THE WORLD BANK

IMPACT EVALUATION

Measuring Impact:

Impact Evaluation Methods for Policy Makers

Christel Vermeersch

The World Bank

Sincere thanks to Sebastian Martinez for all of his work on these materials. The content of this presentation reflects the views of the authors, and not necessarily those of the World Bank Group. This version: August 2009.

Human Development Network

Spanish Impact Evaluation Fund

Impact Evaluation

- Logical Framework
  - How the program works "in theory"
- **<u>Measuring Impact</u>**
  - **<u>Identification Strategy</u>**
- Data
- Operational Plan
- Resources

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled &not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments & Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Difference (Diff-in-diff)
   - Matching (P-score matching)

Our Objective

- Estimate the CAUSAL effect (impact) of
  - intervention **P** (program or treatment)
  - on
  - outcome **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**?

- Answer:

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

- Can we all go home?

Problem of missing data

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

For a program beneficiary:

- we observe $(Y \mid P=1)$:
  - Consumption level (Y) with a cash transfer program (P=1)
- but we do not observe $(Y \mid P=0)$:
  - Consumption level (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……

<u>COUNTERFACTUAL</u>

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

Estimating Impact of P on Y

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

- **OBSERVE** (Y | P=1)
  Outcome with treatment

- **ESTIMATE** (Y | P=0)
  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

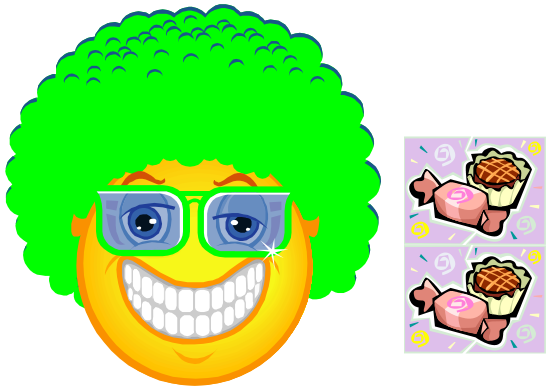
Fulanito's consumption of candies (**Y**)

The perfect “Clone”

Beneficiary

6 Candies

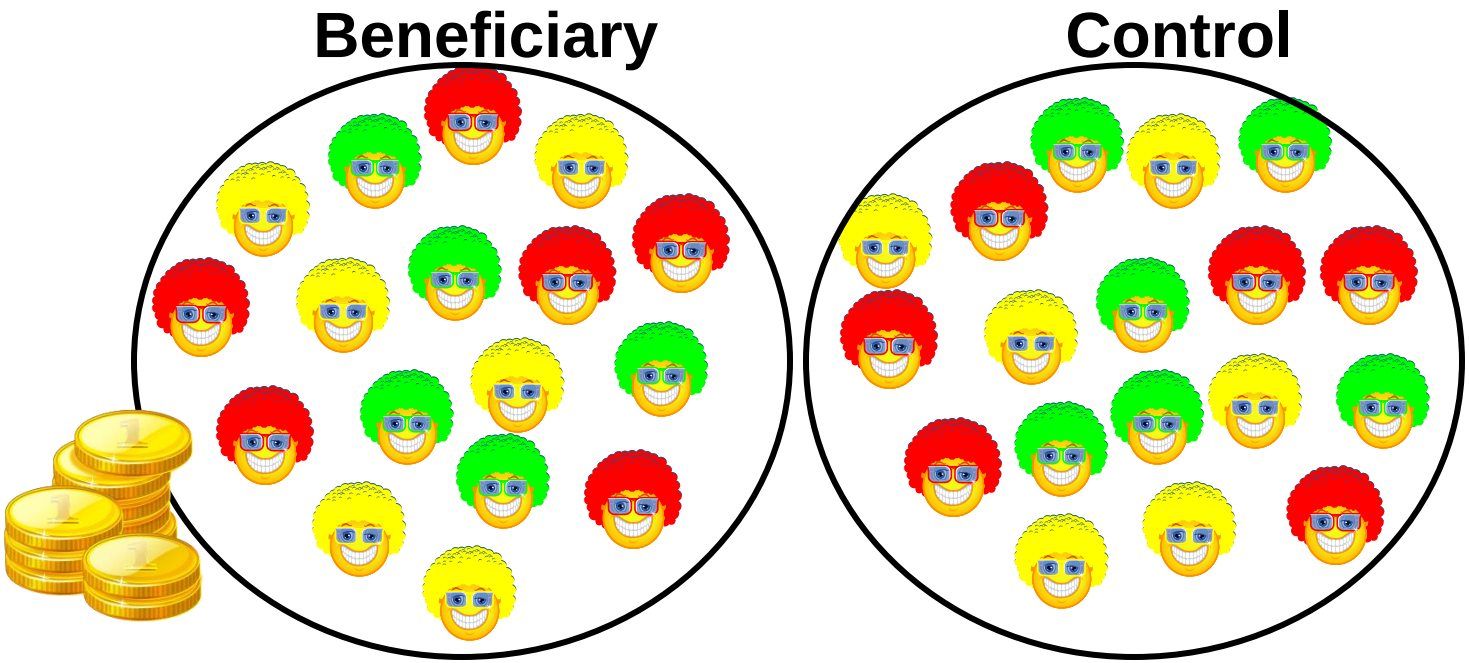
Control



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 Counterfactuals

- We want clones….
- The treated observation and the counterfactual:
  - have identical characteristics,
  - except for benefiting from the intervention
- Understand the DATA GENERATION process
  - What is it that determines who receives the treatment and who does not?
    - What is the procedure?
    - How are benefits assigned?
    - What are the eligibility rules?

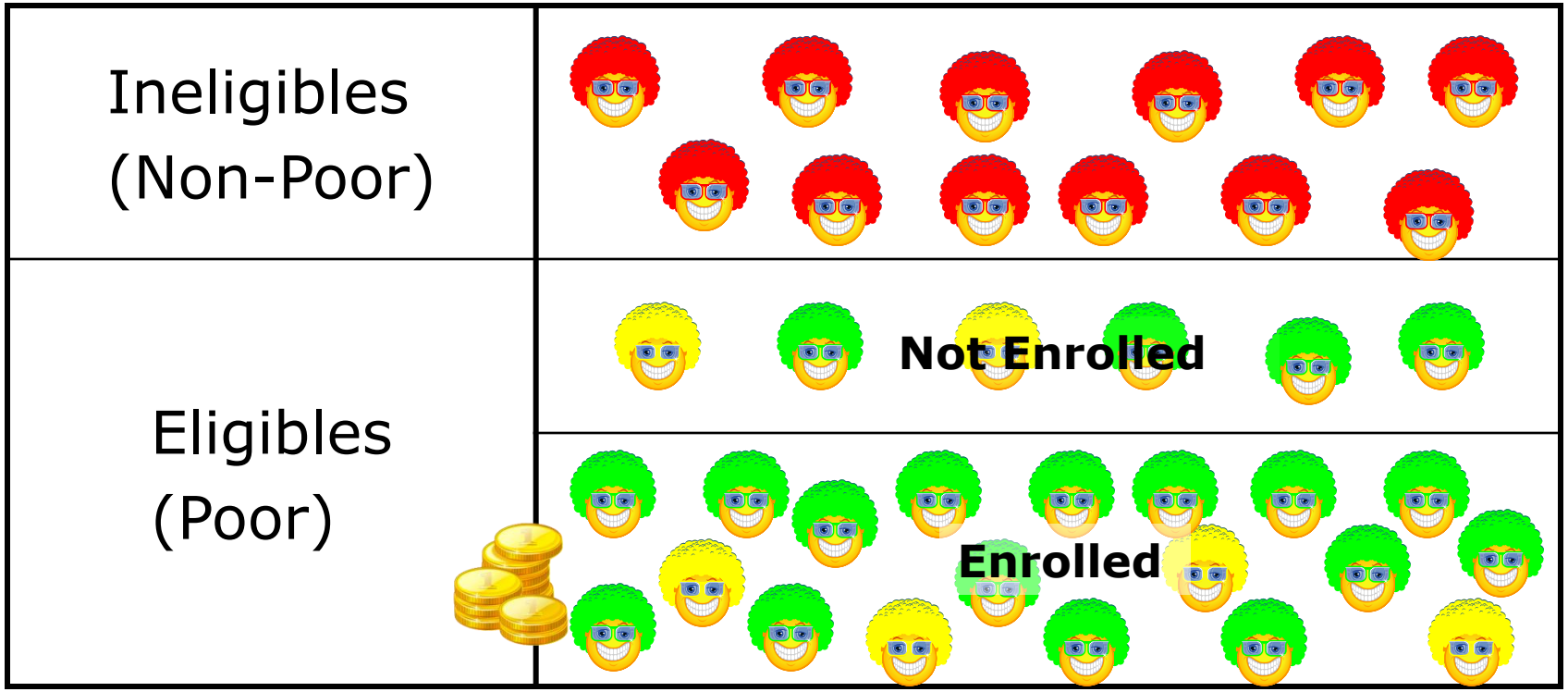
With a good counterfactual, the **only reason** for different outcomes between treatments and controls is the **intervention (P)**

Case Study: PROGRESA

- PROGRESA/OPORTUNIDADES Program
- National anti-poverty program in Mexico
  - Started 1997
  - 5 million beneficiaries by 2004
  - Eligibility – based on poverty index
- Cash transfers
  - conditional on attendance at school and health checkups
- Rigorous impact evaluation with rich data
  - 506 communities, 24K households
  - Baseline 1997, follow-up 1998
- Many outcomes of interest
- What is the effect of a cash transfer program (**P**) on **household consumption** (**Y**)?

Case Study

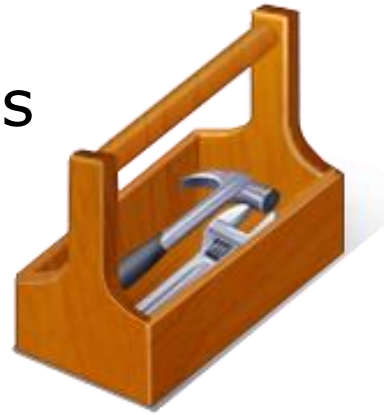
Eligibility and Enrollment



Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Difference (Diff-in-diff)
   - Matching (P-score matching)

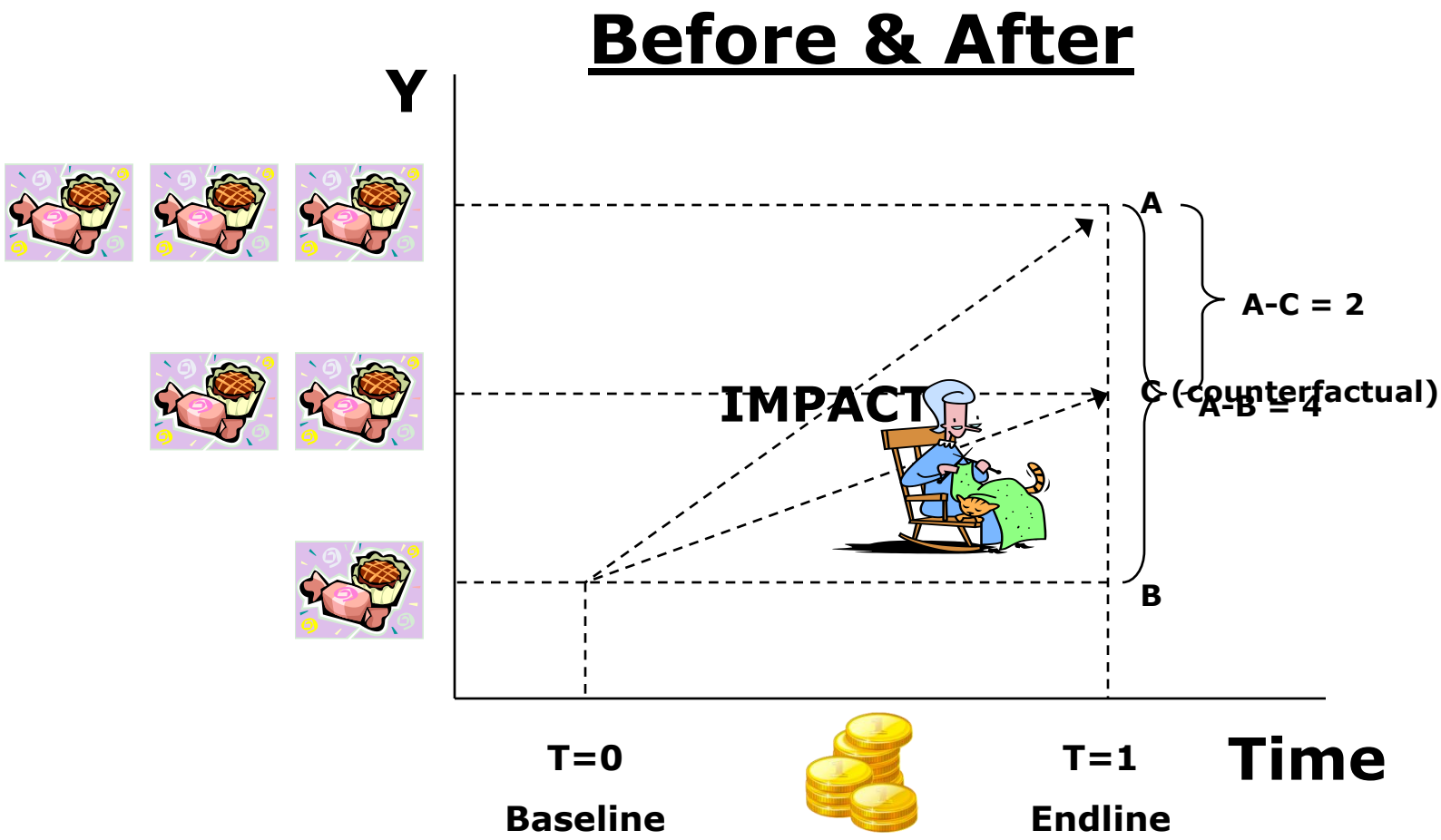


False Counterfactuals

- Two common counterfactuals to be avoided!!
  - **Before & After** (pre & post)
    - Compare: same individuals before and after they receive **P**
    - Problem: other things may have happened
  - **Enrolled & Not enrolled** (apples & oranges)
    - Compare: a group of individuals that enrolled in a program with another group that did not
    - Problem: We don't know why they are not enrolled

- Both counterfactuals may lead to biased estimates of the impact

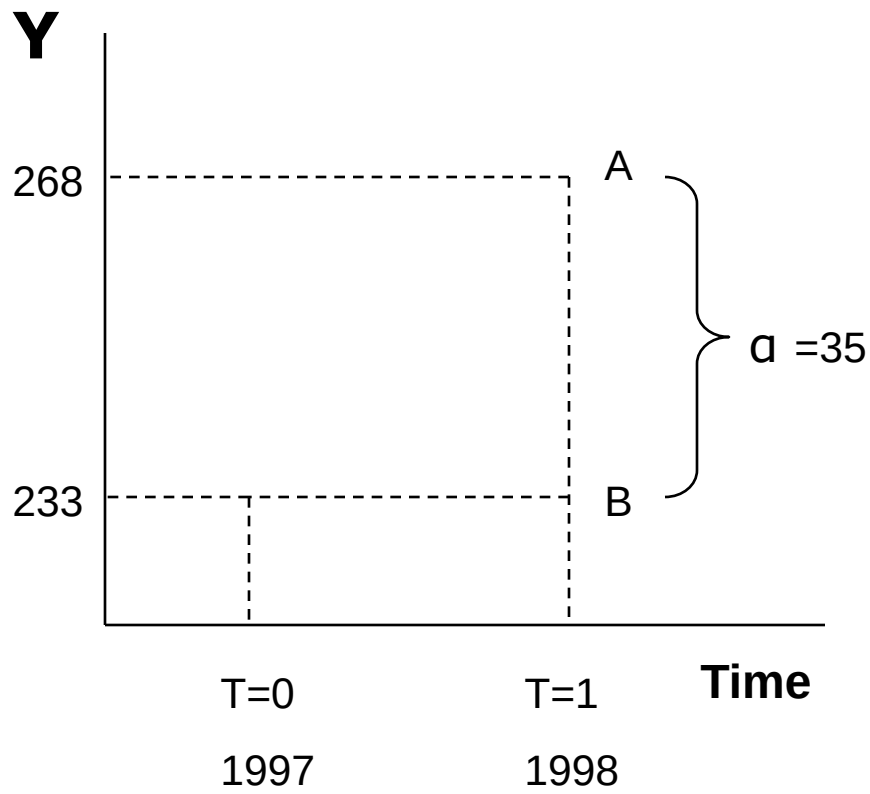
Counterfeit Counterfactual #1



Progresa Case 1: Before & After

What is the effect of **a cash transfer program (P)** on **household consumption (Y)**?

- Can only measure beneficiaries (P=1)
- 2 observations in time
  - Consumption at T=0
  - Consumption at T=1
- Outcome measure under treatment
  $=A= (Y_{t=1}| P=1)$
- Estimate of counterfactual
  $=B=(Y_{t=0}| P=1)$
- "Impact" = α =A-B = 35



Progresa Case 1: Before & After

	With treatment (After)	Counterfactual / comparison (Before)	Difference =After-Before
Mean annual consumption per capita	268,75	233,48	35.27

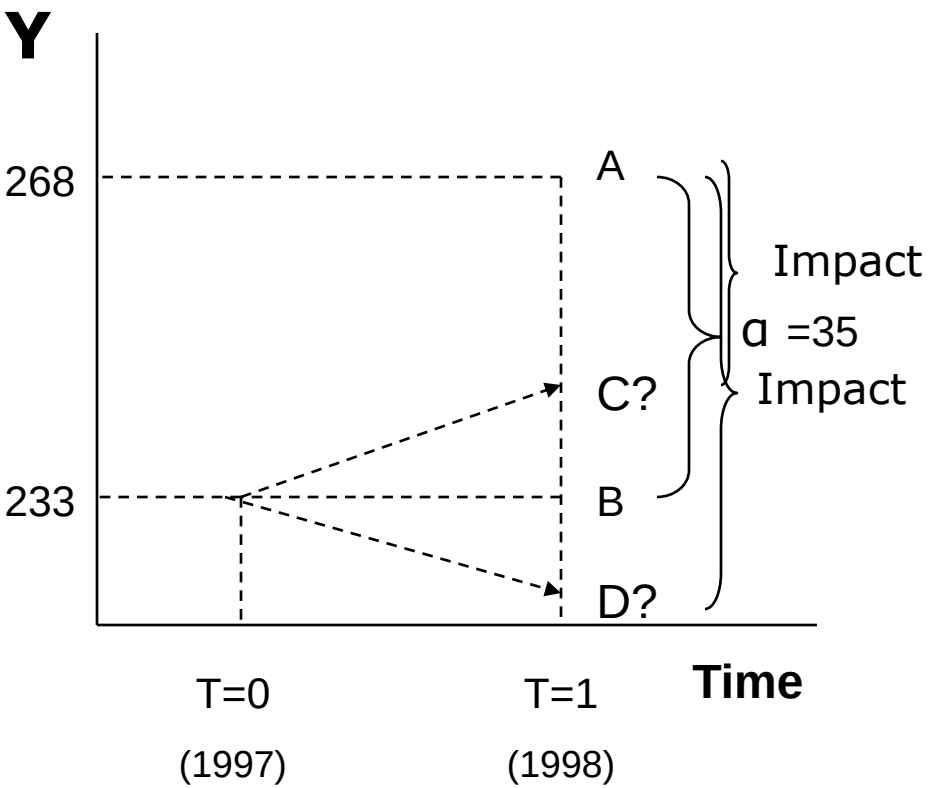
	Linear regression	Multivariate linear regression
Estimated impact on annual consumption per capita	35,27** (2,16)	34,28** (2,11)

Note: Standard errors in parenthesis. If the difference is statistically significant at the 1% significance level, we label the estimated impact with 2 stars.

Progresa Case 1: Before & After

What's the Problem?

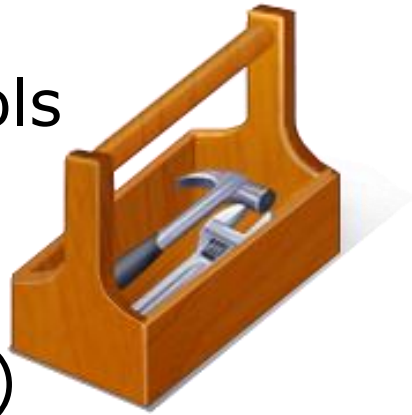
- Outcome measure under treatment =A (after)
- Estimate of counterfactual =B (before)
- "Impact" = α =A-B = 35
- We do not control for time-varying factors
  - Boom: Real Impact = A-C
    - A-B is an overestimate
  - Recession: Real Impact = A-D
    - A-B is an underestimate



Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Difference (Diff-in-diff)
   - Matching (P-score matching)



False Counterfactual #2

Enrolled & not enrolled

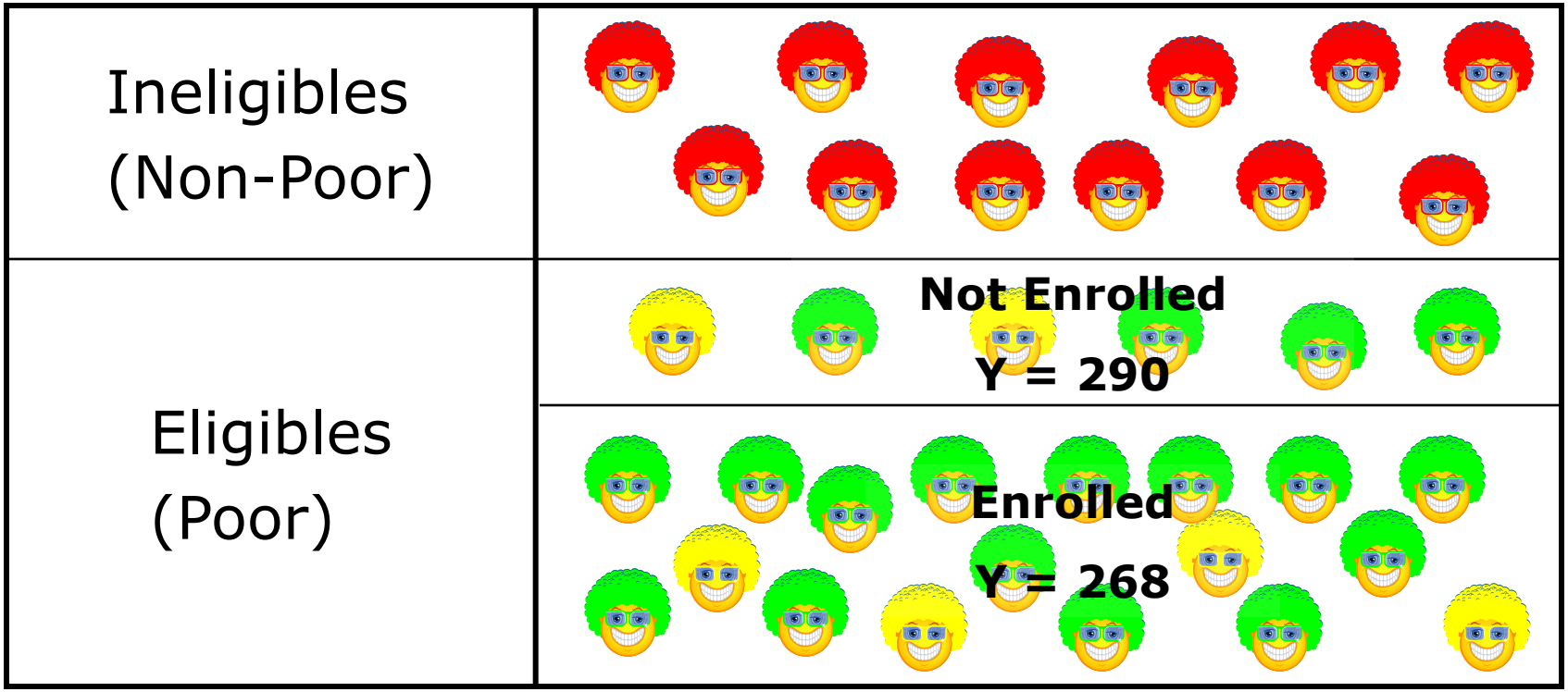
- If we have post-treatment data on
  - Enrolled: treatment group

  - Not-enrolled: "control" group (counterfactual)
    - Those **ineligible** to participate
    - Those that **choose NOT** to participate

- Selection Bias
  - Reason for not enrolling may be correlated with outcome (Y)
    - Control for observables
    - But not unobservables!!
  - Estimated impact is confounded with other things

Progresa Case 2: Enrolled & Not enrolled

Measure outcomes in post-treatment (1998)



In what ways might enrolled & not enrolled be different, other than their enrollment in the program?

Progresa Case 2: Enrolled & Not enrolled

Mean annual consumption per capita

	Comparison (Not enrolled)	Treatment (Enrolled)	Difference
Mean annual consumption per capita	290,16	268,75	-22,7

Estimated impact (treatment versus comparison)

	Linear regression	Multivariate linear regression
Estimated impact on annual consumption per capita	-22,7** (3,78)	-4,15 (4,05)

Note: Standard errors in parenthesis. If the difference is statistically significant at the 1% significance level, we label the estimated impact with 2 stars.

Progresa: Case 1 or Case 2 ?

	Case 1: before & after		Case 2 : Enrolled & not enrolled	
	Linear regression	Multivariate linear regression	Linear regression	Multivariate linear regression
Estimated impact on annual consumption per capita	35,27** (2,16)	34,28** (2,11)	-22,7** (3,78)	-4,15 (4,05)

- Which of these do we believe?
- Problem with Before-After:
  - *Can not control for other time-varying factors*
- Problem with Enrolled-Not Enrolled:
  - *Don't know if other factors, beyond the intervention, are affecting the outcome*
  - *Don't know why some poor enrolled and others not*

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)

Choosing your IE method(s)…..

- Key information you will need for your program:
  - Prospective/retrospective ?
  - Eligibility rules and criteria?
  - 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)…..

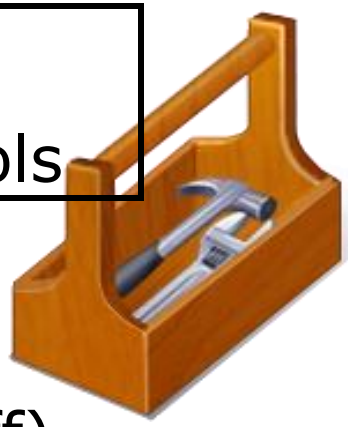
Choose the **"best" possible design given** the operational context

- Best design = best comparison group you can find + least operational risk
- Have we controlled for "everything"?
  - Internal validity
  - "good counterfactual clones"
- Is the result valid for "everyone"?
  - External validity
  - Local versus global treatment effect
  - Evaluation results apply to population we're interested in

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

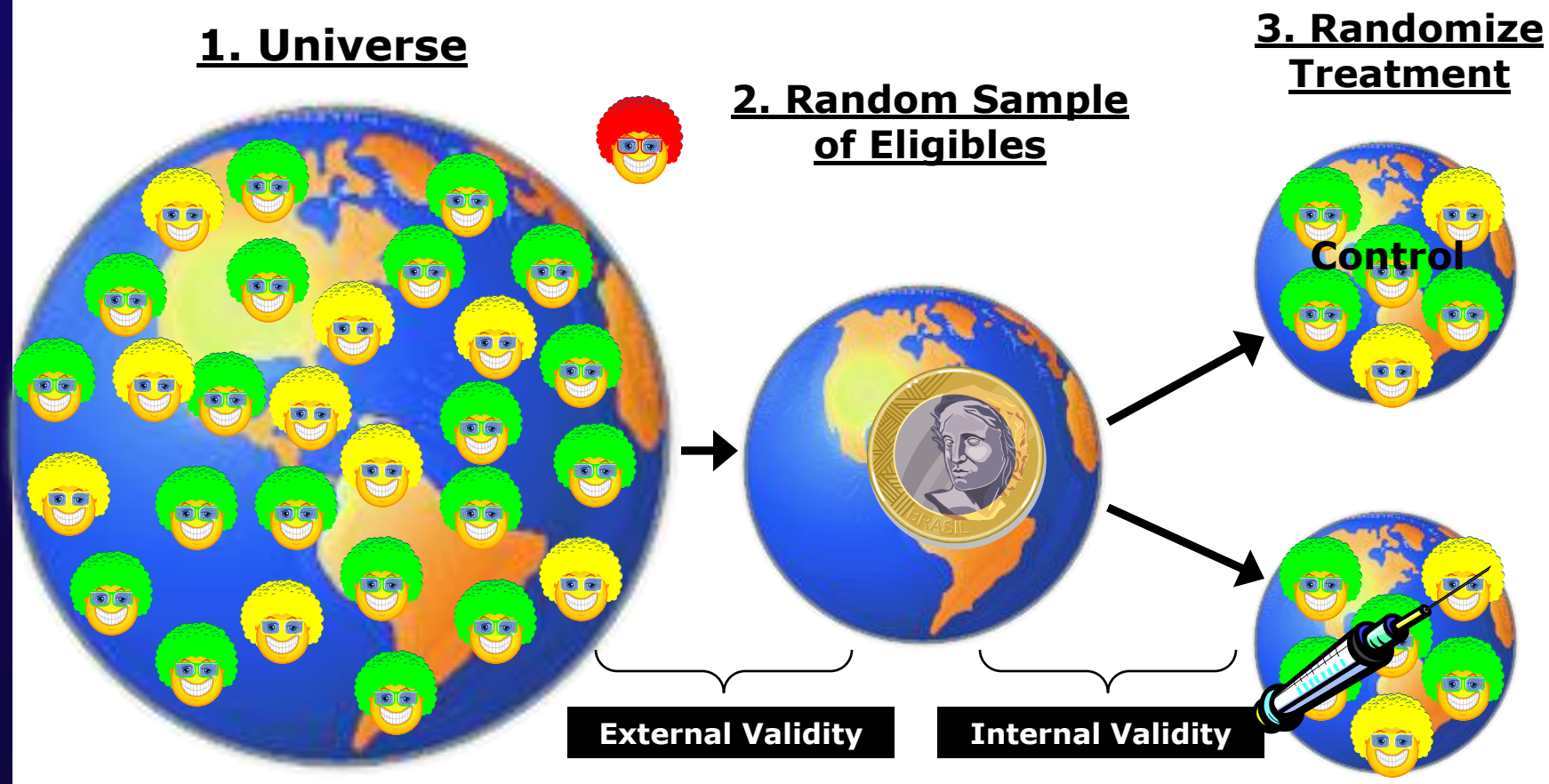
2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)



Randomized Treatments and Controls

- When universe of eligibles > # 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 treatment later (controls)

Randomized treatments and controls



Ineligible = 

Eligible = 

Unit of Randomization

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

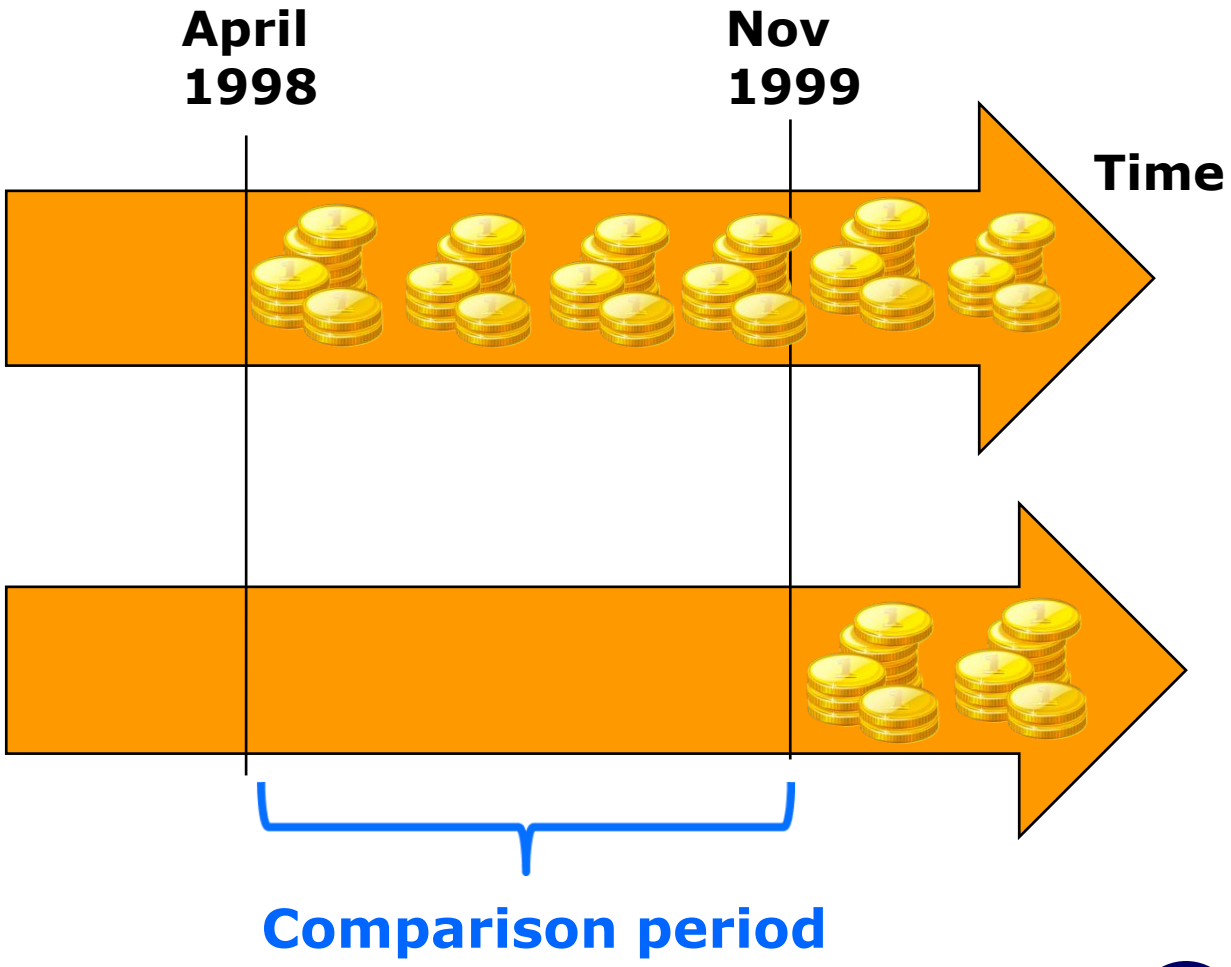
- Keep in mind:
  - Need "sufficiently large" number of units to detect minimum desired impact ➔ power
  - Spillovers/contamination
  - Operational and survey costs

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

Progresa Case 3: Randomized Phase-in

320
Treatment Communities (14,446 HH)

186
Control Communities (9,630HH)



Progresa Case 3: Randomization

- Unit of randomization: Community
- 506 communities in the sample
- Randomized phase-in:
  - 320 treatment communities (14,446 households)
    - Started receiving transfers in April 1998
  - 186 control communities (9,630 households)
    - Started receiving transfers in November 1999

Progresa Case 3: Randomization

How do we know we have good clones?

Let's compare how the "treatment" and "control" units look like

Let's compare their characteristics at baseline

Progresa Case 3: Balance at Baseline

	Progresa Case 3: randomization		
	Treatment	Control	T-stat
Annual per capita consumption	233.47 (1.02)	233.4 (1.3)	-0.04
Age of head of household	41.94 (0.2)	42.35 (0.27)	1.2
Years of Education of head of household	2.95 (0.04)	2.81 (0.05)	-2.16
Age of spouse	37.02 (0.7)	36.96 (0.22)	-0.38
Years of Education of spouse	2.76 (0.03)	2.76 (0.04)	-0.006

**= significant at 1%

Progresa Case 3: Balance at Baseline

	Progresa Case 3: randomization		
	Treatment	Control	T-stat
Female head of household	0.073 (0.03)	0.078 (0.005)	0.66
Household size	5.76 (0.02)	5.7 (0.038)	-1.21
Household has bathroom	0.57 (0.007)	0.56 (0.009)	-1.04
Surface of owned land in ha	1.63 (0.03)	1.72 (0.05)	1.35
Distance to urban center	109.28 (0.6)	106.59 (0.81)	-1.02

**= significant at 1%

Progresa Case 3: Randomization

Mean per capita consumption

	Treatment	Control	Difference
Mean per capita consumption at baseline	233,47	233,40	0.07
Mean per capita consumption at follow up	268,75	239,50	29.25

Estimated impact (treatment versus control)

	Linear regression	Multivariate linear regression
Estimated impact on mean per capita consumption	29,25** (3,03)	29,79** (3,00)

Progresa Case Study: Cases 1-3

	Case 1: Before & After	Case 2: Enrolled & Not Enrolled	Case 3: Randomization
	Multivariate Linear Regression	Multivariate Linear Regression	Multivariate Linear Regression
Estimated impact on mean per capita consumption	34,28** (2,11)	-4,15 (4,05)	29,79** (3,00)

Standard errors in parenthesis.
**= significant at 1%

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)

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,
  - Most will accept
  - A few will not accept

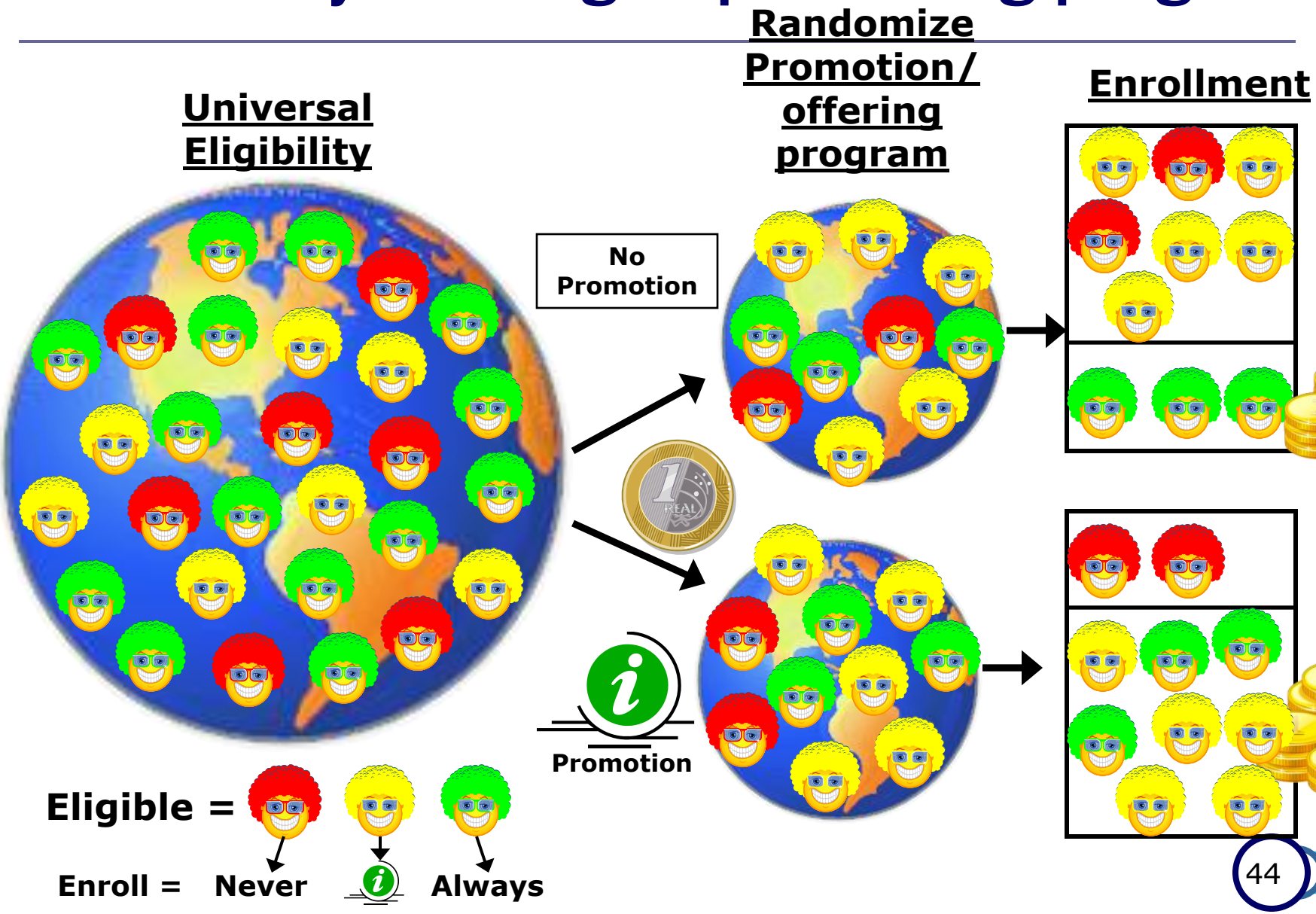
- If you cannot exclude anyone, and cannot force anyone
  - Making the program **available to everyone**
  - But provide additional **promotion**, **encouragement** or **incentives** to a random sub-sample:
    - Additional Information
    - Encouragement
    - Incentives (small gift or prize)
    - Transport (bus fare)

Randomly offering or promoting program

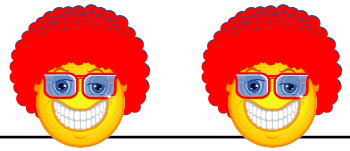
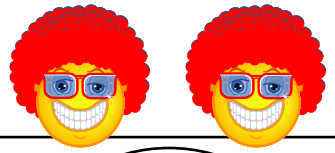
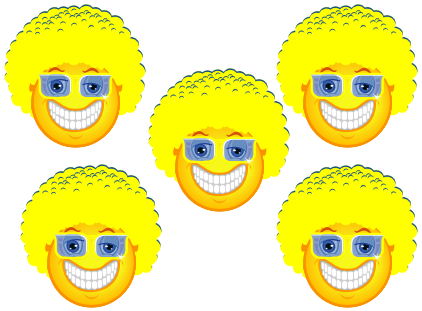
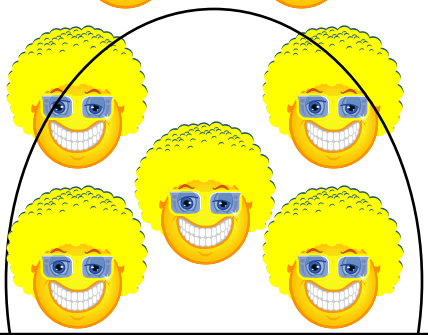
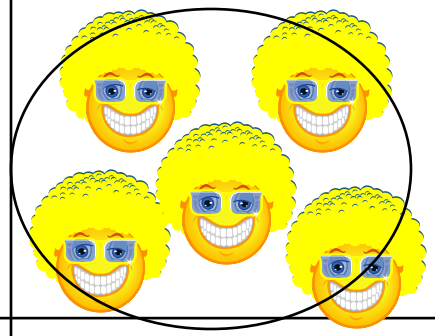
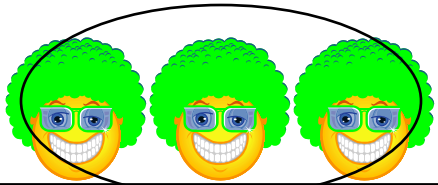
Necessary conditions:

1. Offered/promoted and not-offered/non-promoted groups are comparable:
   - Whether or not you offer or promote is not correlated with population characteristics
   - OK if you randomize
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



Randomly offering or promoting program

	NO OFFERING/ NO PROMOTION GROUP	OFFERED/ PROMOTED GROUP	IMPACT
	% Enrolled = 30% Average Y for entire group= 80	% Enrolled = 80% Average Y for entire group= 100	Δ Enrolled= 50% Δ Y=20 Impact = 20/50% =40
Never Enroll			X
Only Enroll if Encouraged			
Always Enroll			X

Randomized promotion: Examples

- Maternal Child Health Insurance in Argentina
  - Intensive information campaigns
- Employment Program in Argentina
  - Transport voucher
- Community Based School Management in Nepal
  - NGO helps with enrollment paperwork

Randomized Promotion

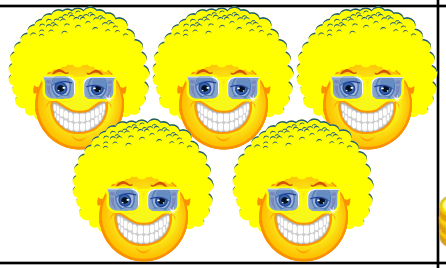
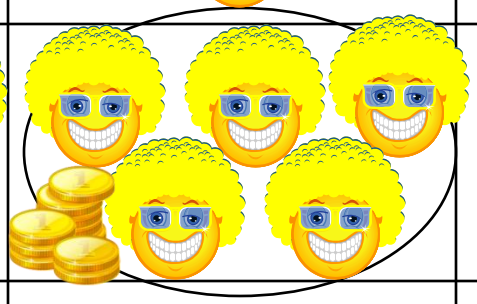
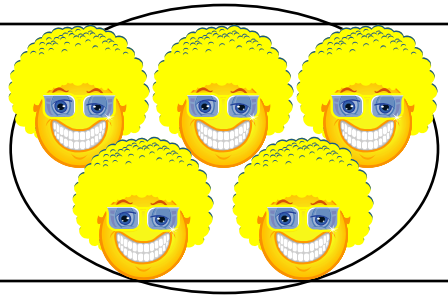
- Your promotion strategy…
  - Needs to be tested in advance!
  - Testing will help understand how to increase enrollment
- Don’t have to “exclude” anyone, but…..
  - Strategy depends on success and validity of promotion
  - We estimate a **local** average treatment effect
  - Impact estimate valid only for the “**yellow**” type of beneficiaries

Randomized Promotion

- Randomized Promotion is an "Instrumental Variable" (IV)
  - A variable correlated with treatment but nothing else (i.e. random 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

Progresa Case 4: Randomized promotion

	Not offered treatment "NOT PROMOTED"	**Offered treatment "PROMOTED"**	**IMPACT**
	Enrolled = 0% Average Y = 239	Enrolled = 92% Average Y = 268	Δ Enrolled= 92% Δ Y=29 Impact = 29/0.92 = 31
Never Enroll			X
Enroll if Encouraged			
Always enroll	X	X	X

Progresa Case 4: Random offering/ Treatment on treated

- Estimate TOT effect of Oportunidades on consumption

	Case 4: Random offering / Treatment on the treated	Case 4: Random offering / Treatment on the treated
	Instrumental variable	Instrumental variable with additional controls
Estimated impact on mean per capita consumption	29,88** (3,09)	30,44** (3,07)

Standard errors in parenthesis.
**= significant at 1%

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)

Discontinuity Design

Many social programs select beneficiaries using an "index" or "score":

- Anti-poverty programs:

  ➔ targeted to households below a given poverty index/score

- Pension programs:

  ➔ targeted to population above a certain age

- Scholarships:

  ➔ targeted to students with high scores on standardized test

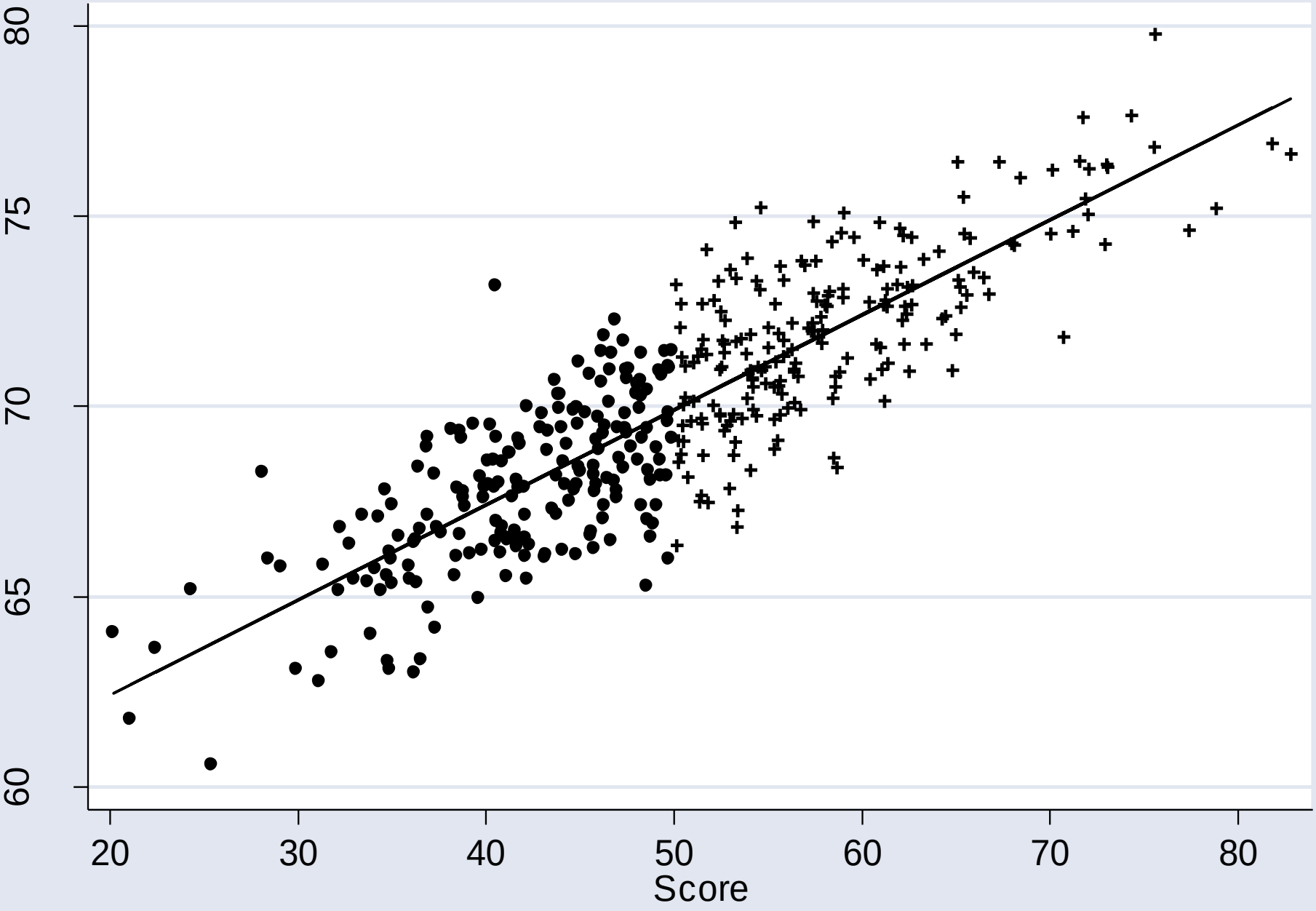
Example: effect of cash transfer on consumption

- **Goal:** Target transfer to poorest households

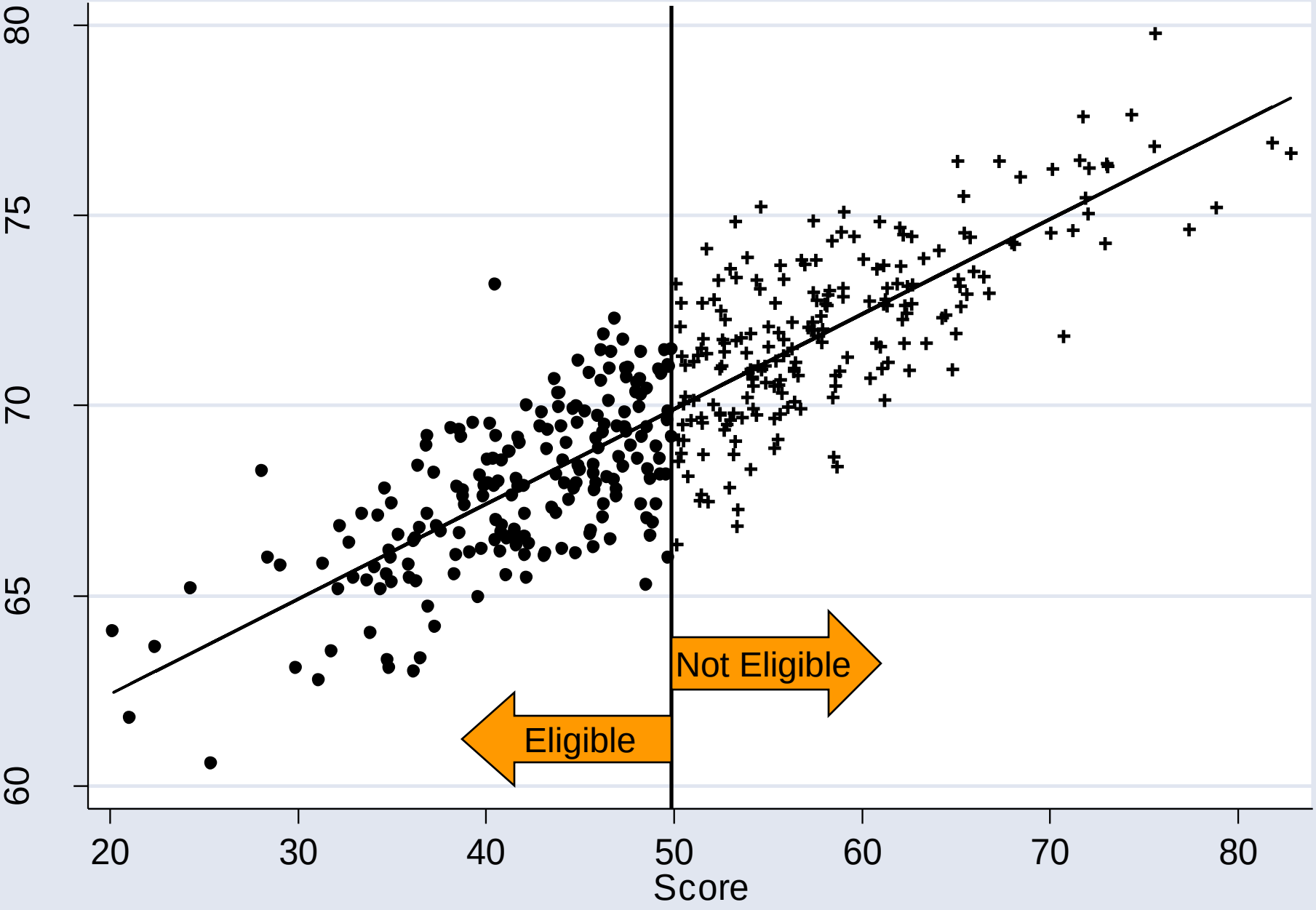
- **Method:**
  - Construct poverty index from 1 to 100 with pre-intervention characteristics
  - Households with a score <=50 are poor
  - Households with a score >50 are non-poor

- **Implementation**:
  - Cash transfer to poor households

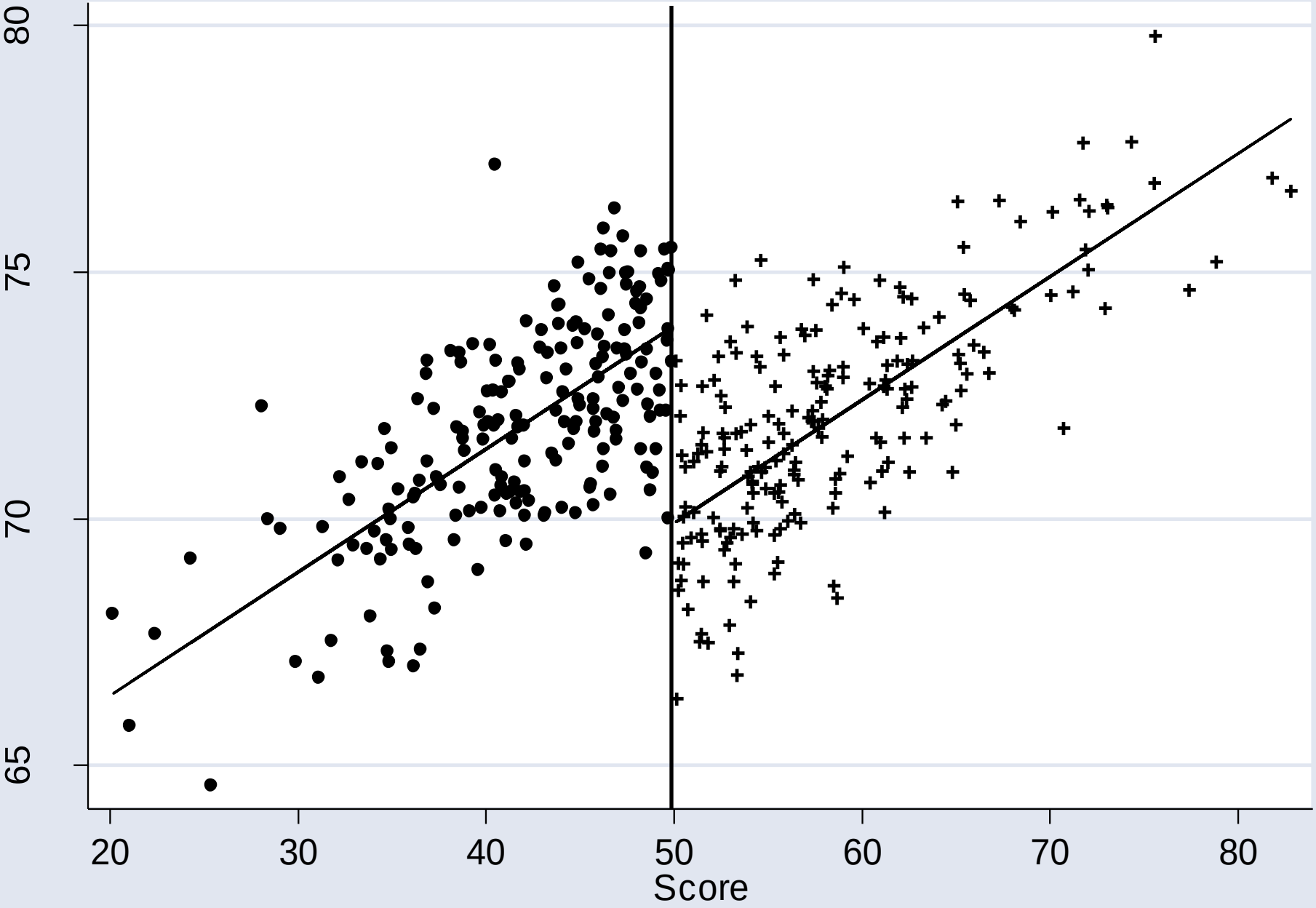
Regression Discontinuity Design - Baseline



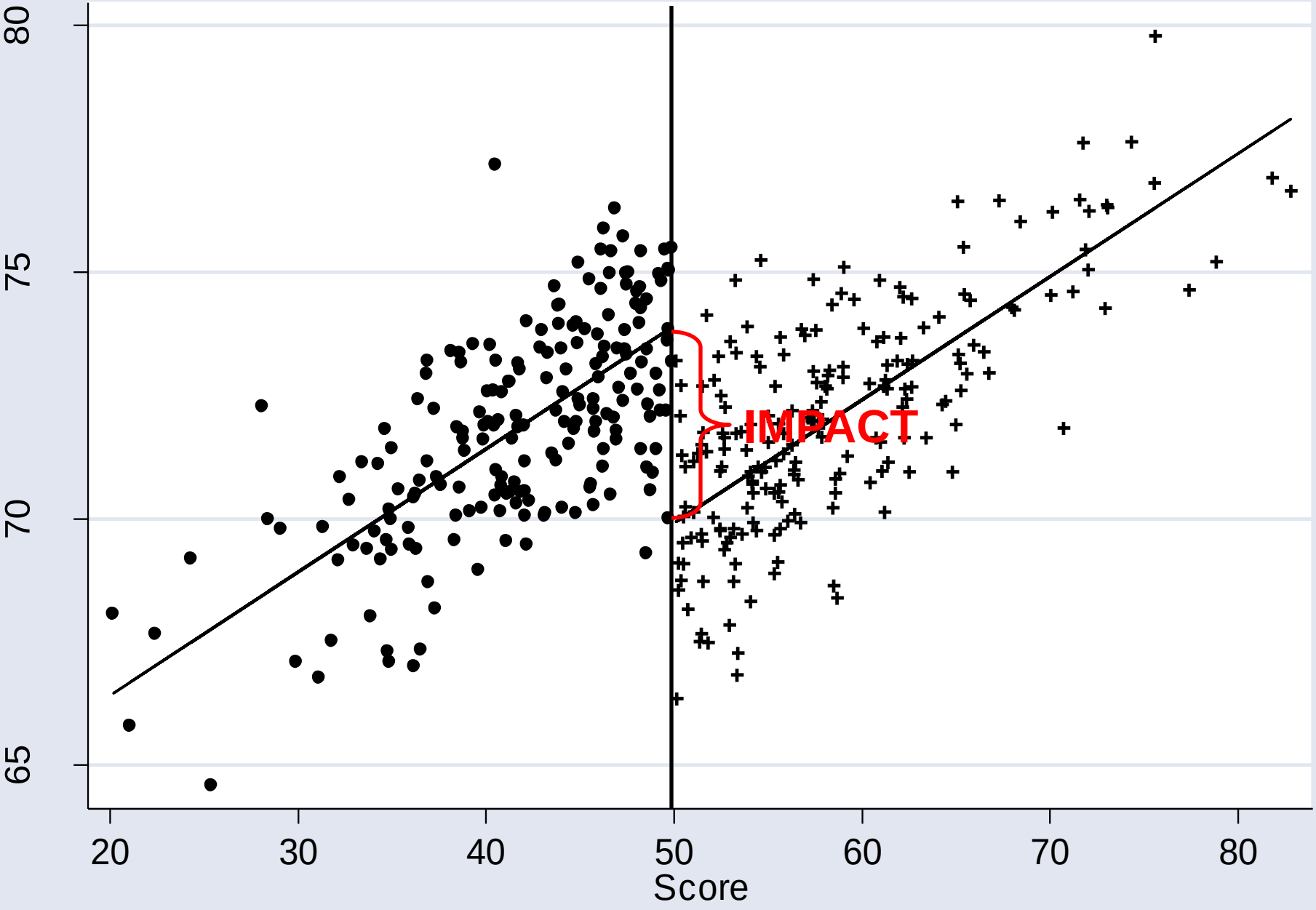
Regression Discontinuity Design - Baseline



Regression Discontinuity Design - Post Intervention



Regression Discontinuity Design - Post Intervention



Discontinuity Design

- We have: poverty score from 1 to 100
  - Households with a score <=50 are poor
  - Households with a score >50 are non-poor

- Intuitive explanation of the method:
  - Units just above the cut-off point (50) are very similar to units just below it.
  - Compare outcome **Y** for units just above and below the cut-off point (50).

For a discontinuity design, you need:

-Continuous eligibility index

-Clearly defined eligibility cut-off

Progresa Case 5: Discontinuity Design

- Progresa assigned benefits based on a poverty index

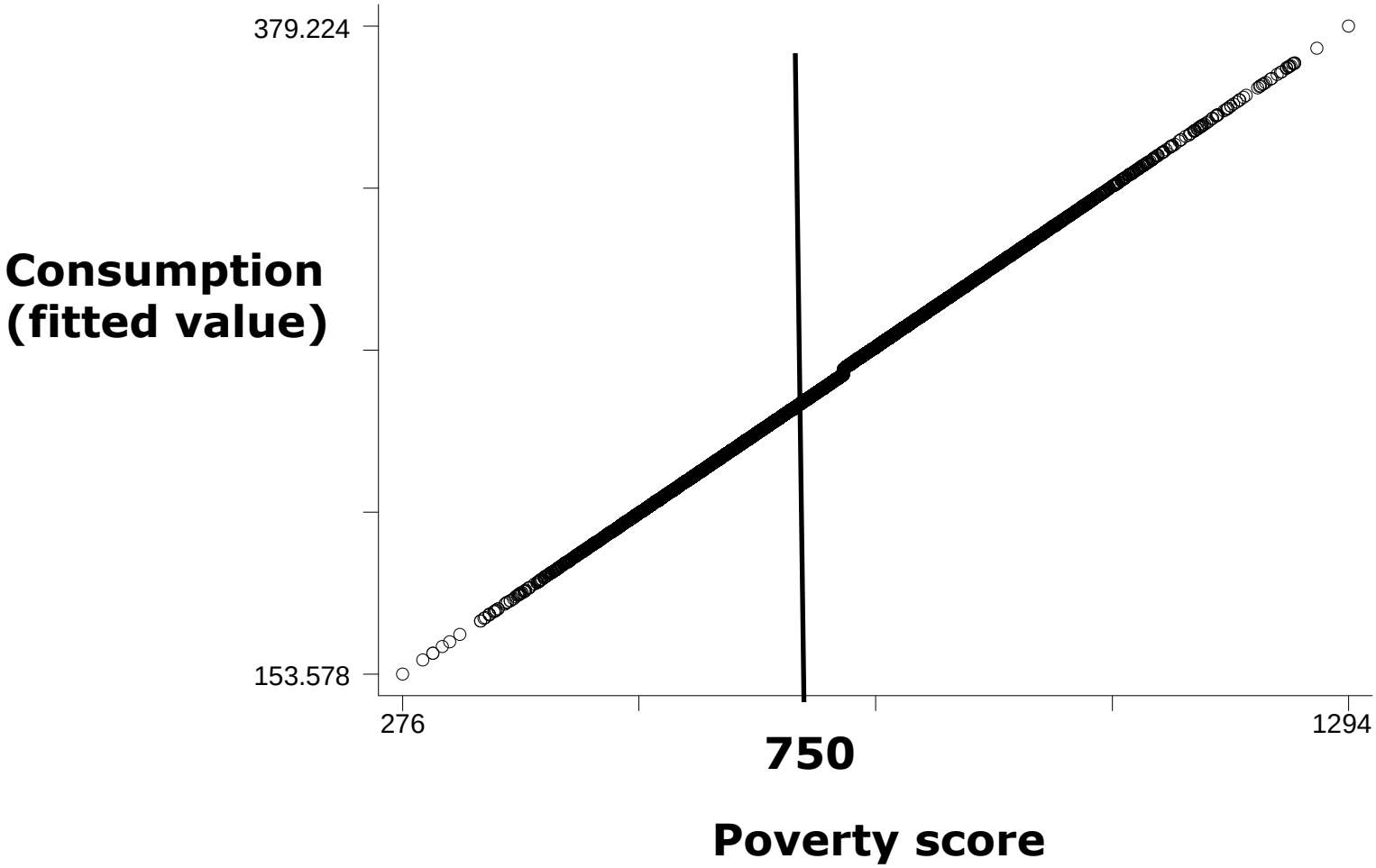
- Where

  Treatment = 1 if score <=750

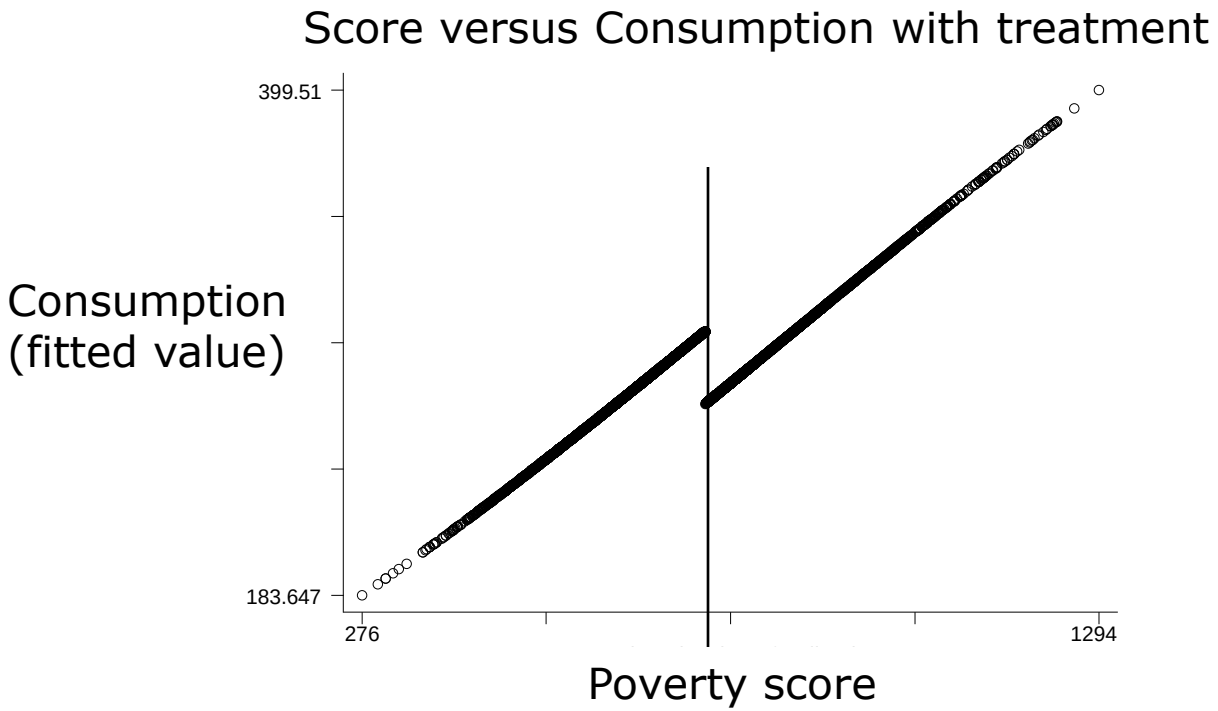
  Treatment = 0 if score >750

Progresa Case 5: Discontinuity Design

Score vs. Consumption at Baseline – No treatment



Case 5: Discontinuity Design



	Case 5: Discontinuity Design
	Multivariate linear regression
Estimated impact on mean per capita consumption	30,58** (5,93)

Potential Disadvantages of Discontinuity Design

- It's a <u>local</u> estimate
  - We estimate the effect of the program around the cut-off point / discontinuity
  - This is not always generalizable
- Power
  - Need many observations around the cut-off point
- It's easy to make mistakes in the statistical model
  - Sometimes what looks like a discontinuity in the graph, is something else…

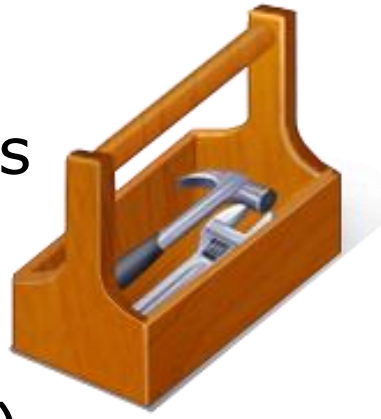
Advantages of Discontinuity Design

- Discontinuity Design gives an unbiased ("valid")estimate of the effect of treatment
  - Though only around the cut-off point
- No need to "exclude" a group of eligible households/individuals from treatment
- Can sometimes use it for programs that are already ongoing

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)



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

Y= yield of soybeans, tons per acre
P= new type of inoculant

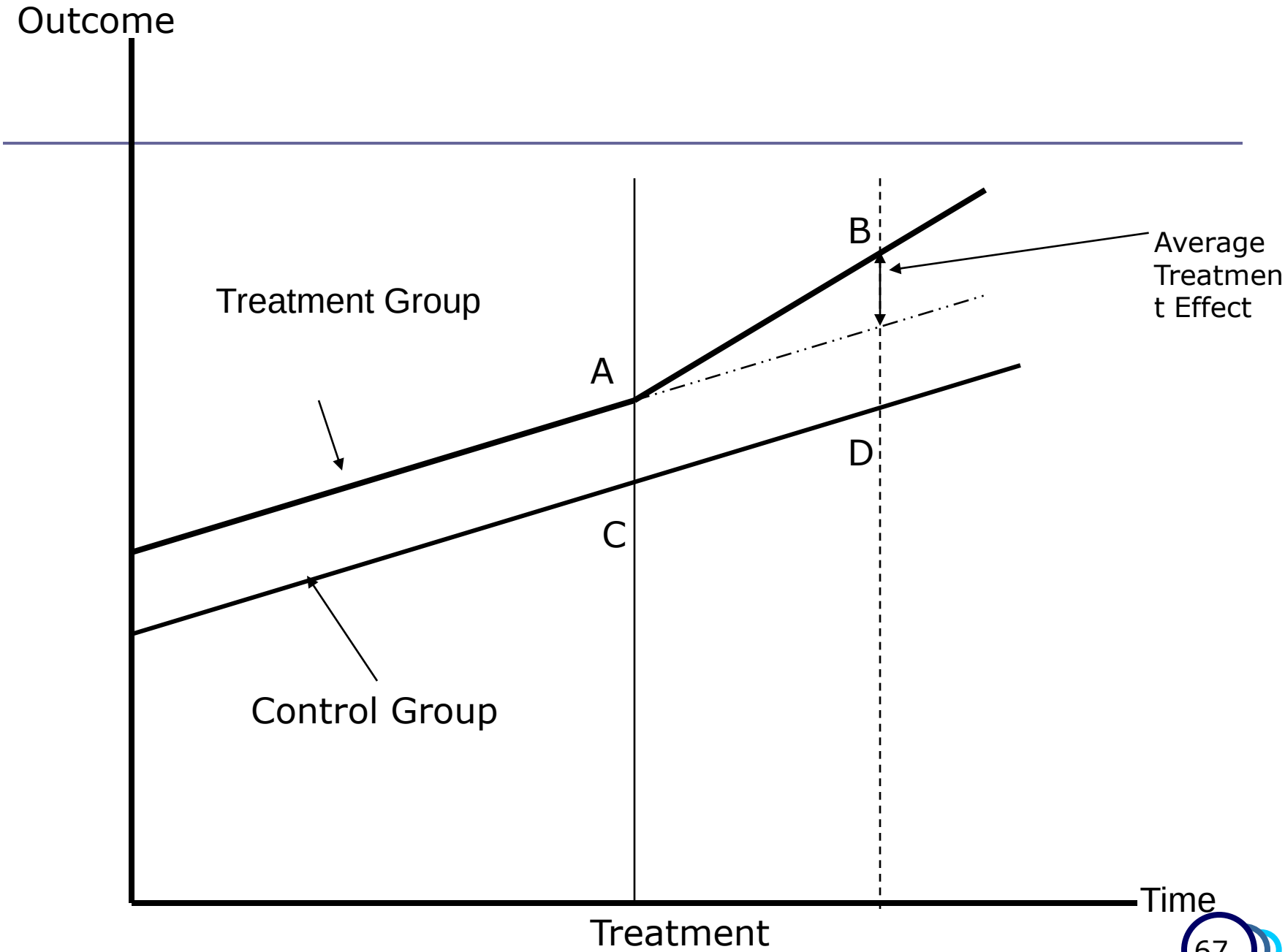
	Treatment	Comparison
After	8.49	9.76
Before	8.02	9.4
Difference	**+0.47**	**+0.36**

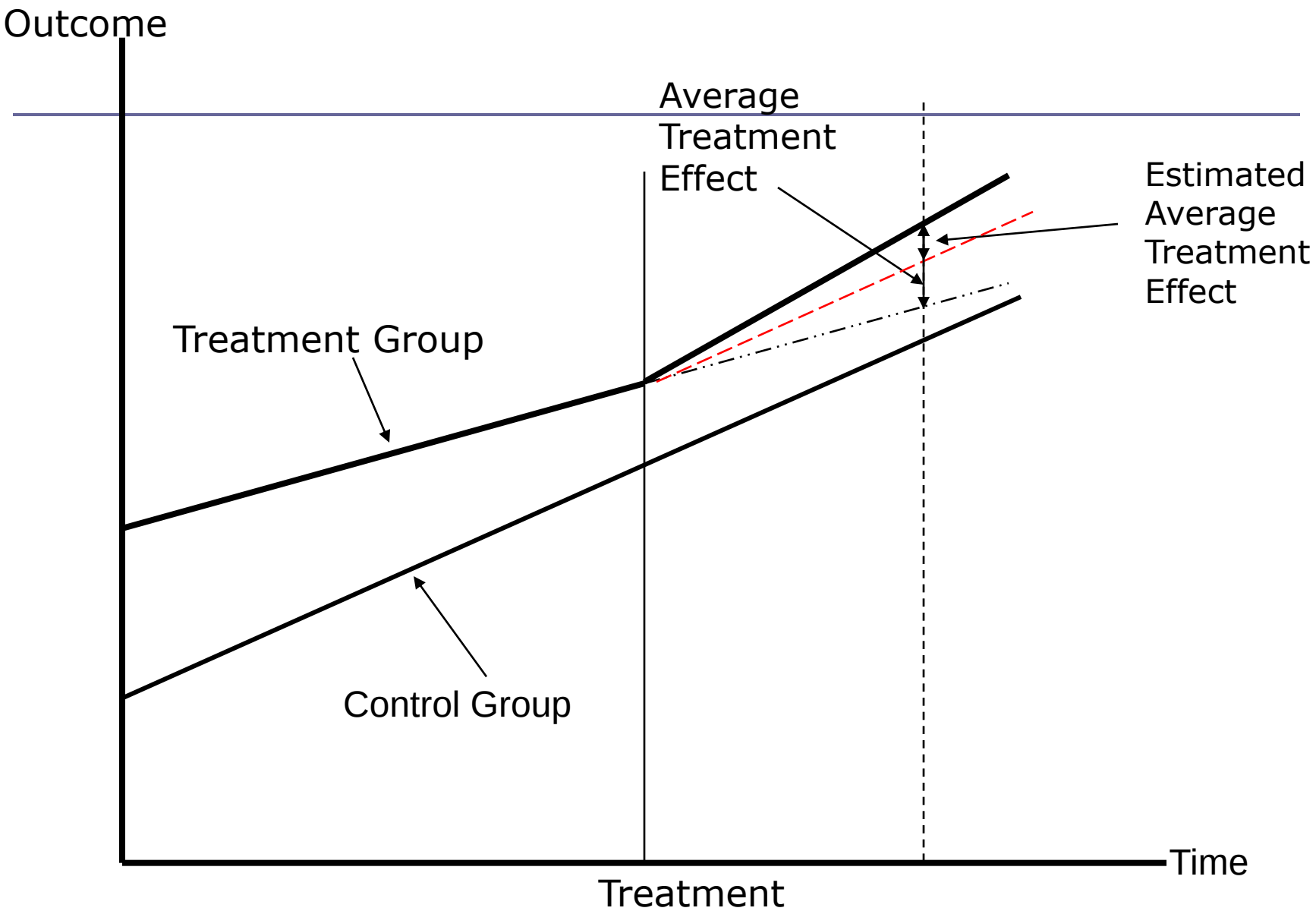
0.11

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

Y= yield of soybeans, tons per acre
P= new type of inoculant

	Treatment	Comparison	Difference
After	8.49	9.76	**-1.27**
Before	8.02	9.4	**-1.38**
			0.11


Outcome
Treatment Group
B
Average
Treatmen
t Effect
A
D
C
Control Group
Time
Treatment


Outcome
Average
Treatment
Effect
Estimated
Average
Treatment
Effect
Treatment Group
Control Group
Time
Treatment

Differences in differences (Diff-in-diff)

- Fundamental assumption:
  - Trends (slopes) are the same in treatments and controls

- Need at least three observations in time
  - Two observations “before”
  - One observation “after”

Progresa Case 6: Difference-in-Differences

	Case 6:diff-in-diff	Case 6:diff-in-diff
	Linear regression	Multivariate linear regression
Estimated impact on mean per capita consumption	26,66** (2,68)	25,53** (2,77)

Standard errors in parenthesis.
**= significant at 1%

Progresa Case Study: Cases 1-6

Estimated impact on mean per capita consumption

Case 1: Before & After	Case 2: Enrolled & not enrolled	Case 3: Randomization	Case 4: Random offering / Treatment on the treated	Case 5: Discontinuity design	Casae 6: Diff-in-Diff
Multivariate linear regression	Multivariate linear regression	Multivariate linear regression	2SLS	Multivariate linear regression	Multivariate linear regression
34,28** (2,11)	-4,15 (4,05)	29,79** (3,00)	30,44** (3,07)	30,58** (5,93)	25,53** (2,77)

Standard errors in parenthesis.
**= significant at 1%

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)

2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)

Matching

- Idea:
  - For each treated unit
  - Pick up the “best” comparison unit (“match”)
  - from a larger survey
- How?
  - Matches are selected on the basis of similarities in observed characteristics
- Issue?
  - If there are differences unobservable characteristics
  - And those unobservables influence participation
  - => selection bias!

Propensity-Score Matching (PSM)

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

- Solution proposed by Rosenbaum and Rubin:

  Match on the basis of the "propensity score":
  - Compute everyone's probability of participating, based on their observable characteristics
  - Choose matches that have the same probability of participation as the treatments

Matching versus an experiment

- Experiment ensures that participation is uncorrelated with:
  - Observable characteristics
  - AND unobservable characteristics
  - => no selection bias

- Matching
  - allows to control for the correlation between participation and <u>observable</u> characteristics
  - But if participation is also correlated with observable characteristics
  - Then we may have SELECTION BIAS

Lessons on Matching

- Matching ex-post
  - When randomization, RD, other options are not possible
  - Because there is no baseline
  - Be careful: Matching on endogenous variables gives BAD results
- Matching at baseline can be very useful:
  - combine with other techniques (i.e. diff in diff)
  - Know the assignment rule and match based on it
- Matching requires large samples and good data
  - Common support can be a problem (see annex)

Progresa Case 7: P-Score Matching

Case 7 - PROPENSITY SCORE: Pr(treatment=1)

Variable	Coef.	Std. Err.
Age Head	-0.0282433	0.0024553
Educ Head	-0.054722	0.0086369
Age Spouse	-0.0171695	0.0028683
Educ Spouse	-0.0643569	0.0093801
Ethnicity	0.4166998	0.0397539
Female Head	-0.2260407	0.0714199
_cons	1.6048	0.1013011

P-score Quintiles

	Quintile 1			Quintile 2			Quintile 3			Quintile 4			Quintile 5		
Xi	T	C	t-score	T	C	t-score	T	C	t-score	T	C	t-score	T	C	t-score
Age Head	68.04	67.45	-1.2	53.61	53.38	-0.51	44.16	44.68	1.34	37.67	38.2	1.72	32.48	32.14	-1.18
Educ Head	1.54	1.97	3.13	2.39	2.69	1.67	3.25	3.26	-0.04	3.53	3.43	-0.98	2.98	3.12	1.96
Age Spouse	55.95	55.05	-1.43	46.5	46.41	0.66	39.54	40.01	1.86	34.2	34.8	1.84	29.6	29.19	-1.44
Educ Spouse	1.89	2.19	2.47	2.61	2.64	0.31	3.17	3.19	0.23	3.34	3.26	-0.78	2.37	2.72	1.99
Ethnicity	0.16	0.11	-2.81	0.24	0.27	-1.73	0.3	0.32	1.04	0.14	0.13	-0.11	0.7	0.66	-2.3
Female Head	0.19	0.21	0.92	0.42	0.16	-1.4	0.092	0.088	-0.35	0.35	0.32	-0.34	0.008	0.008	0.83

Progresa Case 7: P-Score Matching

	Case 7: Propensity Score Matching	Case 7: Propensity Score Matching
	Linear Regression	Multivariate Linear Regression
Estimated impact on mean per capita consumption	1,16 (3,59)	7,06* (3,65)

Standard errors in parenthesis.
**= significant at 1%

Progresa Case Study: Summary

Estimated impact on mean per capita consumption

Case 1: Before & After	Case 2: Enrolled & not enrolled	Case 3: Randomization	Case 4: Random offering / Treatment on the treated	Case 5: Discontinuity design	Case 6: Diff-in-Diff	Case 7: Matching
Multivariate linear regression	Multivariate linear regression	Multivariate linear regression	2SLS	Multivariate linear regression	Multivariate linear regression	Multivariate linear regression
34,28** (2,11)	-4,15 (4,05)	29,79** (3,00)	30,44** (3,07)	30,58** (5,93)	25,53** (2,77)	7,06* (3,65)

Standard errors in parenthesis.
**= significant at 1%, *=significant at 5%

Measuring Impact

1) Causal Inference
   - Counterfactuals
   - False Counterfactuals:
     - Before & After (pre & post)
     - Enrolled & Not enrolled (apples & oranges)
2) IE Methods Toolbox:
   - Randomized Treatments and Controls
   - Randomized Promotion
   - Discontinuity Design
   - Difference in Differences (Diff-in-diff)
   - Matching (P-score matching)

Combination of the above

Choosing your Method

		Limited resources (cannot treat all target population)	Full resources (can treat all target population)
Roll out in Stages	**Without** eligibility cut-off	Randomization	Randomized Rollout
	With eligibility cut-off	Randomization Discontinuity design	Randomized Rollout Discontinuity design
Roll out Immediately or *Already rolled out*	**Without** eligibility cut-off	Randomization Random Promotion Match + Diff in Diff	Random Promotion Match + Diff in Diff
	With eligibility cut-off	Randomization Discontinuity design	Randomized Promotion Discontinuity design

Remember.....

- Objective of impact evaluation is to estimate the *CAUSAL* effect or *IMPACT* of a program on outcomes of interest
- In designing the program we must understand the data generation process
  - behavioral process that generates the data
  - how benefits are assigned
- Fit the best evaluation design to the operational context

THANK YOU!

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*

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.*

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.

Appendix 2: Common support in propensity score matching