Human Development *Africa*

SIEF

DIME

Measuring Impact II – Non-experimental Methods

David Evans, World Bank

Dakar, Senegal
Wednesday, October 2, 2013

Lessons from Yesterday

- What's a counterfactual?

Lessons from Yesterday

- Why is Before-After an incorrect counterfactual?

- Why is Enrolled-Not Enrolled an incorrect counterfactual?

Lessons from Yesterday

- What does randomization deliver?

Randomized Assignment

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods

Toolbox

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 school attendance for poor students

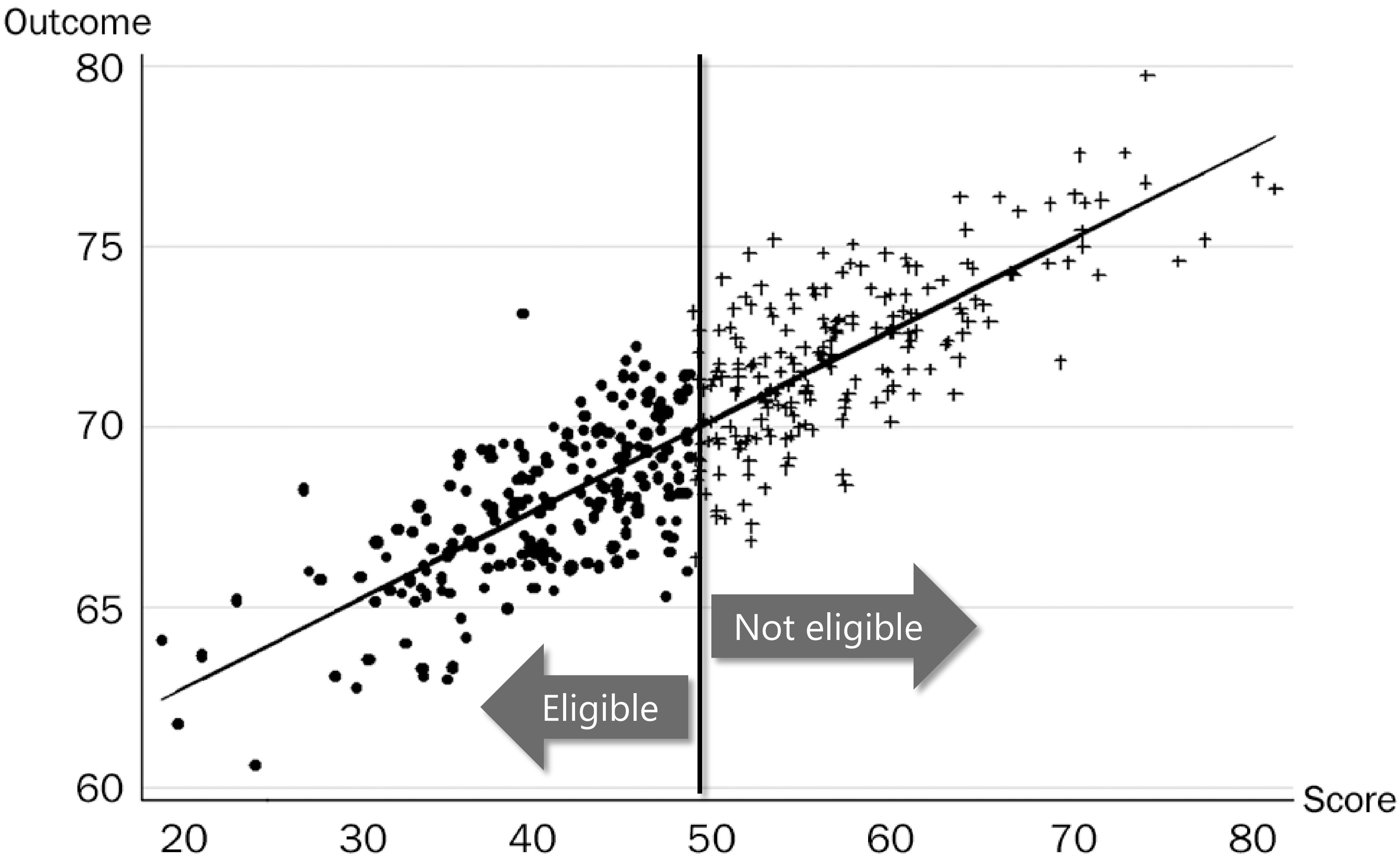
Method

- o Households with a score (Pa) of assets ≤50 are poor
- o Households with a score (Pa) of assets >50 are not poor

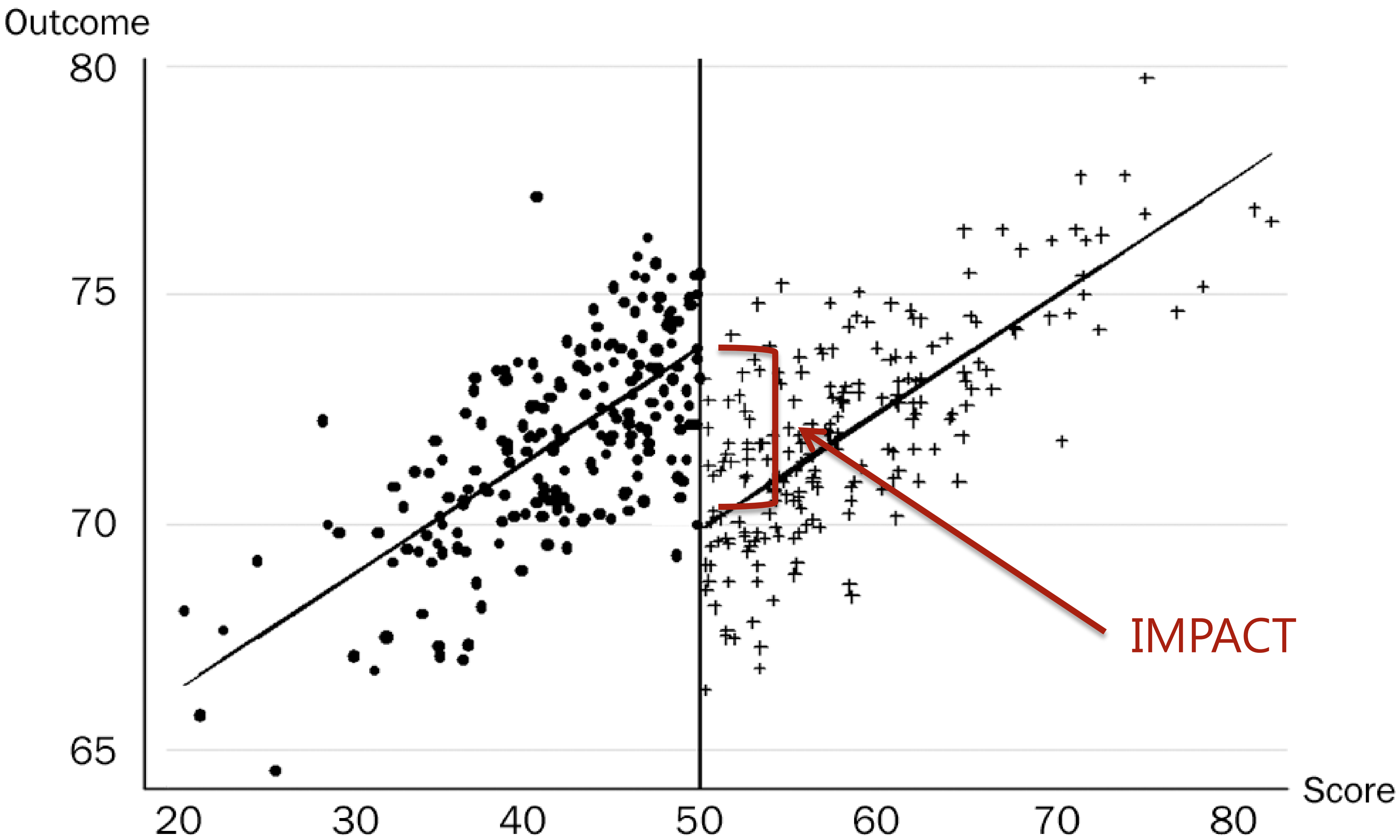
Intervention

Poor households receive scholarships to send children to school

Regression Discontinuity Design-Baseline



Regression Discontinuity Design-Post Intervention



Case 5: Discontinuity Design

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

For a discontinuity design, you need:

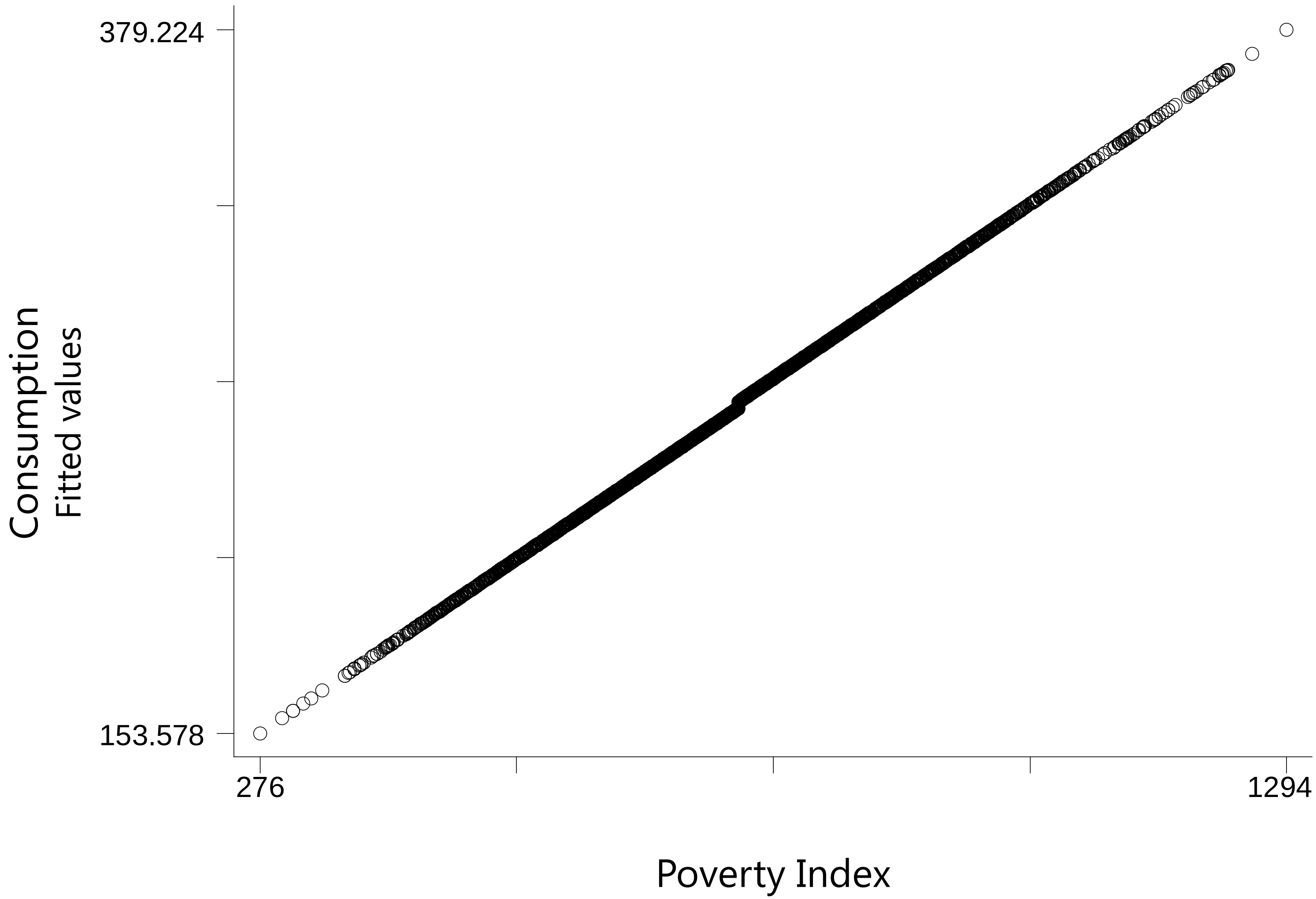
1) Continuous eligibility index
2) Clearly defines eligibility cut-off.

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

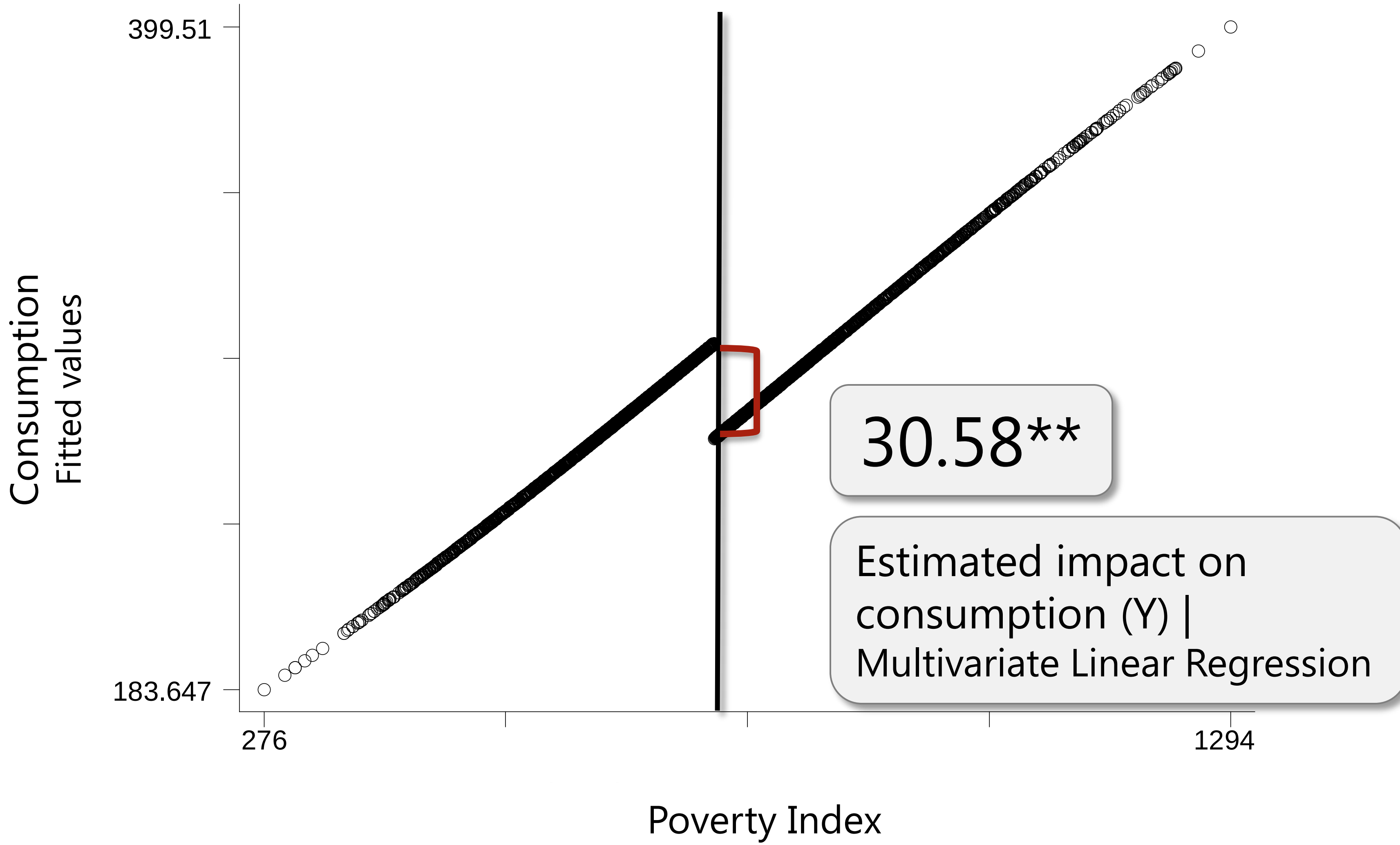
Case 5: Discontinuity Design

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 Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

2

IE Methods Toolbox

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

Y=Wages
P=Youth employment program

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

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

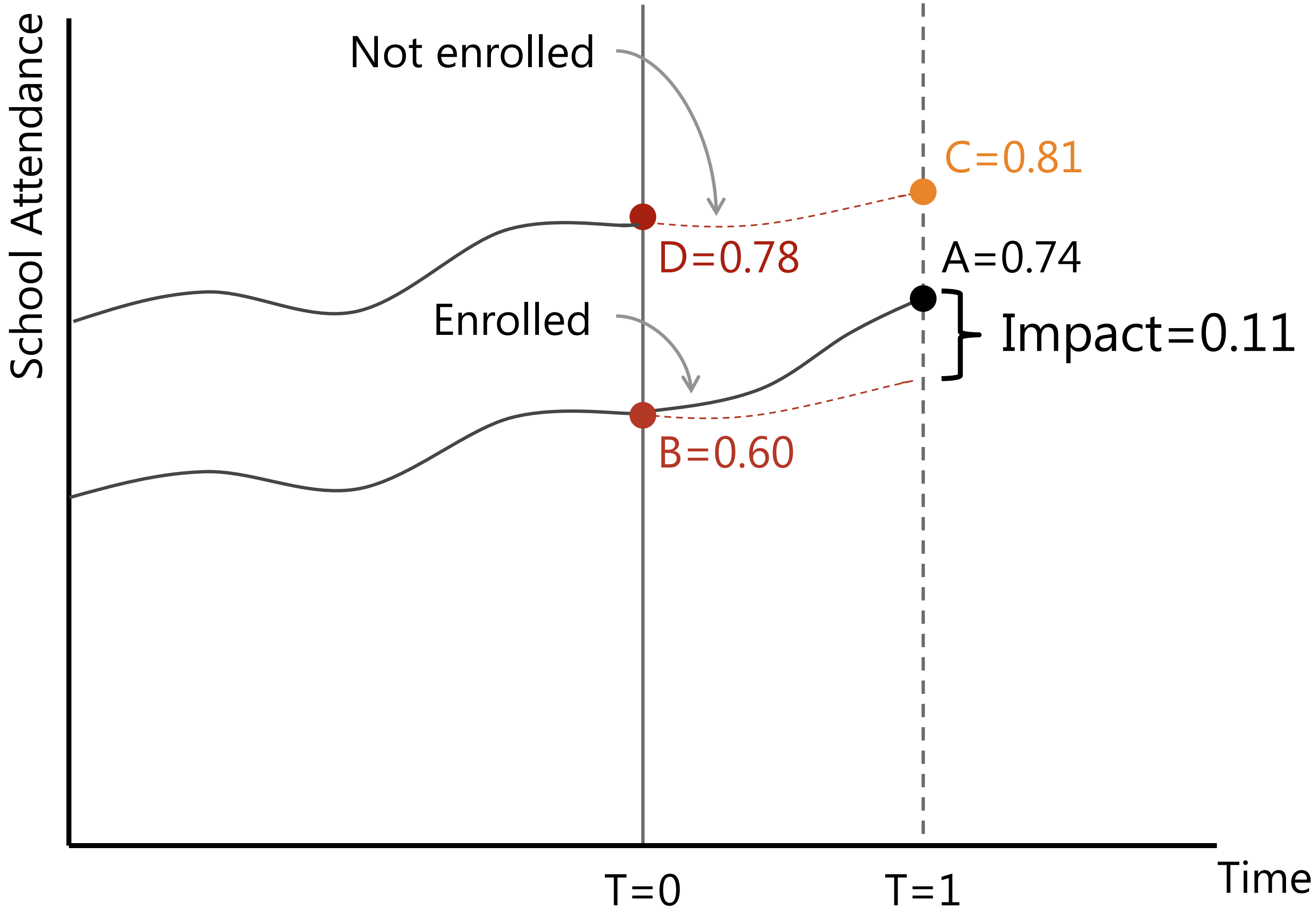
Difference-in-differences
(Diff-in-diff)

Y=School attendance
P=Girls' scholarship program

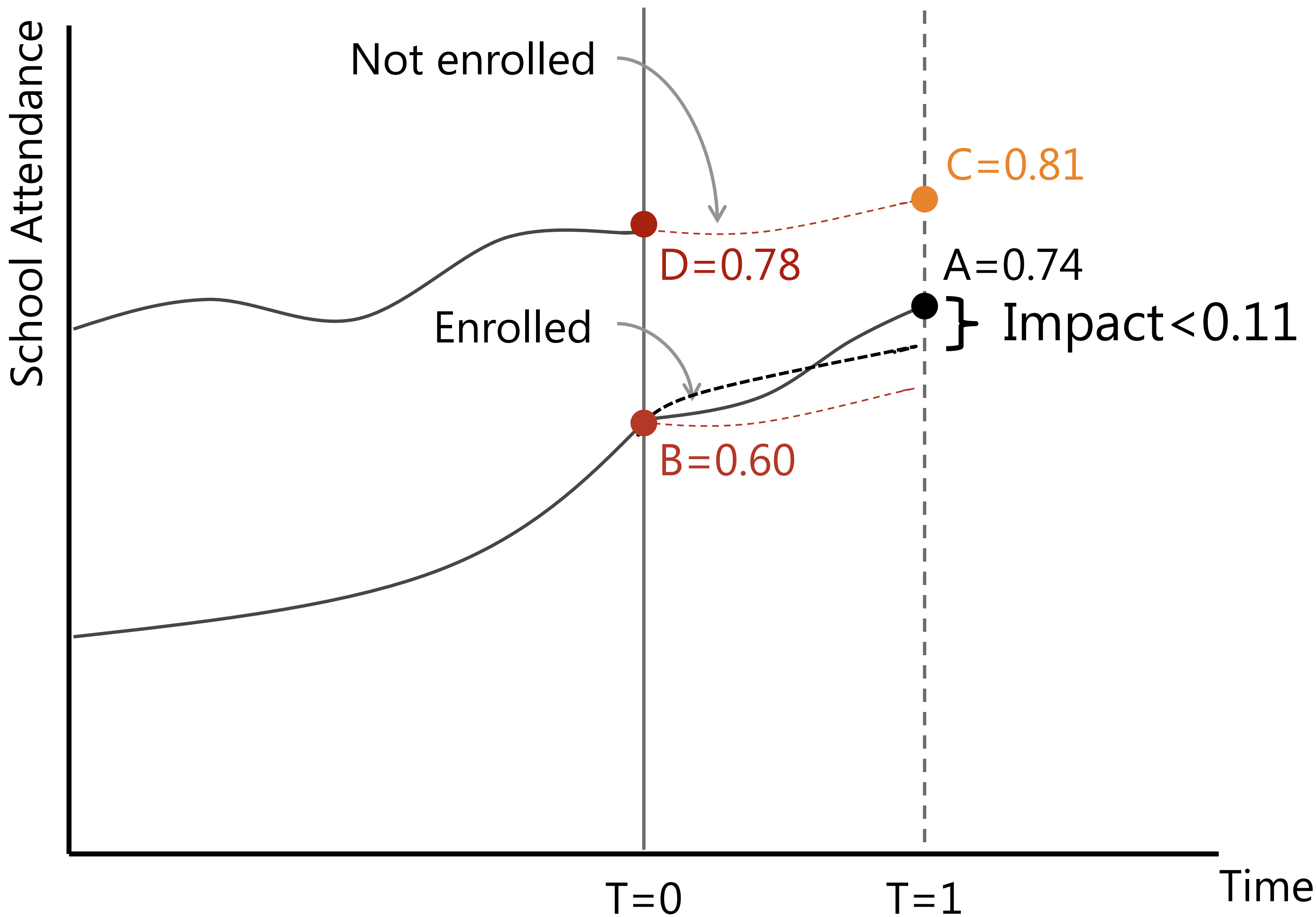
	Enrolled		Not Enrolled		Difference
After	0.74	-	0.81		**-0.07**
					-
Before	0.60	-	0.78		**-0.18**
					=
					0.11

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

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



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



Case 6: Difference in 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 (**).

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**.

Randomized Assignment

Randomized 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?

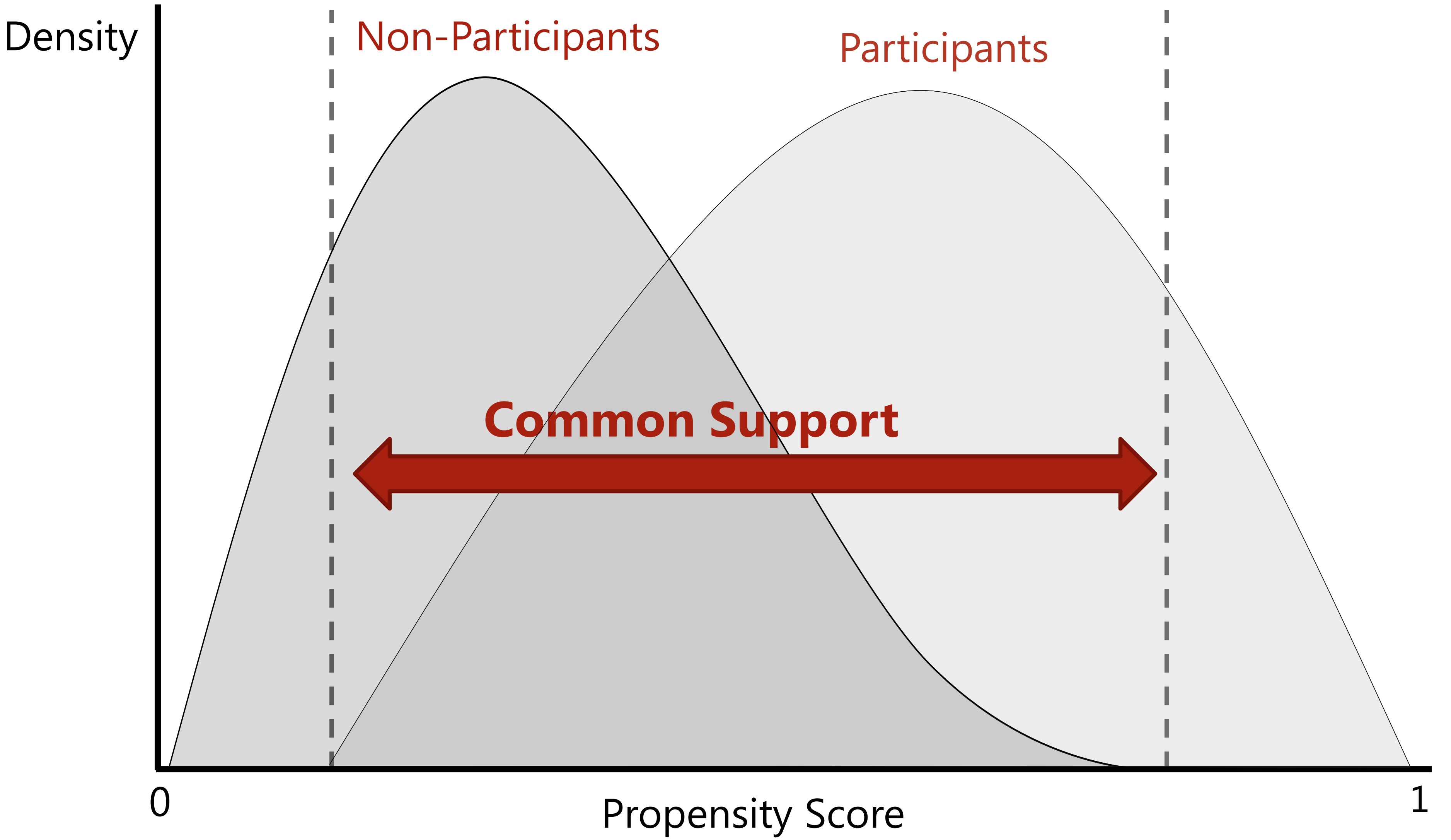
If there are *unobservable* characteristics and those unobservables influence participation: Selection bias!

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

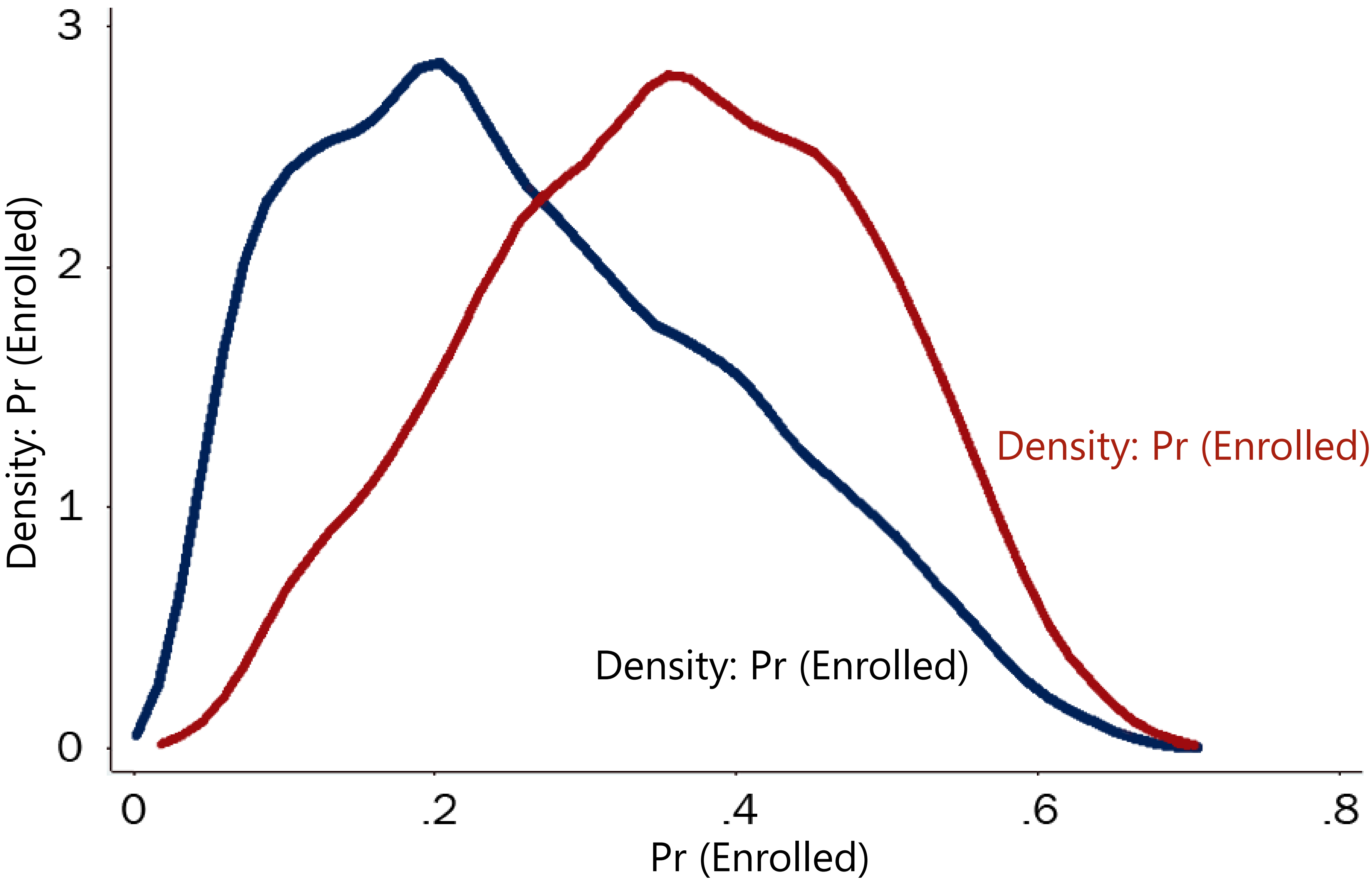


Case 7: Progresa Matching *(P-Score)*

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:

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

Ex-post matching is risky:

- o If there is no baseline, be careful!
- o 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 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 +

Randomized Assignment

Randomized Promotion

Discontinuity Design

Difference-in-Differences

Diff-in-Diff

Matching

P-Score matching

Combinations of methods

IE Methods Toolbox

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.

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?

Choosing your IE method(s)

Money → / Targeting → / Timing ↓	Excess demand		No Excess demand	
	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

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

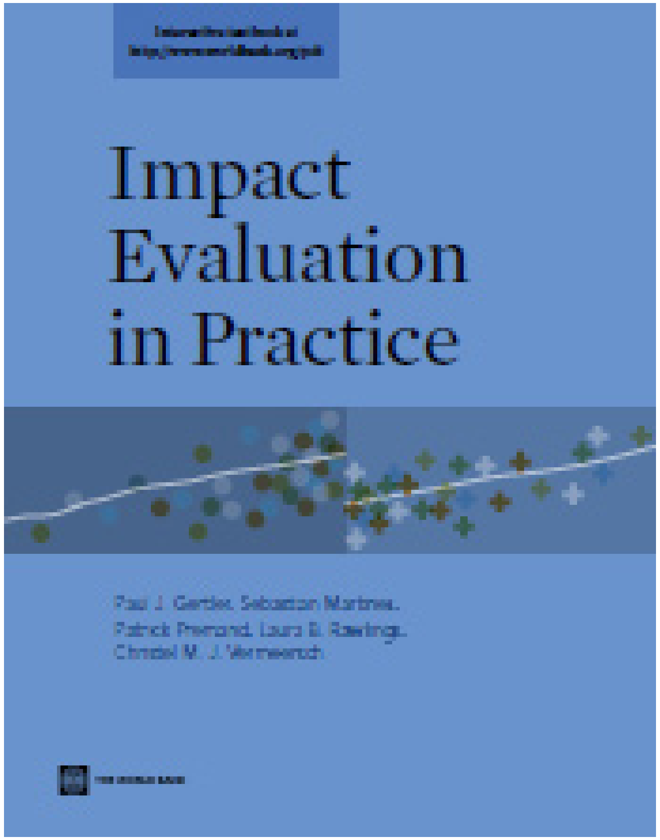
“

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



Spanish Version
& French Version
also available

www.worldbank.org/ieinpractice

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.