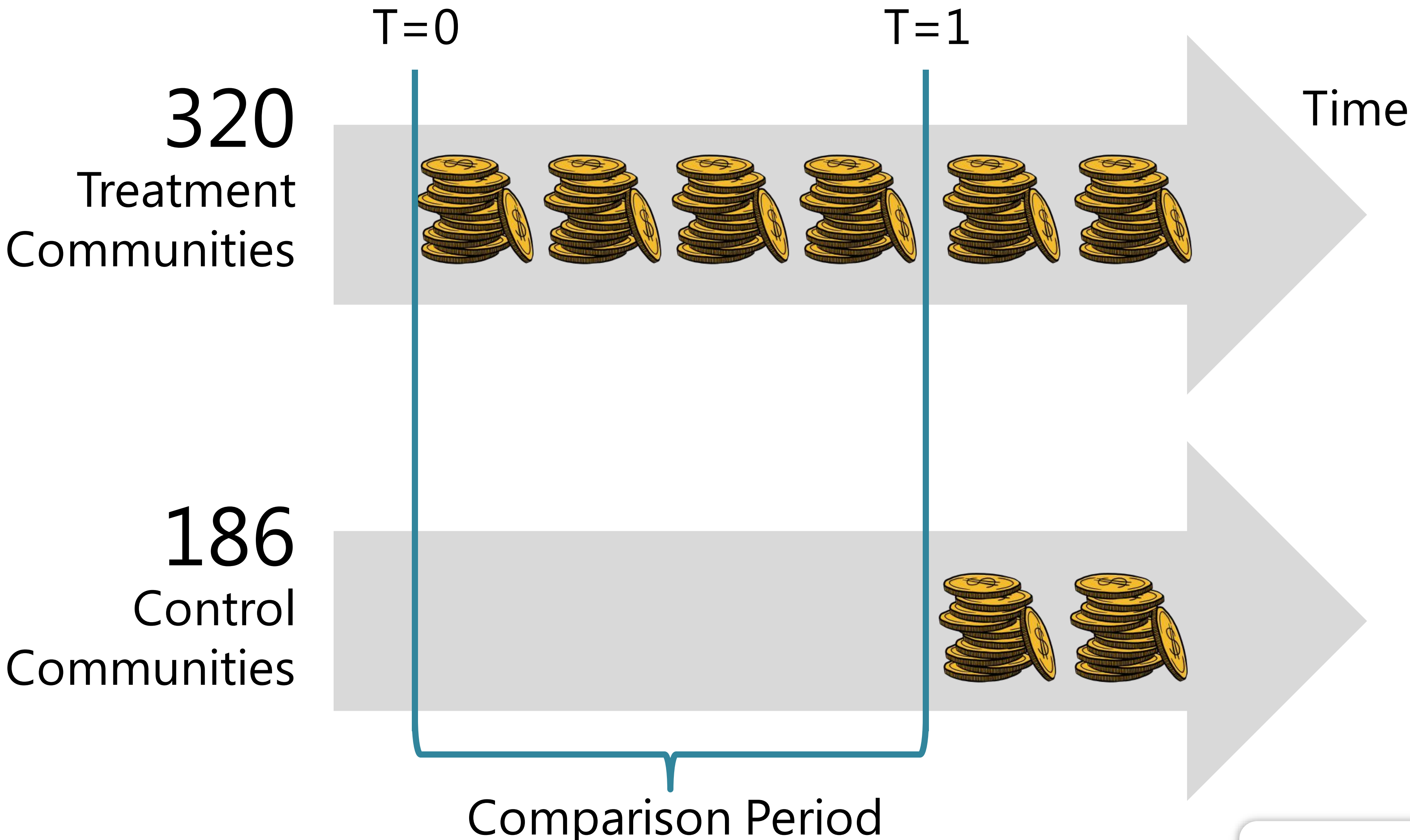
> As a rule of thumb, randomize at the smallest viable unit of implementation.

- 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 control 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			
	Control	Treatment	*T-stat*
Consumption ($ monthly per capita)	233.47	233.4	-0.39
Head's age (years)	42.3	41.6	1.2
Spouse's age (years)	36.8	36.8	-0.38
Head's education (years)	**2.8**	**2.9**	**-2.16****
Spouse's education (years)	2.6	2.7	-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			
	Control	Treatment	*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	0.56	0.57	-1.04
Hectares of Land	1.71	1.67	1.35
Distance to Hospital *(km)*	106	109	-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 (**).

Keep in Mind

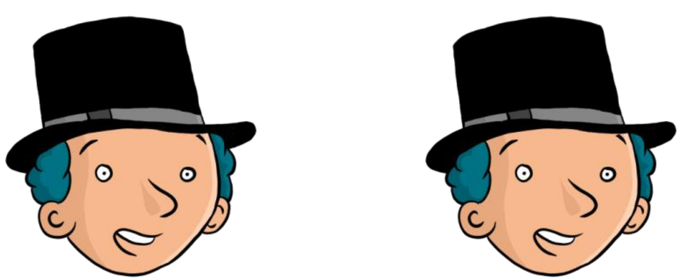
Randomized Assignment

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 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:

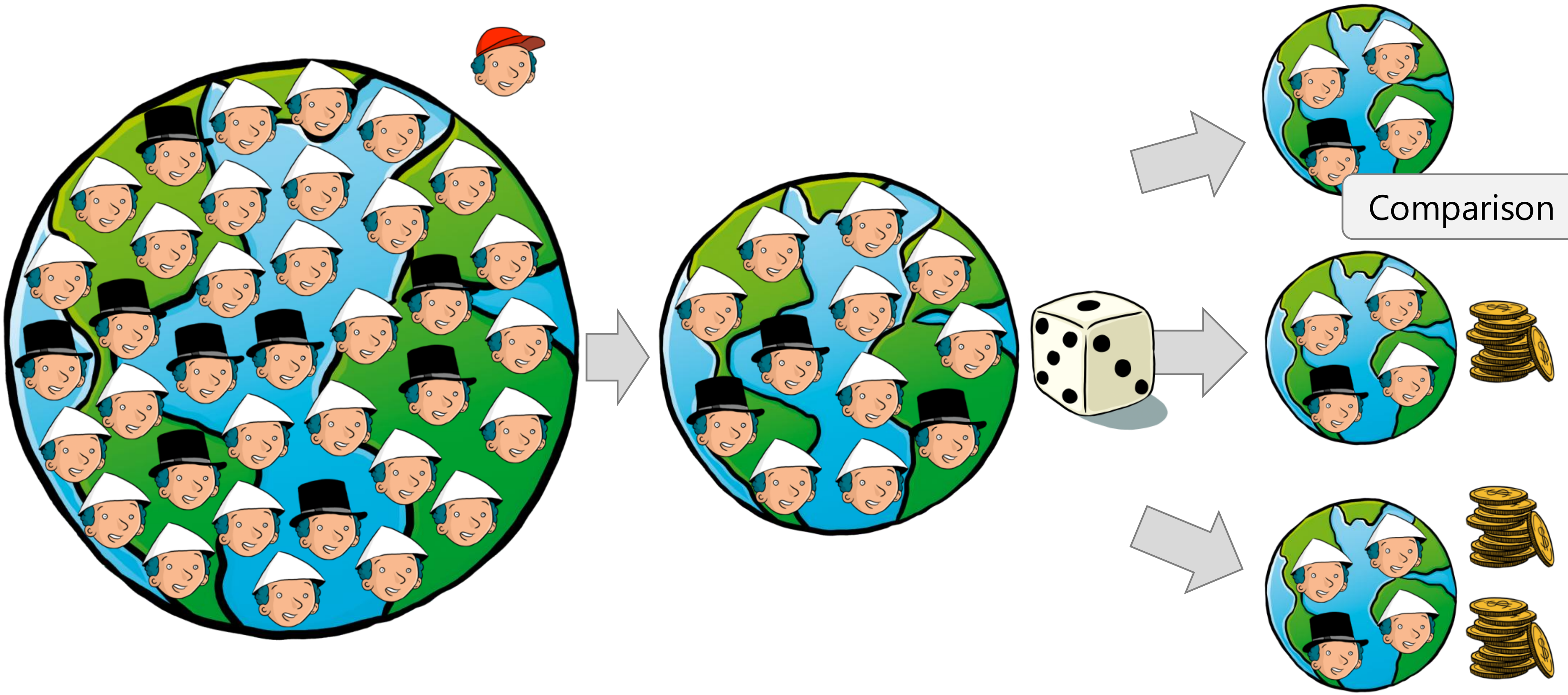
Comparison	Low Benefit	High Benefit
X		 

Randomized assignment with different benefit levels

1. Eligible Population

2. Evaluation sample

3. Randomize treatment (2 benefit levels)



= Ineligible

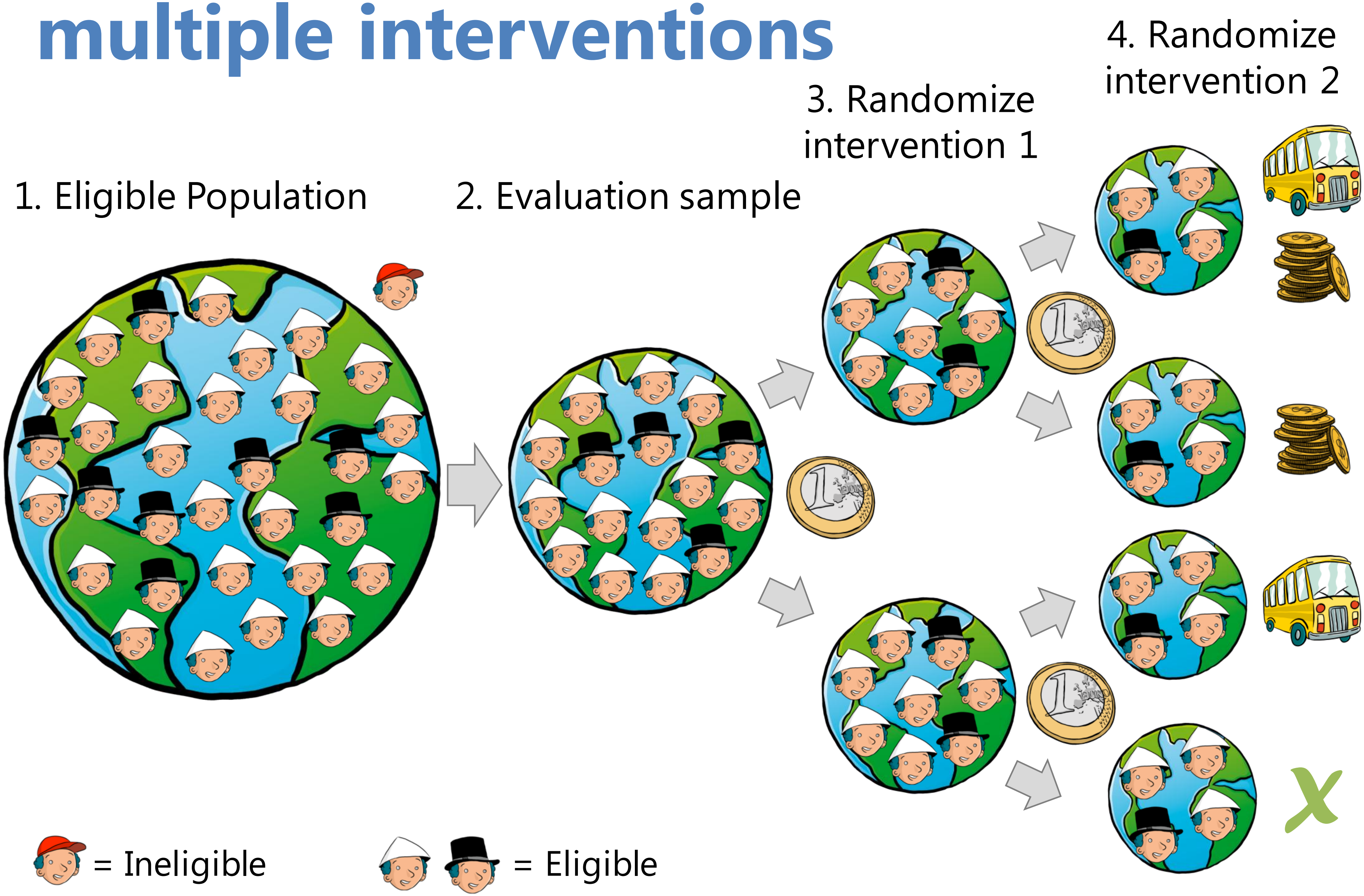
= Eligible

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

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

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

Y=Girl's school attendance
P=Tutoring program

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

+0.14 - +0.03 = 0.11

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

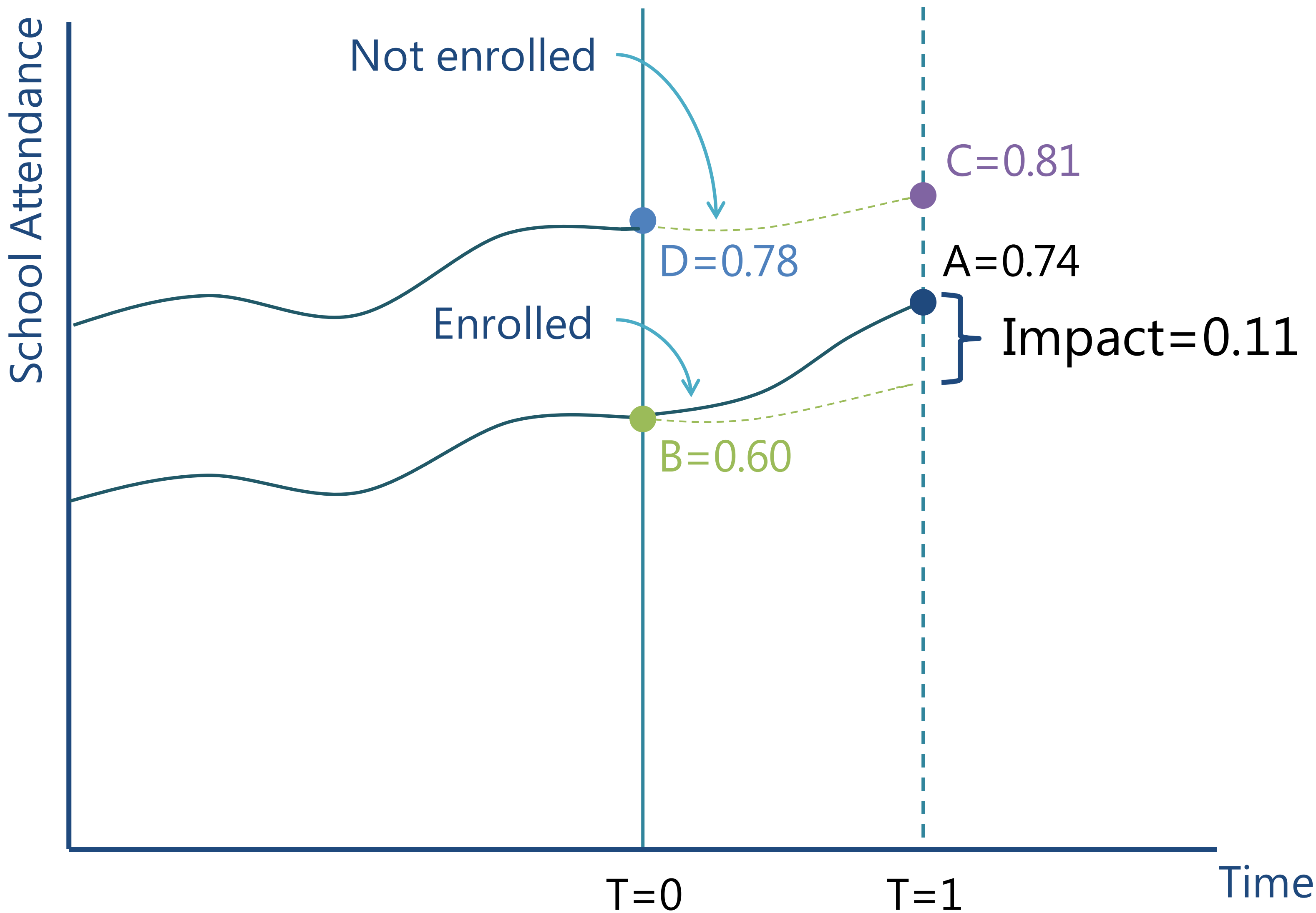
Difference-in-differences
(Diff-in-diff)

Y=Girl's school attendance
P=Tutoring program

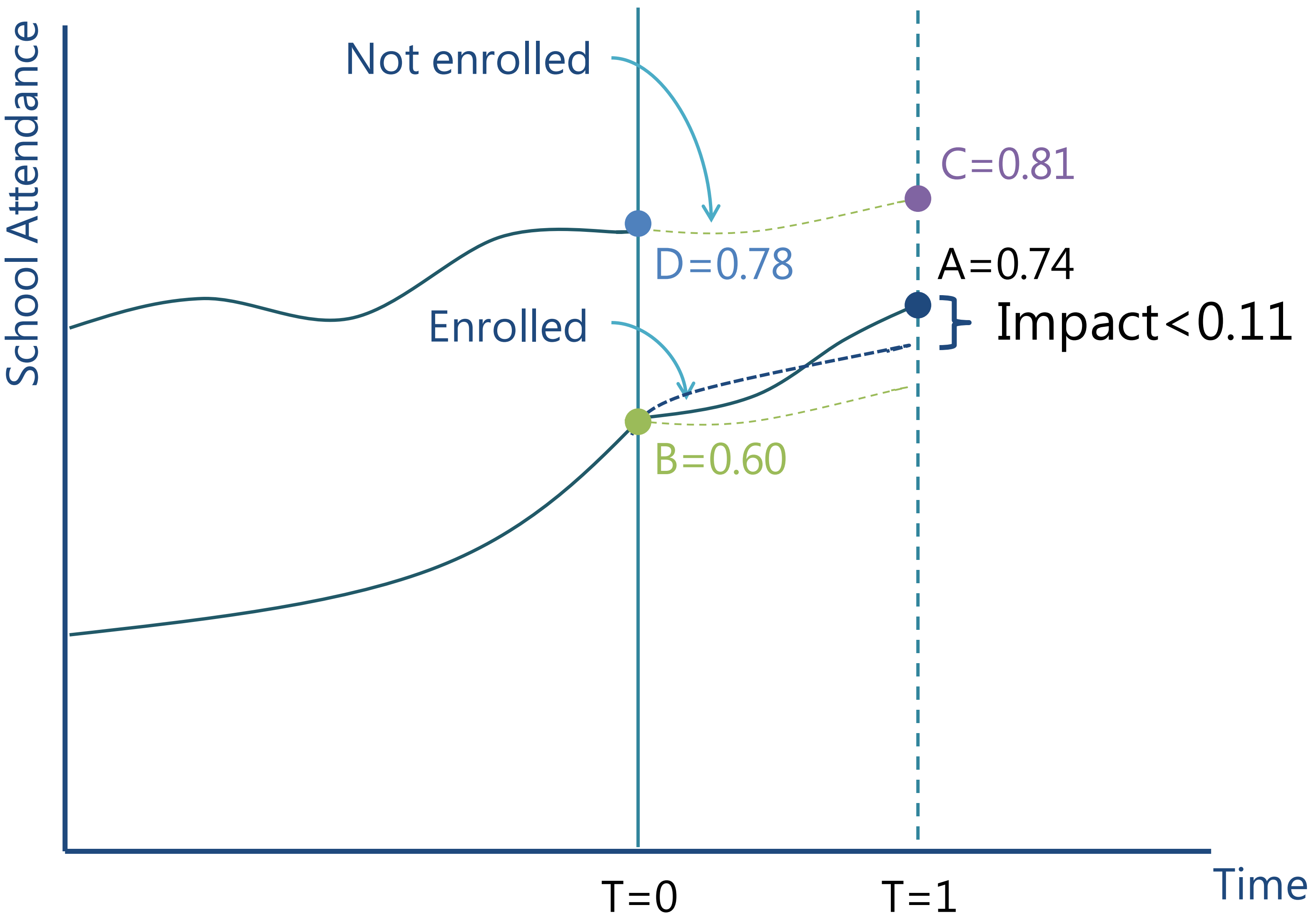
	Enrolled		Not Enrolled		Difference
After	0.74	-	0.81		**-0.07**
					-
Before	0.60	-	0.78		**-0.18**
					=
					0.11

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

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



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



Case 6: Difference in differences

	Enrolled	Not Enrolled	Difference
Baseline ***(T=0)*** Consumption (Y)	233.47	281.74	-48.27
Follow-up ***(T=1)*** Consumption (Y)	268.75	290	-21.25
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 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 controls *(Fundamental assumption)*.

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

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

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

Combinations of methods

IE Methods Toolbox

Choosing your IE method(s)

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

- Prospective/Retrospective Evaluation?
- Eligibility rules and criteria?
  - Poverty targeting?
  - Geographic targeting?
- 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?
  - Etc.

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

Choosing your method

	Targeted (Eligibility Cut-off)		**Universal (No Eligibility Cut-off)**	
	Limited Resources *(Never Able to Achieve Scale)*	Fully Resourced *(Able to Achieve Scale)*	Limited Resources *(Never Able to Achieve Scale)*	Fully Resourced *(Able to Achieve Scale)*
Phased Implementation Over Time	○ Randomized Assignment ○ RDD	○ Randomized Assignment (roll-out) ○ RDD	○ Randomized Assignment ○ Matching with DiD	○ Randomized Assignment (roll-out) ○ Matching with DiD
Immediate Implementation	○ Random Assignment ○ RDD	○ Random Promotion ○ RDD	○ Random Assignment ○ Matching with DiD	○ Random Promotion

Remember

"

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

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.

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

This material constitutes supporting material for the "Impact Evaluation in Practice book. This additional material is made freely but please acknowledge its use as follows:

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."

Thank You