Impact Evaluation Toolbox

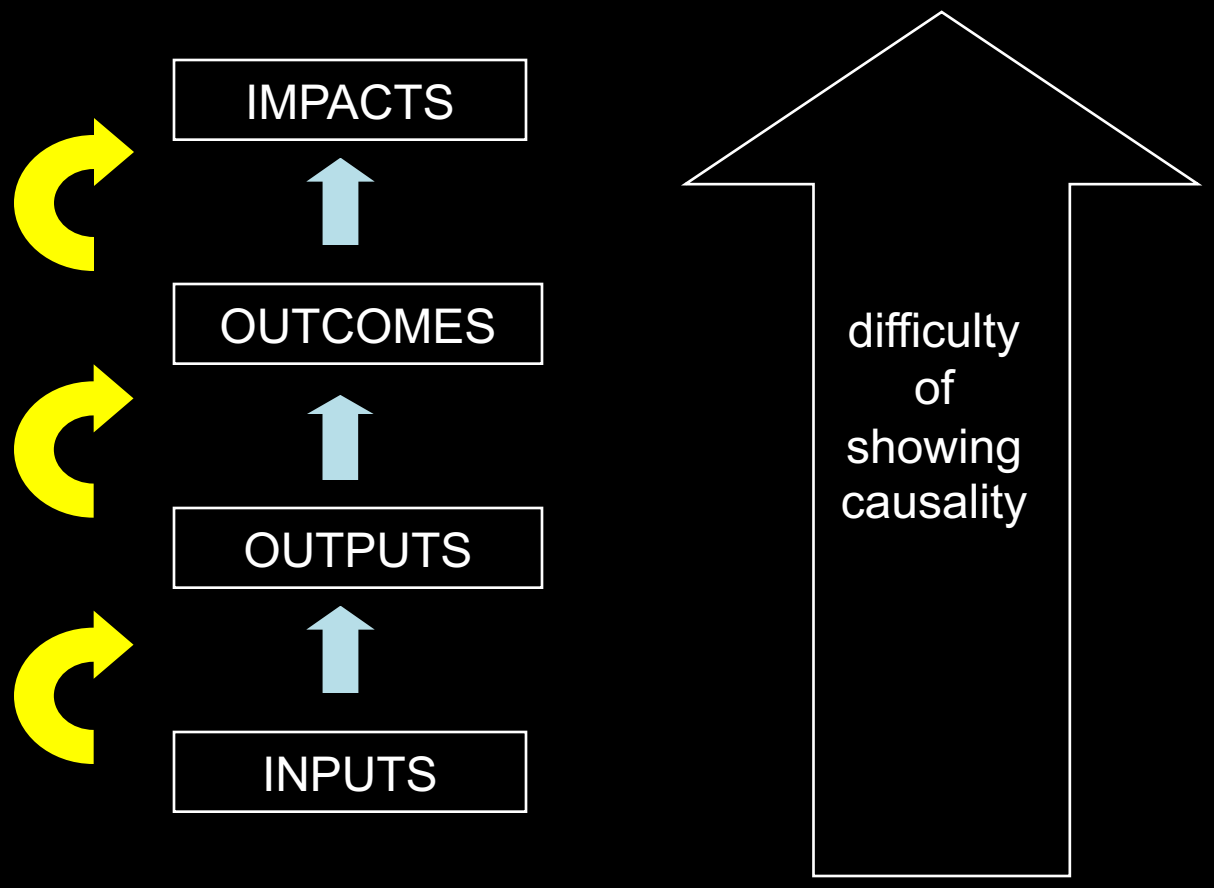
Gautam Rao
University of California, Berkeley

* ** Presentation credit: Temina Madon

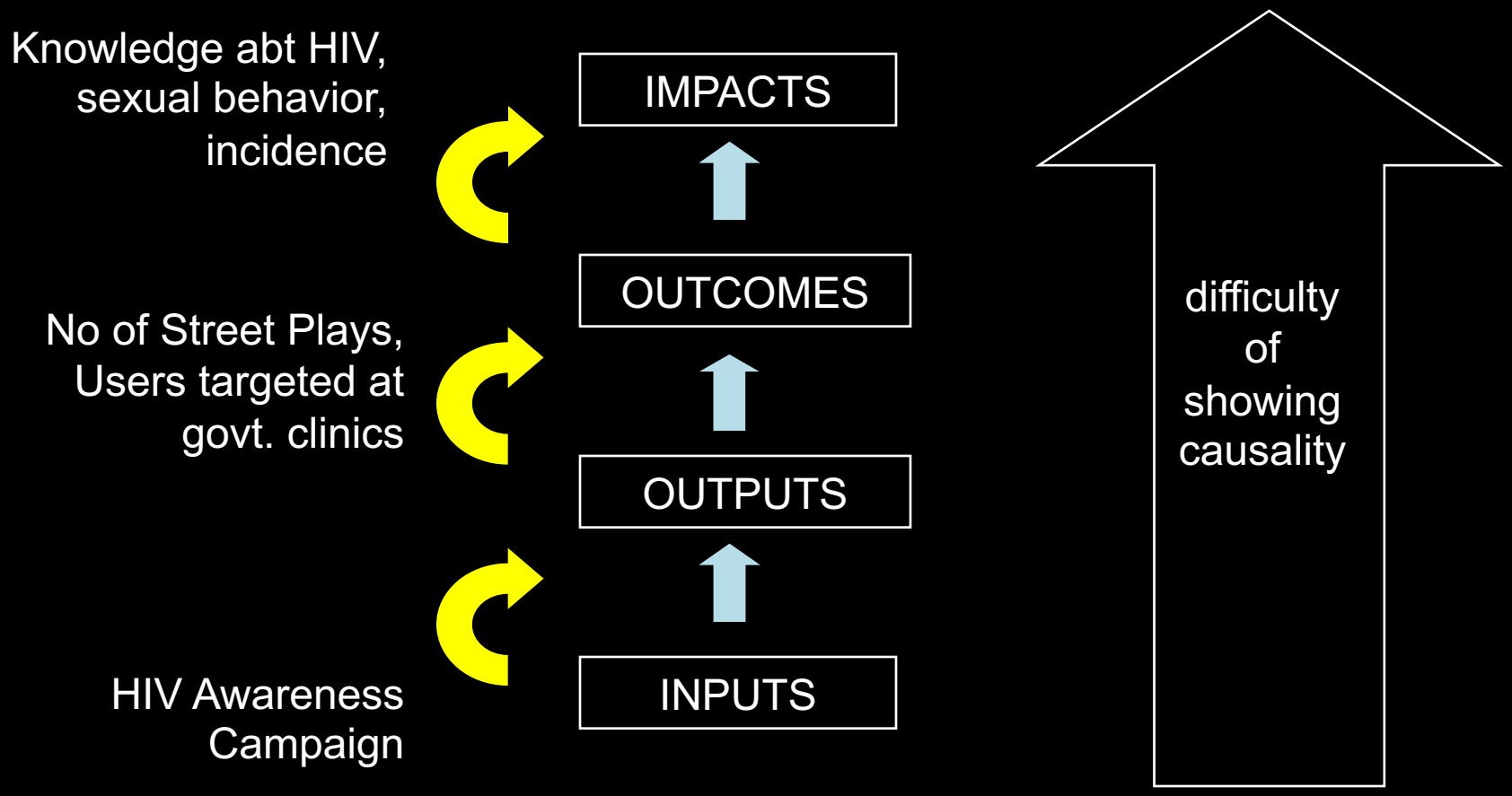
Impact Evaluation

1) The “final” outcomes we care about
   - Identify and measure them

Measuring Impacts



Measuring Impacts



Impact Evaluation

1) The "final" outcomes we care about
   - Identify and measure them

2) True "causal" effect of the intervention
   - *Counterfactual*: What would have happened in the absence of the intervention?
   - Compare measured outcomes with counterfactual → Causal effect

Toolbox for Impact Evaluation

<u>Non or Quasi-Experimental</u>

1) Before vs. After
2) With / Without Program
3) Difference –in-Difference
4) Discontinuity Methods
5) Multivariate Regression
6) Instrumental Variable

<u>Experimental Method (Gold Standard)</u>

7) Randomized Evaluation

Naïve Comparisons

1) Before-After the program

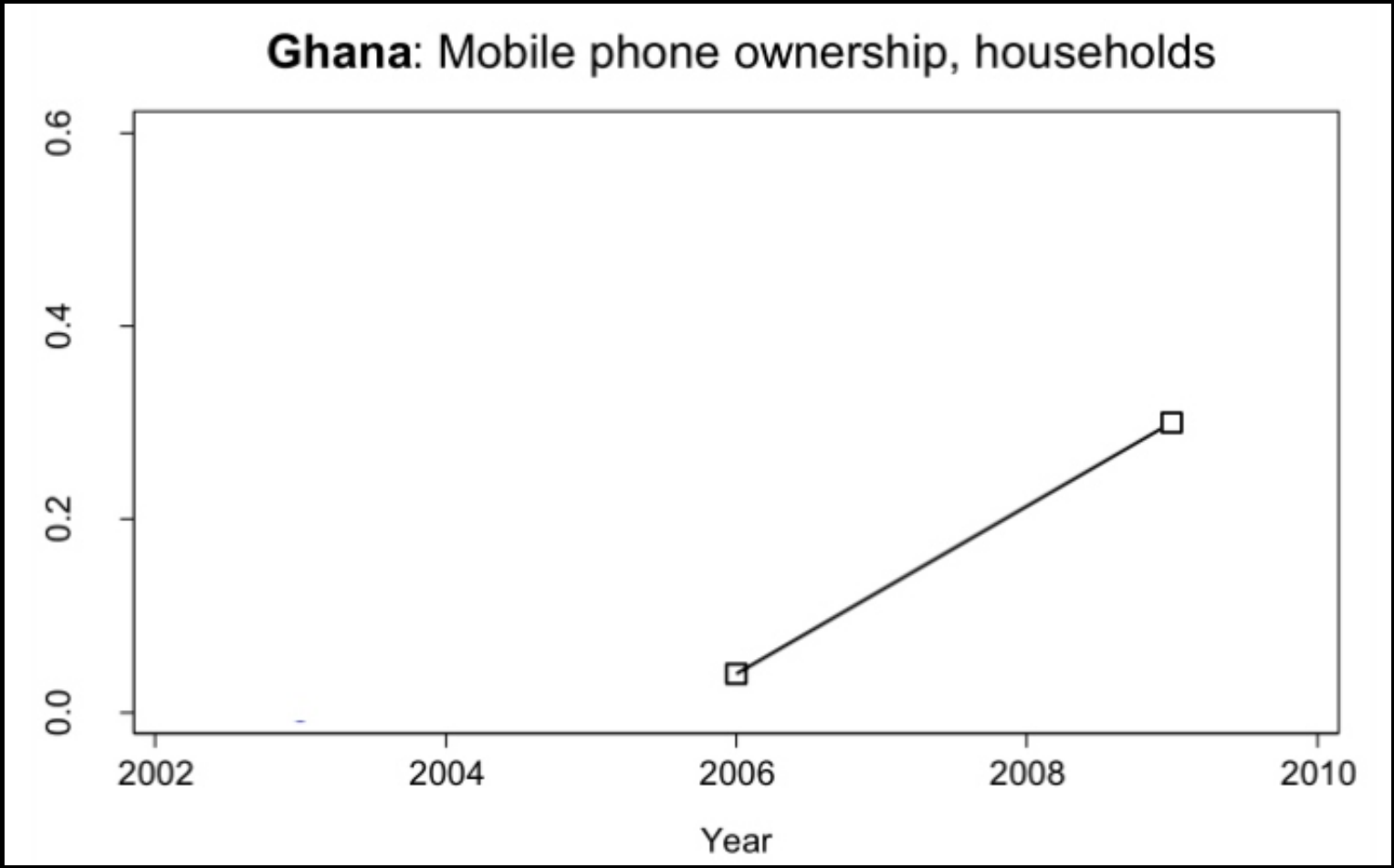
2) With-Without the program

An Example – Millennium Village Project (MVP)

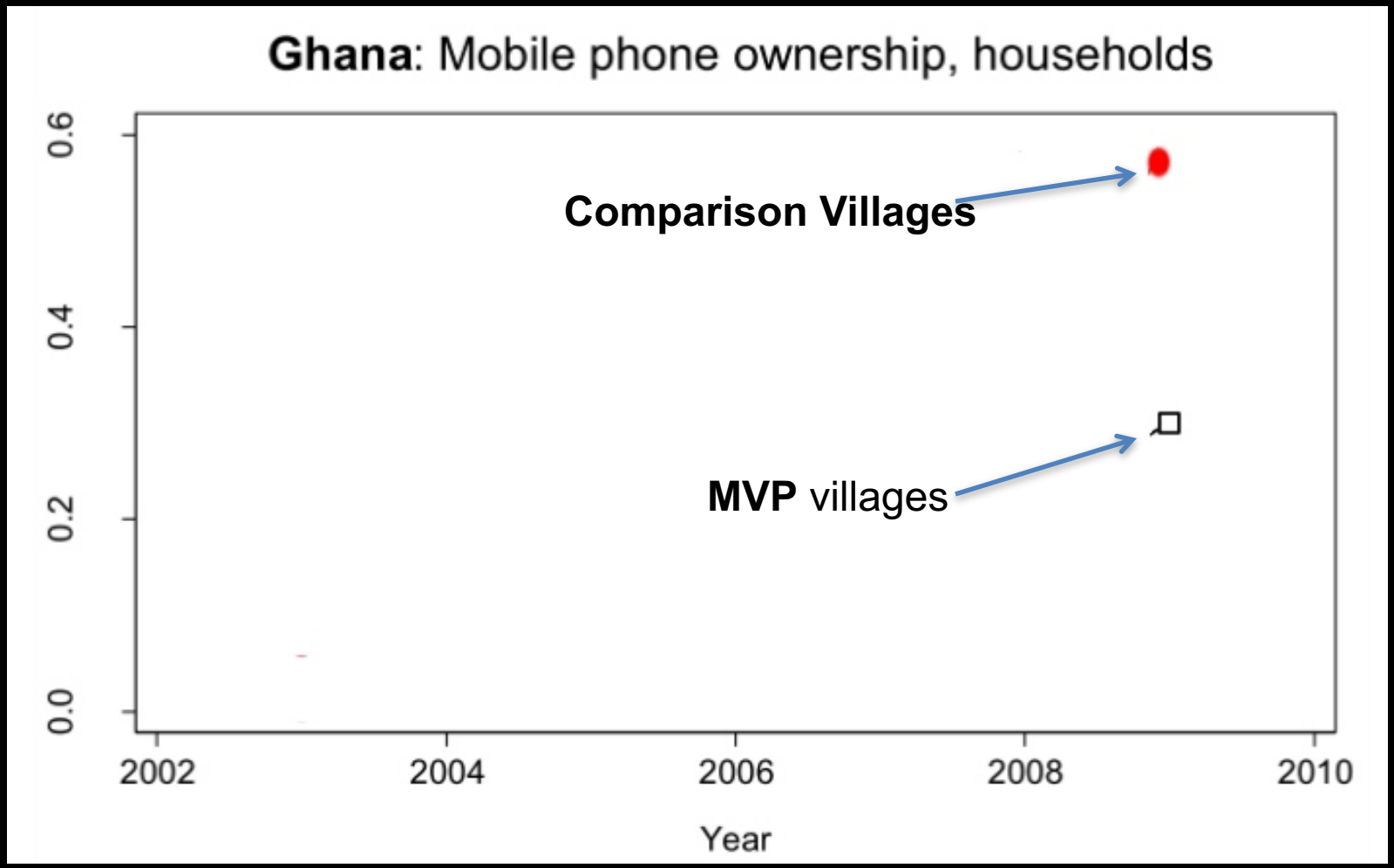
Intensive package intervention to spark development in rural Africa

- 2005 to 2010 (and ongoing)
- Non randomly selected sites
  - Poorer villages targeted

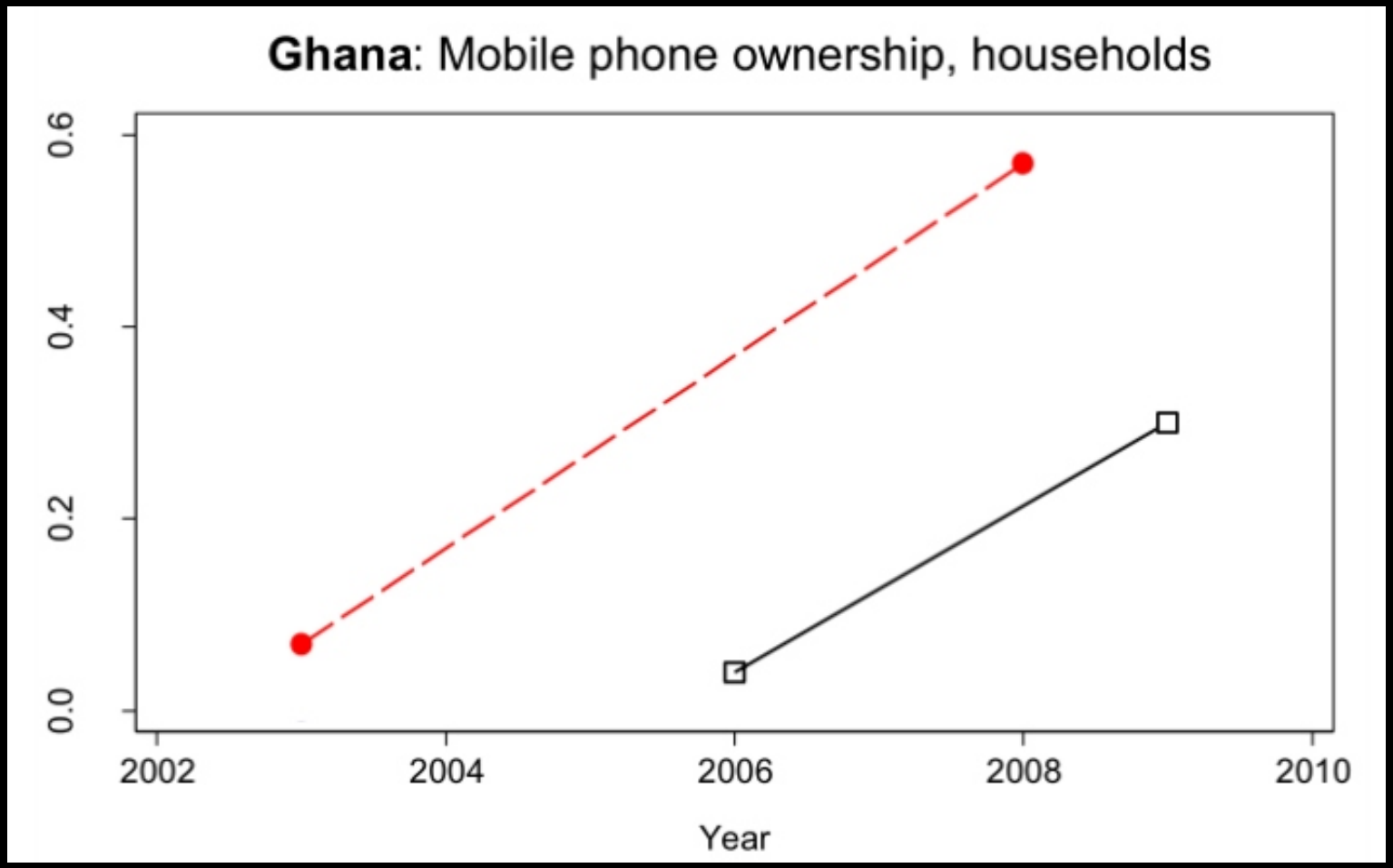
Before vs After



With vs. Without Program



With a little more data



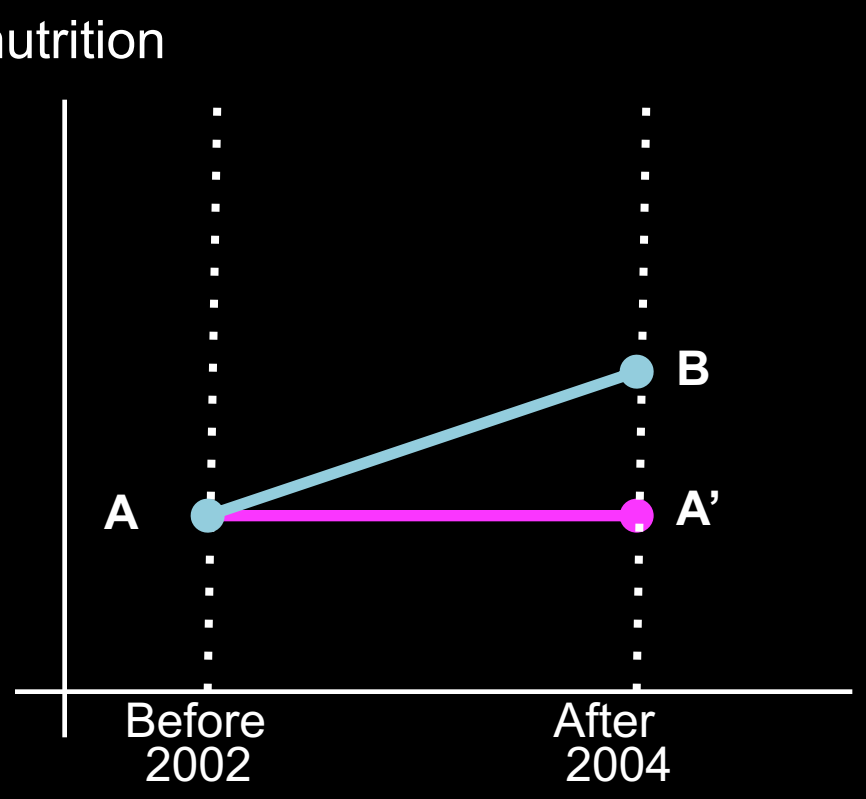
Before-After Comparison

Compare before and after intervention:

A′ = Assumed outcome with no intervention

B = Observed outcome

B-A = Estimated impact



We observe that villagers provided with training experience an ***increase*** in malnutrition from 2002 to 2004.

With-Without Comparison

Now compare villagers in the program region A to those in another region B. We find that our "program" villagers have a larger increase in malnutrition than those in region B.

Did the program have a negative impact?

–Not necessarily (remember program placement!)

- Region B is closer to the food aid distribution point, so villagers can access food easily (observable)
- Region A is less collectivized–villagers are less likely to share food with their malnourished neighbors (unobservable)

Problems with Simple Comparisons

Before-After

Problem: Many things change over time, not just the project. ***Omitted Variables.***

With-Without

Problem: Comparing oranges with apples. Why did the enrollees decide to enroll? ***Selection bias.***

Another example of Selection Bias

Impact of Health Insurance

- Can we just compare those with and without health insurance?
  - Who purchases insurance?
  - What happens if we compare health of those who sign up, with those who don't?

Multivariate Regression

- Change in outcome in the treatment group controlling for observable characteristics
- requires theorizing on what observable characteristics may impact the outcome of interest besides the programme
- Issues:
  - how many characteristics can be accounted for? (omission variable bias)
  - requires a large sample if many factors are to be controlled for

Multivariate Regression

Compare before and after intervention:

A′ = Assumed outcome with no intervention

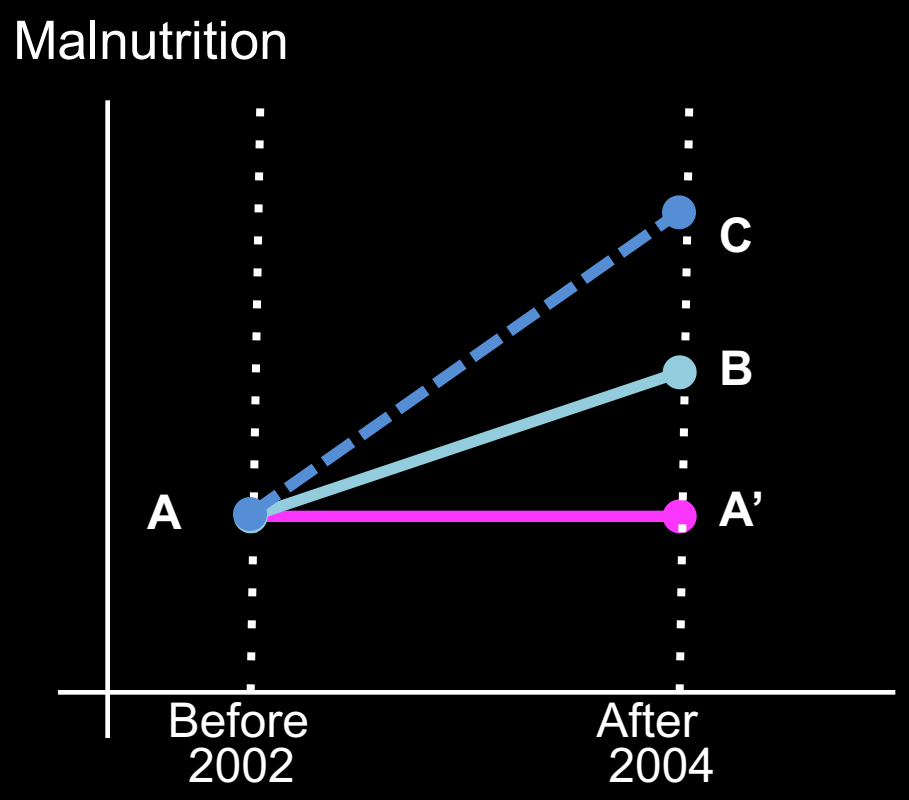
B = Observed outcome

B-A = Estimated impact

Control for other factors

C = Actual outcome with no intervention (control)

B-C = True Impact



To Identify Causal Links:

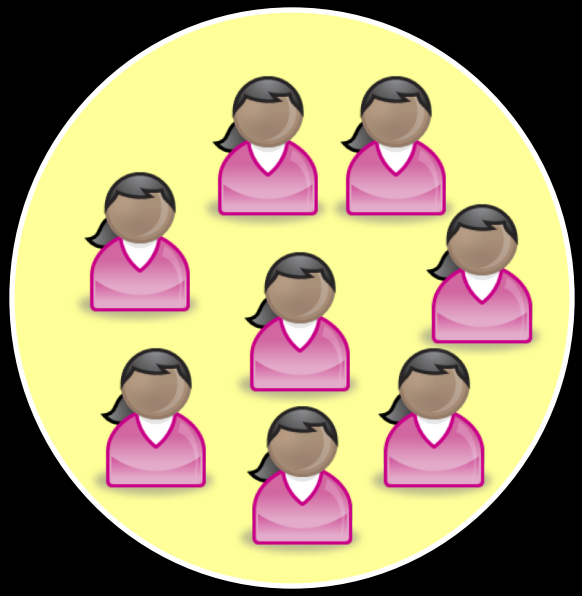
We need to know:

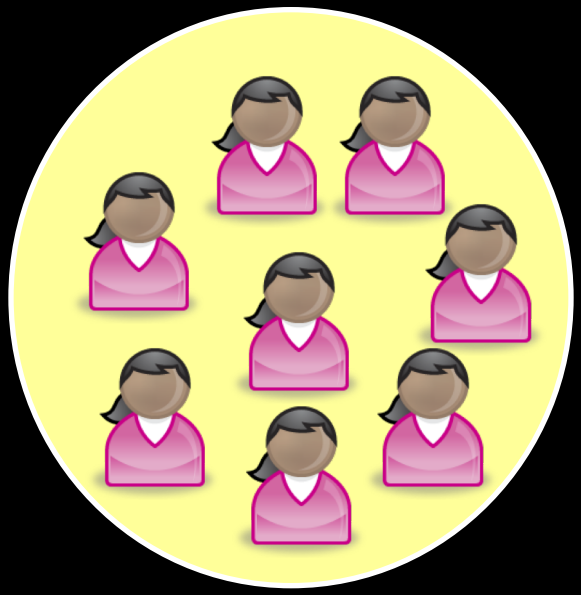
- The change in outcomes for the treatment group
- What would have happened in the absence of the treatment ("counterfactual")

At baseline, comparison group must be identical (in observable and unobservable dimensions) to those receiving the program.

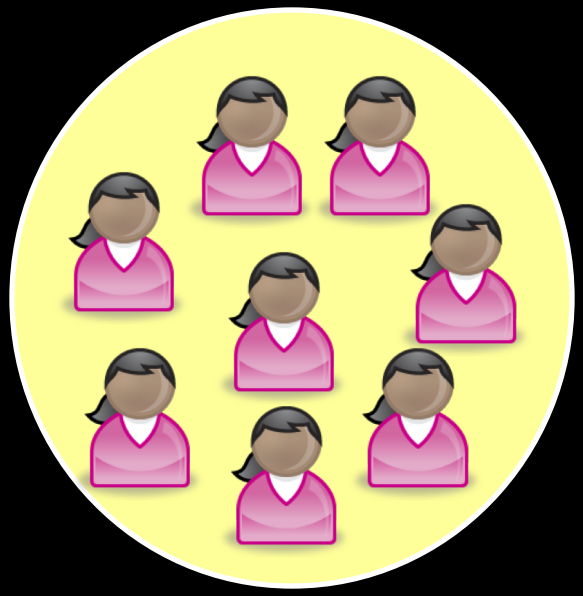
Creating the Counterfactual

Women aged 14-25 years





What we can observe



Women aged 14-25 years

Screened by:
- income
- education level
- ethnicity
- marital status
- employment

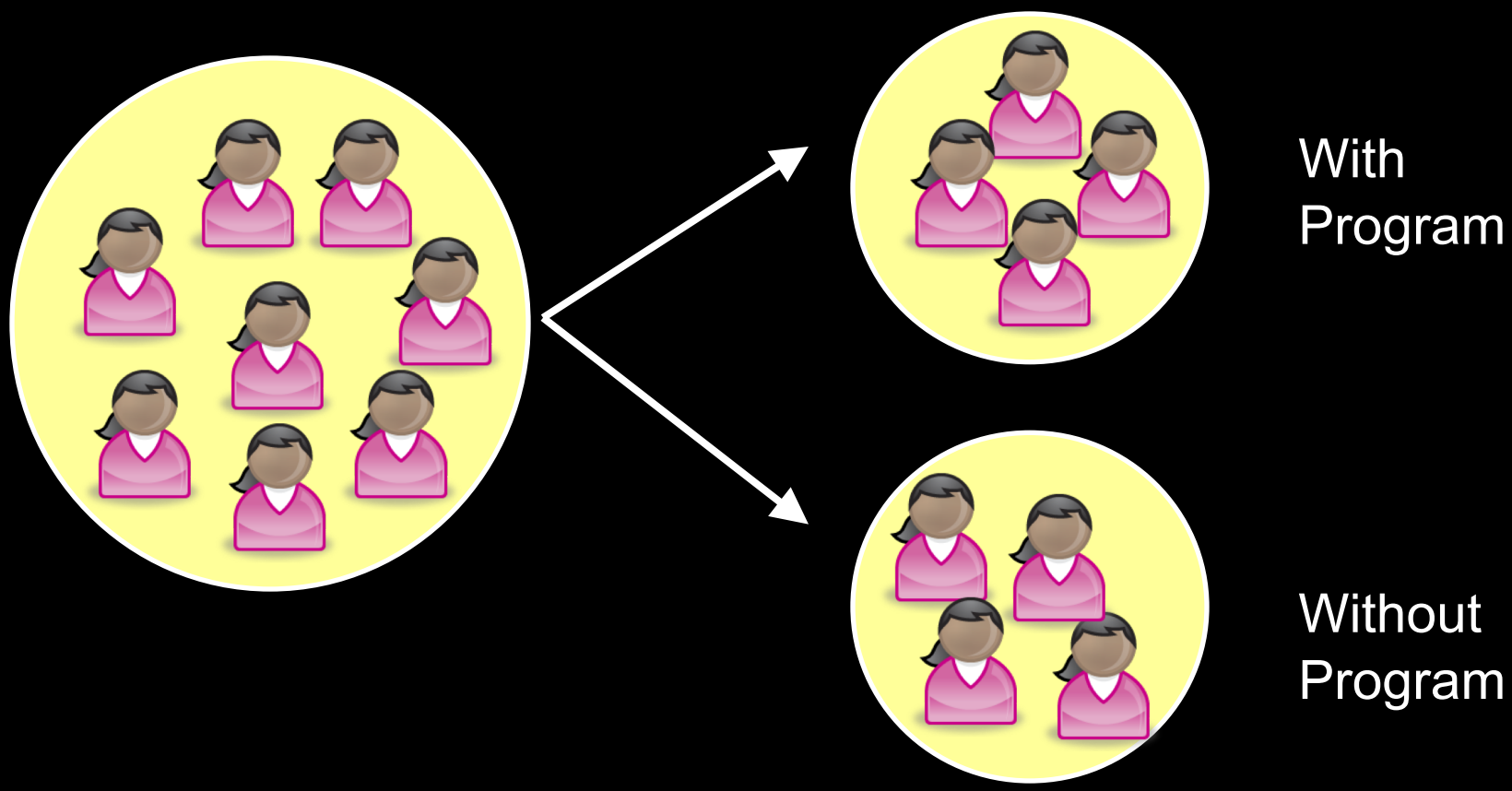
What we can’t observe

Women aged 14-25 years

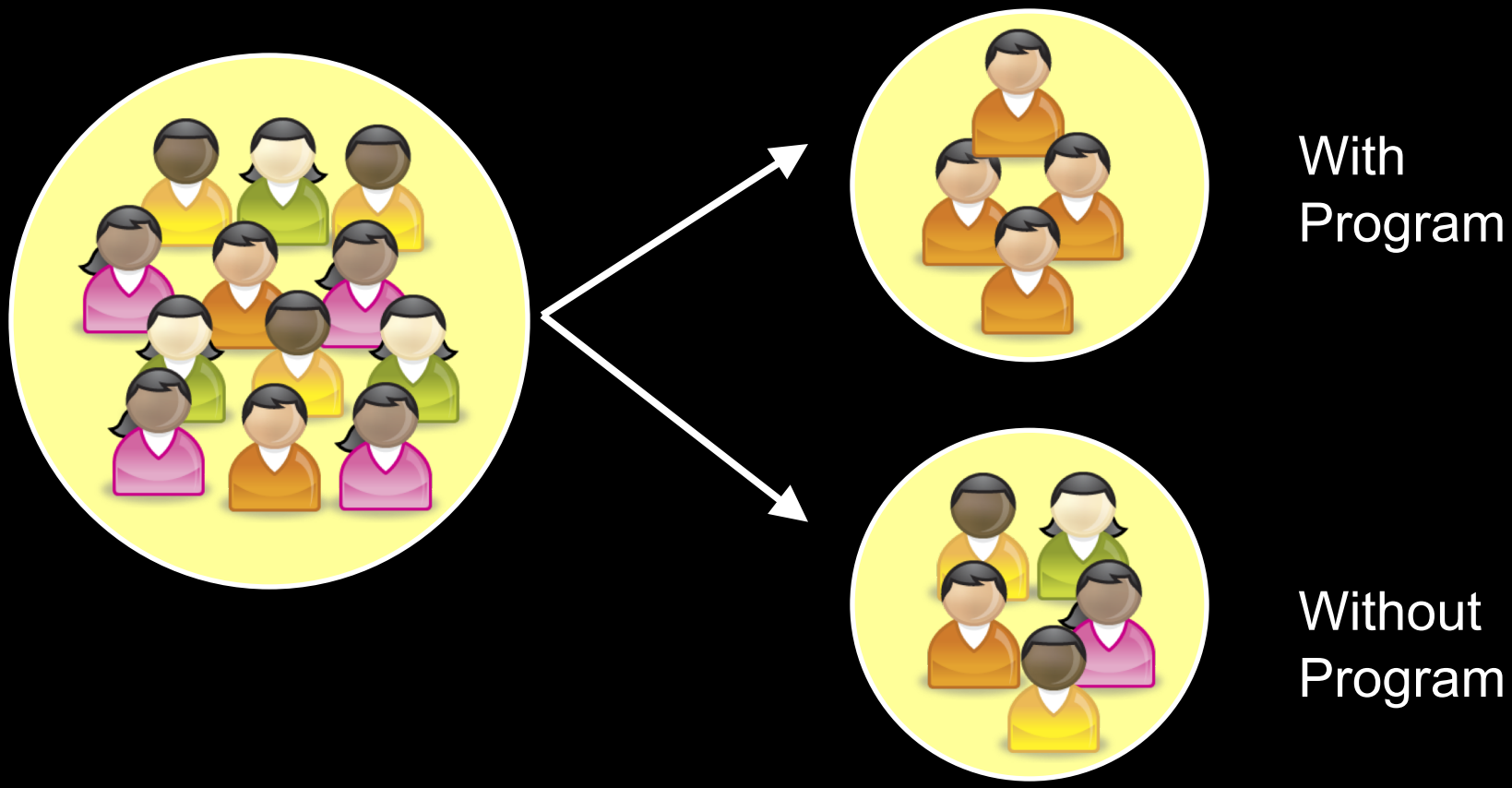
What we can’t observe:

- risk tolerance
- entrepreneurialism
- generosity
- respect for authority

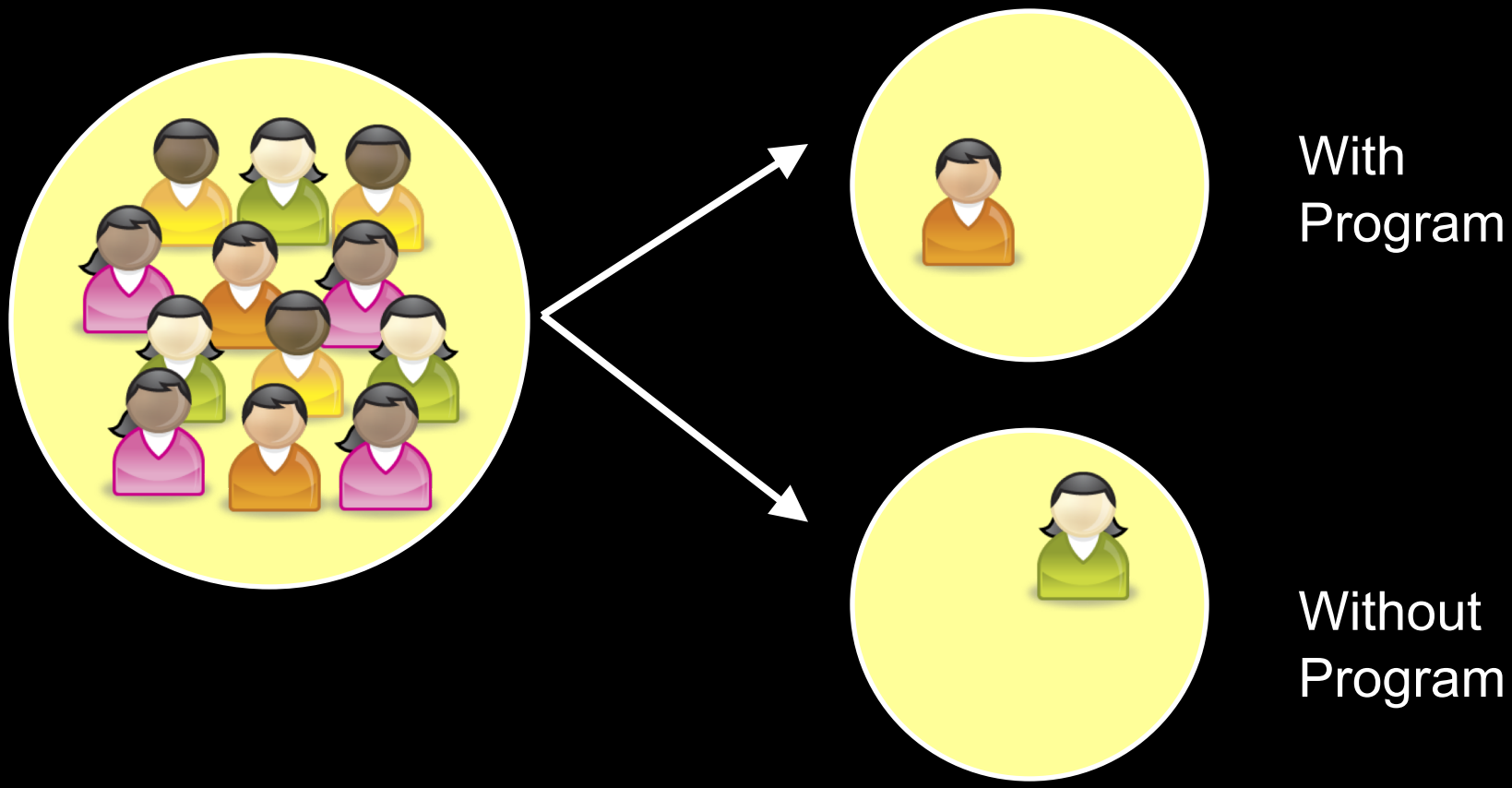
Selection based on Observables



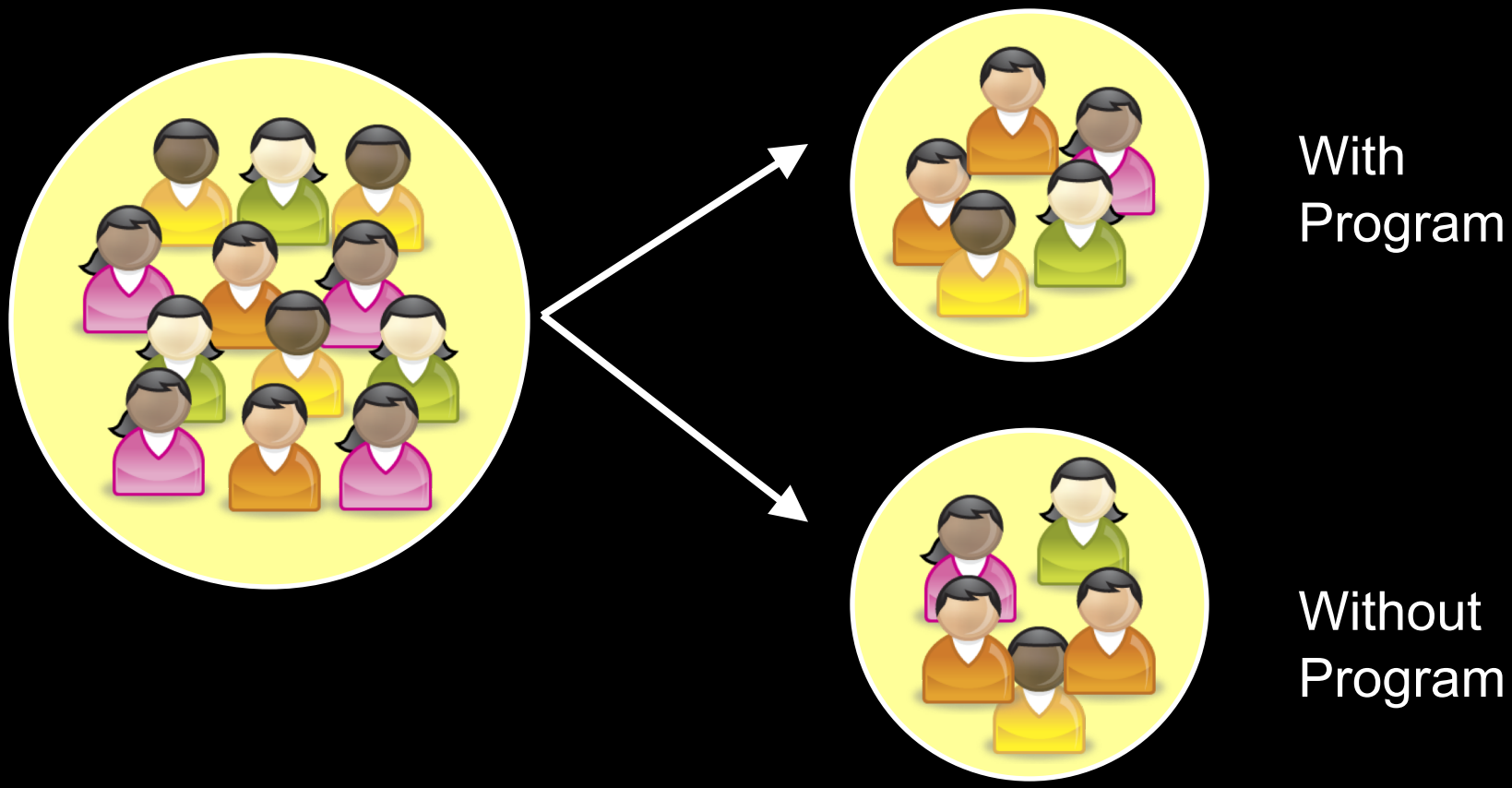
What can happen (unobservables)



Randomization



Randomization



Methods & Study Designs

Methods Toolbox	Study Designs
Randomization	Clustering
Regression Discontinuity	Phased Roll-Out
Difference in Differences	Selective promotion
Matching	Variation in Treatments
Instrumental Variables	

Methods Toolbox

Randomization

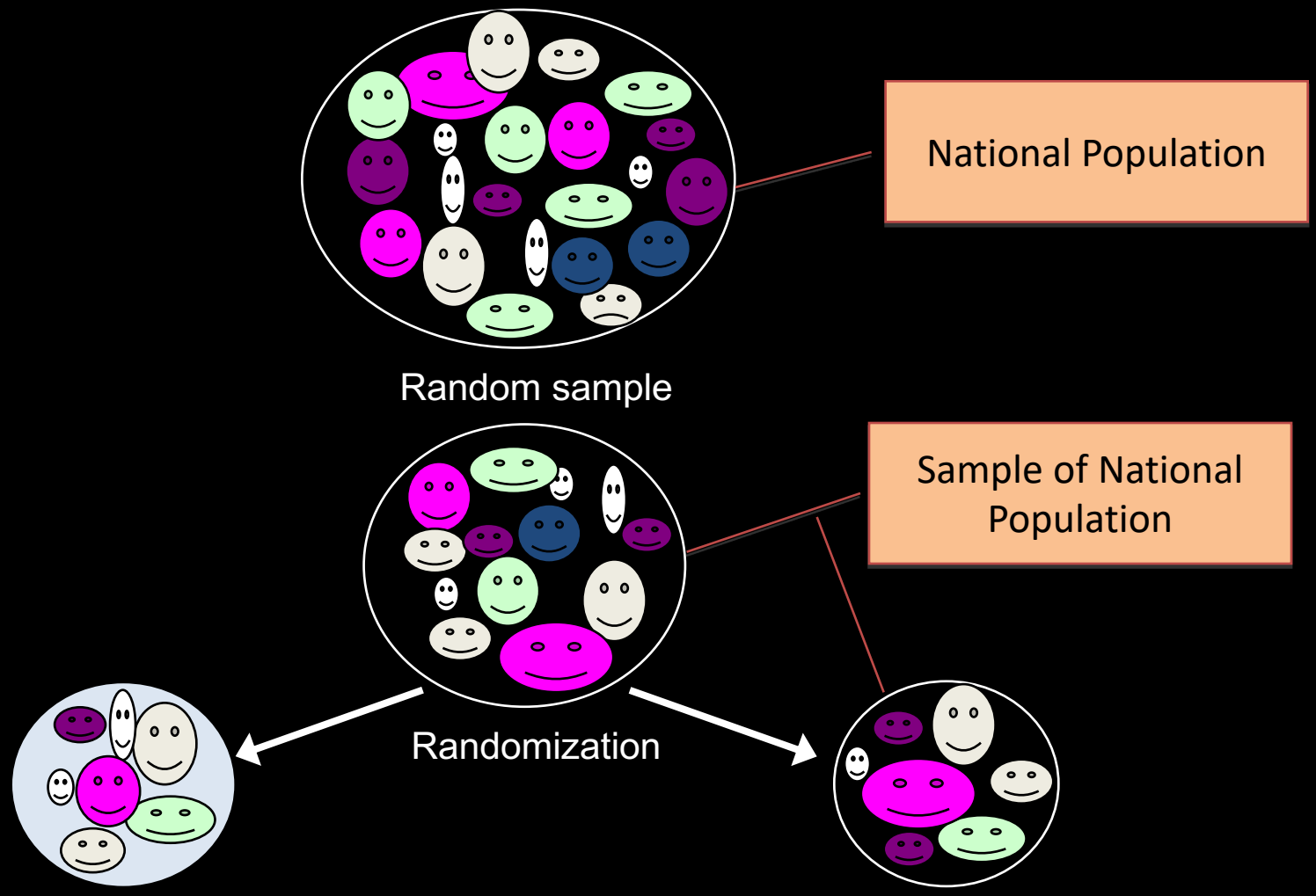
- Use a lottery to give all people an equal chance of being in control or treatment groups
- With a large enough sample, it guarantees that all factors/characteristics will be equal between groups, on average
- Only difference between 2 groups is the intervention itself

"Gold Standard"

Ensuring Validity

EXTERNAL VALIDITY

INTERNAL VALIDITY



Ensuring Validity

Internal Validity:

- Have we estimated a causal effect?
- Have we identified a valid counterfactual?

External Validity:

- Is our estimate of the impact "generalizable"
- e.g. will IE results from Kerala generalize to Punjab?

Methods Toolbox

Regression discontinuity

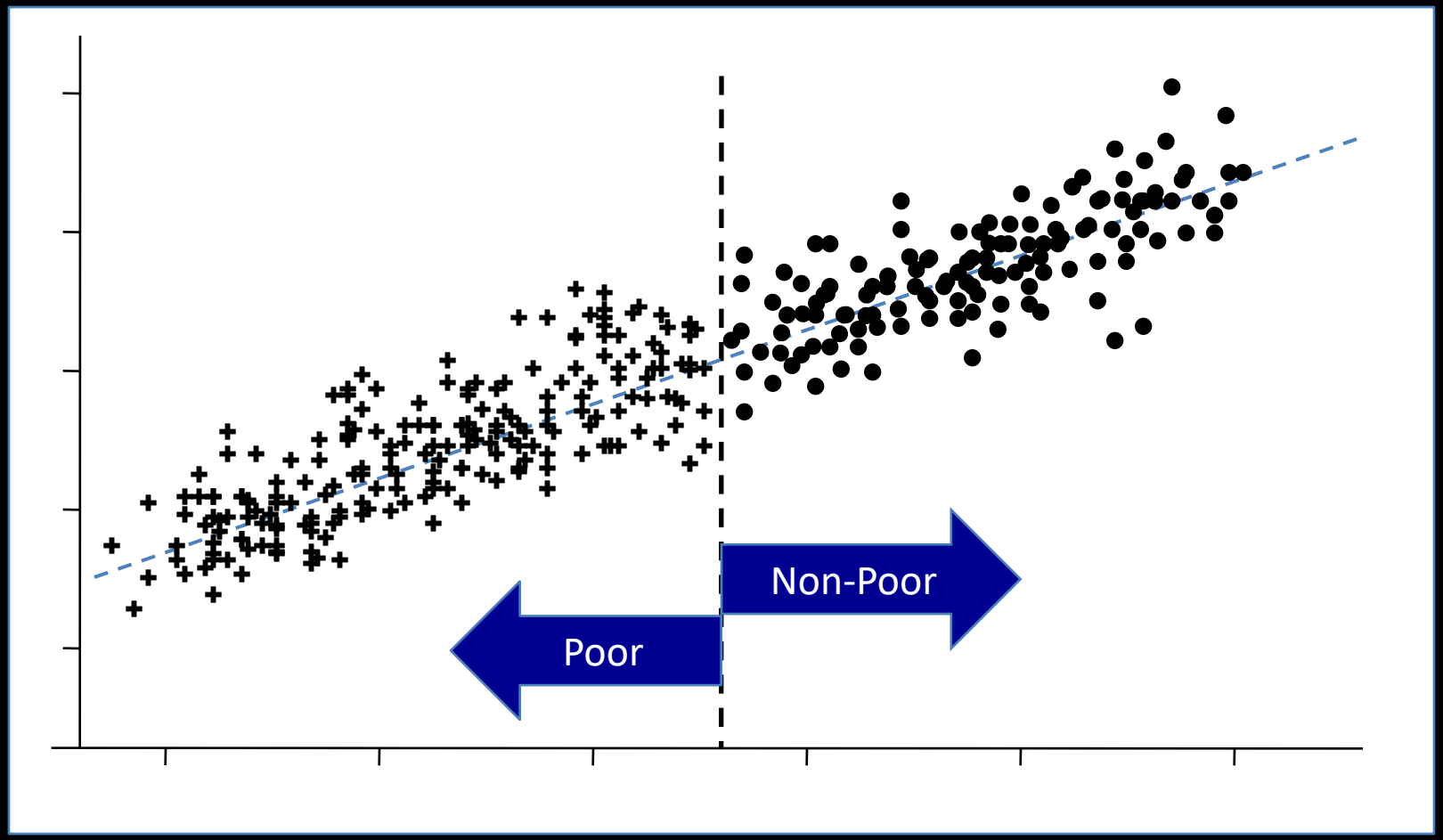
When you can't randomize… use a transparent & observable criterion for who is offered a program

Examples: age, income eligibility, test scores (scholarship), political borders

<u>Assumption</u>: There is a discontinuity in program participation, but we assume it doesn't impact outcomes of interest.

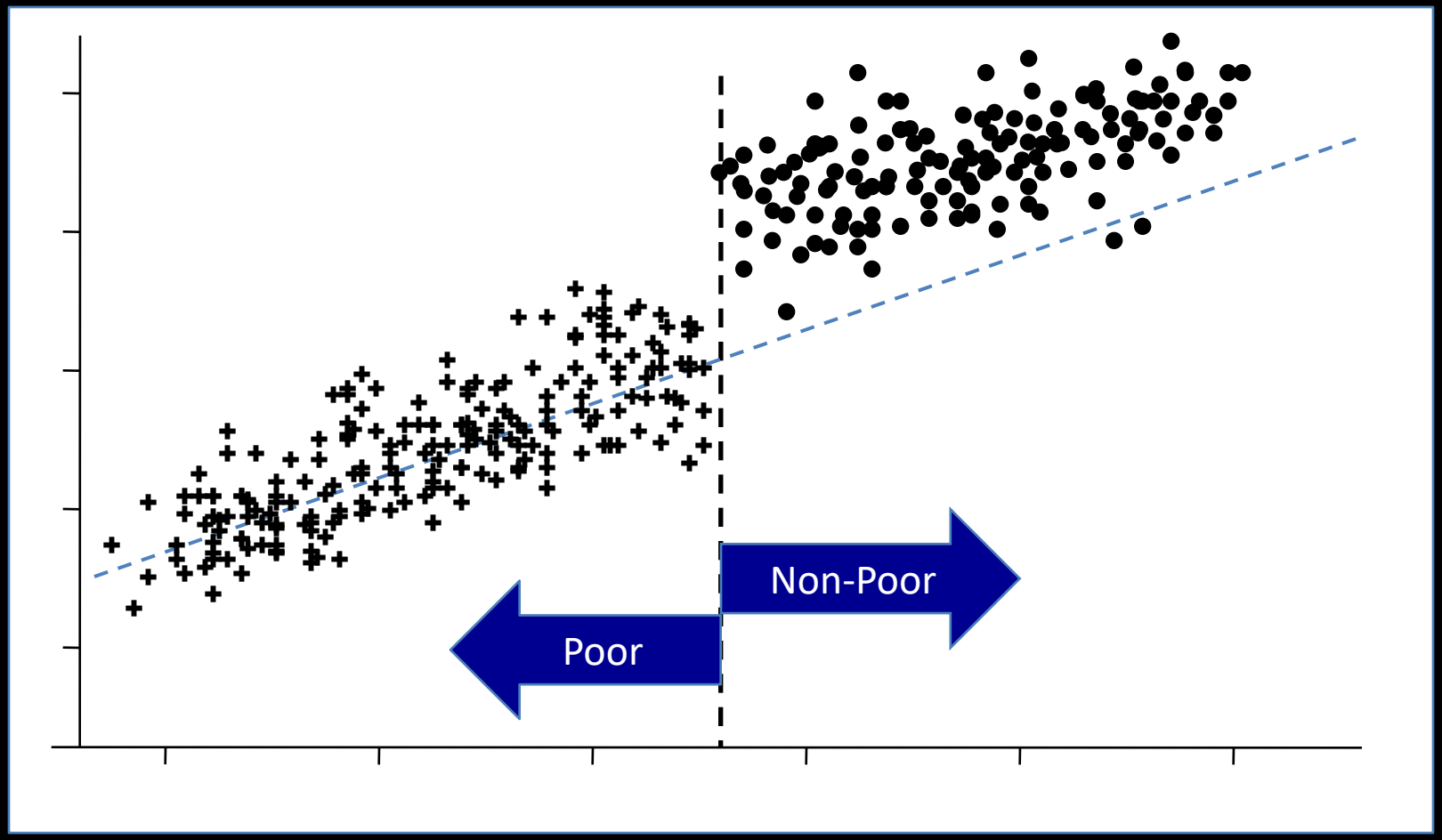
Methods Toolbox

Regression discontinuity



Methods Toolbox

Regression discontinuity



Methods Toolbox

Regression discontinuity

Example: South Africa pensions program

Age	Women	Men
55-59	16%	5%
60-64	77%	22%
65+		60%

What happens to children living in these households?

GIRLS: pension-eligible women vs. almost eligible	0.6 s.d. heavier for age
BOYS: pension-eligible women vs. almost-eligible	0.3 s.d. heavier for age
With pension-eligible men vs almost-eligible	No diff.

Methods Toolbox

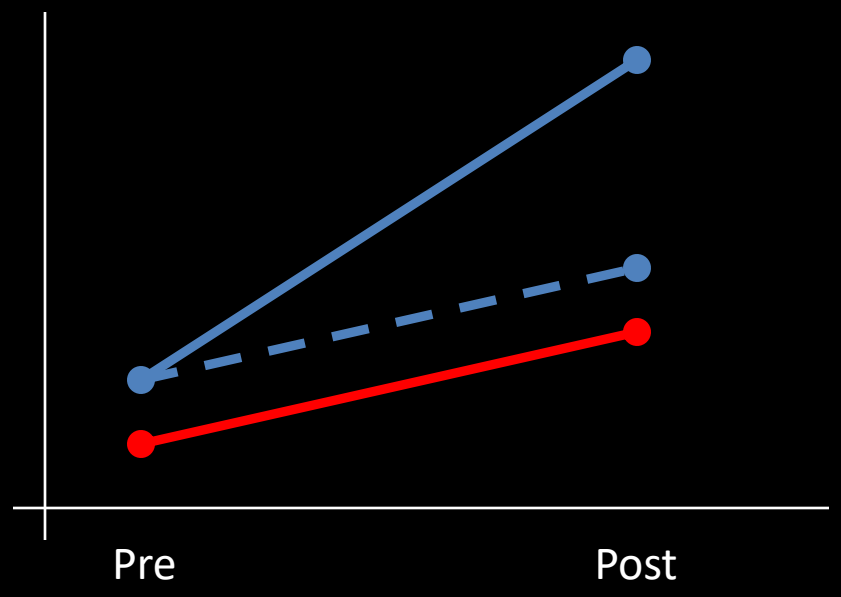
Regression discontinuity

Limitation: Poor generalizability. Only measures those near the threshold.

Must have well-enforced eligibility rule.

Methods Toolbox

Difference in differences



2-way comparison of observed changes in outcomes (Before-After) for a sample of participants and non-participants (With-Without)

Methods Toolbox

Difference in differences

Big assumption: In absence of program, participants and non-participants would have experienced the same changes.

Robustness check: Compare the 2 groups across several periods prior to intervention; compare variables unrelated to program pre/post

Can be combined with randomization, regression discontinuity, matching methods

Methods Toolbox

Time Series Analysis

Also called: Interrupted time series, trend break analysis

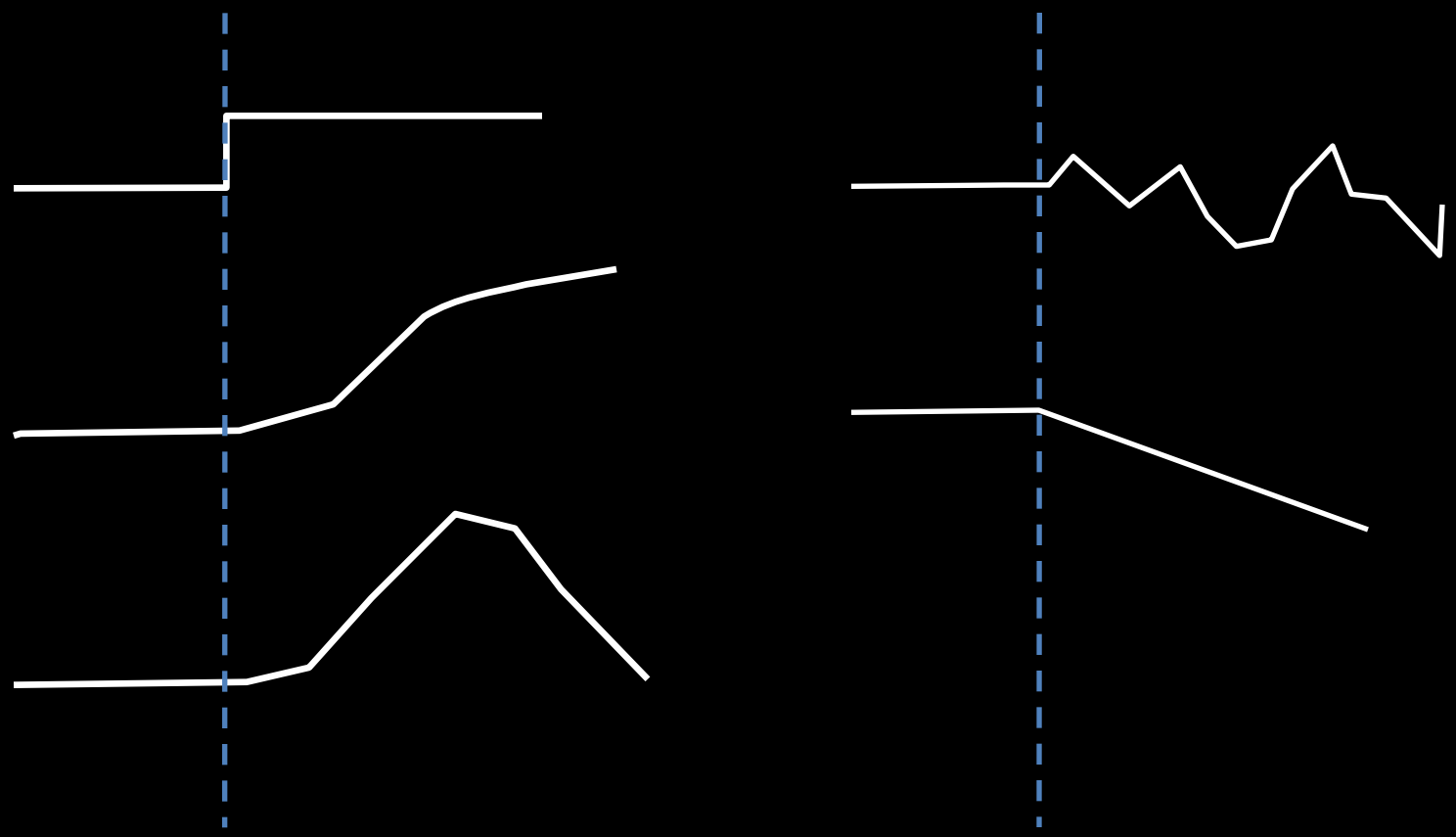
- Collect data on a continuous variable, across numerous consecutive periods (high frequency), to detect changes over time correlated with intervention.

- Challenges: Expectation should be well-defined upfront. Often need to filter out noise.

- Variant: Panel data analysis

Methods Toolbox

Time Series Analysis



Methods Toolbox

Matching

Pair each program participant with one or more non-participants, based on observable characteristics

Big assumption required: In absence of program, participants and non-participants would have been the same. (No unobservable traits influence participation)

**Combine with difference in differences

Methods Toolbox

Matching

Propensity score assigned:

score(xi) = Pr(Zi = 1|Xi = xi)

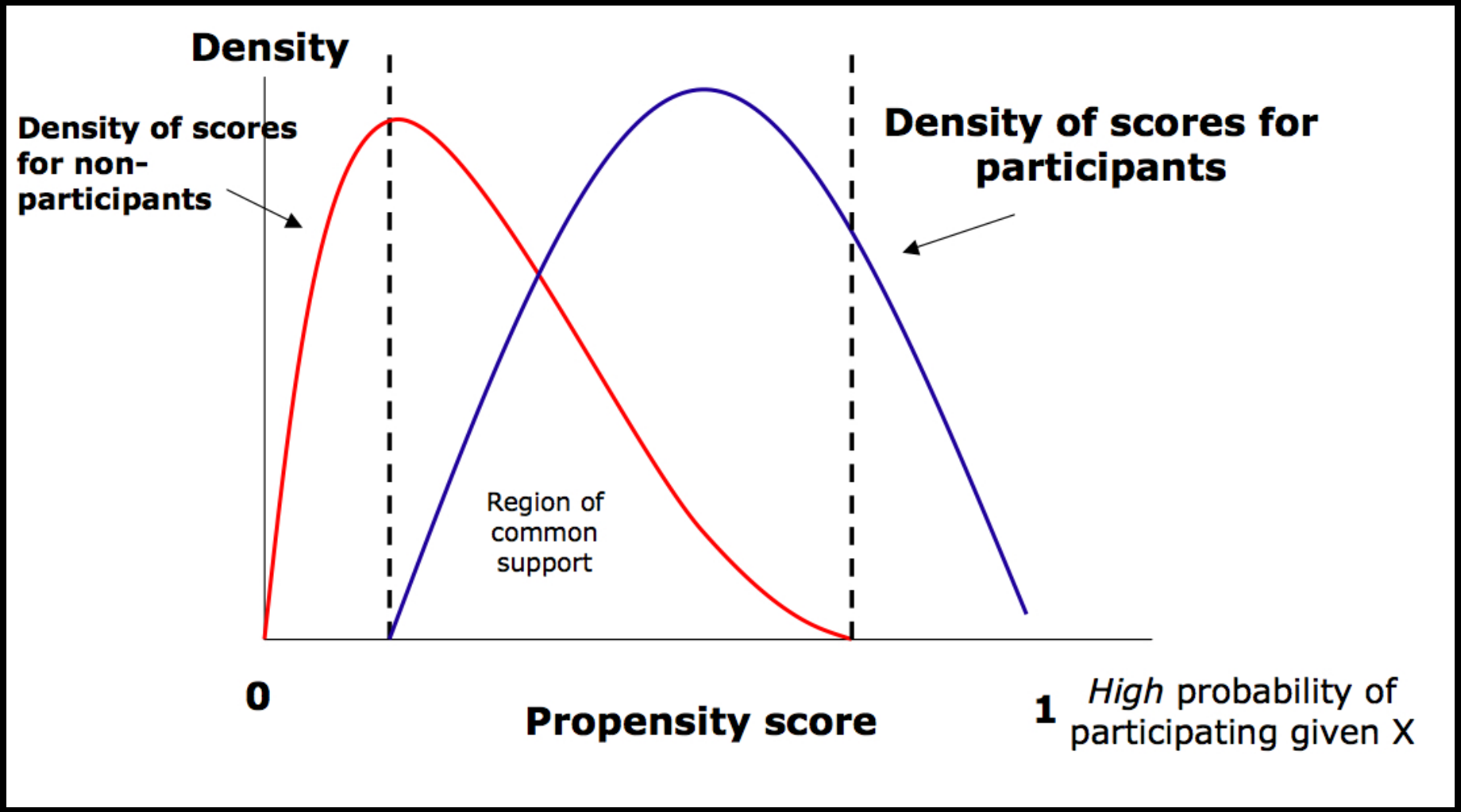
Z=treatment assignment
x=observed covariates

Requires Z to be independent of outcomes…

Works best when x is related to outcomes and selection (Z)

Methods Toolbox

Matching



Methods Toolbox

Matching

Often need large data sets for baseline and follow-up:

- Household data
- Environmental & ecological data (context)

Problems with ex-post matching: More bias with "covariates of convenience" (i.e. age, marital status, race, sex)

Prospective if possible!

Methods Toolbox

Instrumental Variables

Example: Want to learn the effect of access to health clinics on maternal mortality

Find a variable – "Instrument" – which affects access to health clinics but not maternal mortality directly.

e.g. changes in distance of home to nearest clinic.

Overview of Study Designs

Study Designs

Clusters

Phased Roll-Out

Not everyone has access to the intervention at the same time (supply variation)

Selective promotion

Variation in Treatments

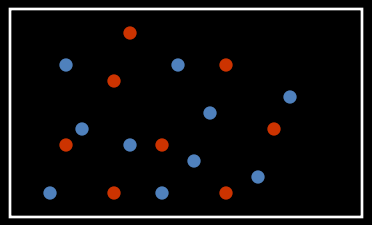
The program is available to everyone (universal access or already rolled out)

Study Designs

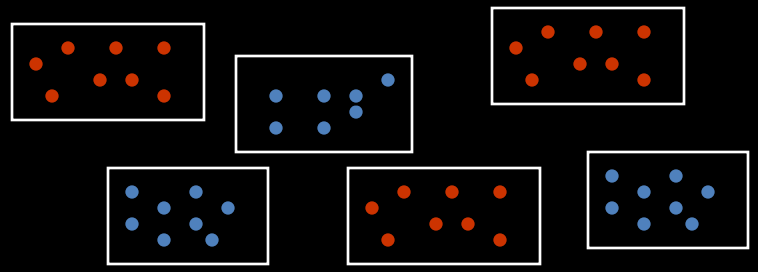
Cluster vs. Individual

- Cluster when spillovers or practical constraints prevent individual randomization.
- But… easier to get big enough sample if we randomize individuals

Individual randomization



Group randomization



Issues with Clusters

- Assignment at higher level sometimes necessary:
  - Political constraints on differential treatment within community
  - Practical constraints—confusing for one person to implement different versions
  - Spillover effects may require higher level randomization

- Many groups required because of within-community ("intraclass") correlation
- Group assignment may also require many observations (measurements) per group

Example: Deworming

- School-based deworming program in Kenya:
  - Treatment of individuals affects untreated (spillovers) by interrupting oral-fecal transmission
  - Cluster at school instead of individual
  - Outcome measure: school attendance
- Want to find cost-effectiveness of intervention even if not all children are treated
- Findings: 25% decline in absenteeism

Possible Units for Assignment

- Individual
- Household
- Clinic
- Hospital
- Village level
- Women's association
- Youth groups
- Schools

Unit of Intervention vs. Analysis

20 Intervention Villages

20 Comparison Villages

Unit of Analysis

Target Population

Random Sample

Evaluation Sample

Random Assignment

Control

Treatment

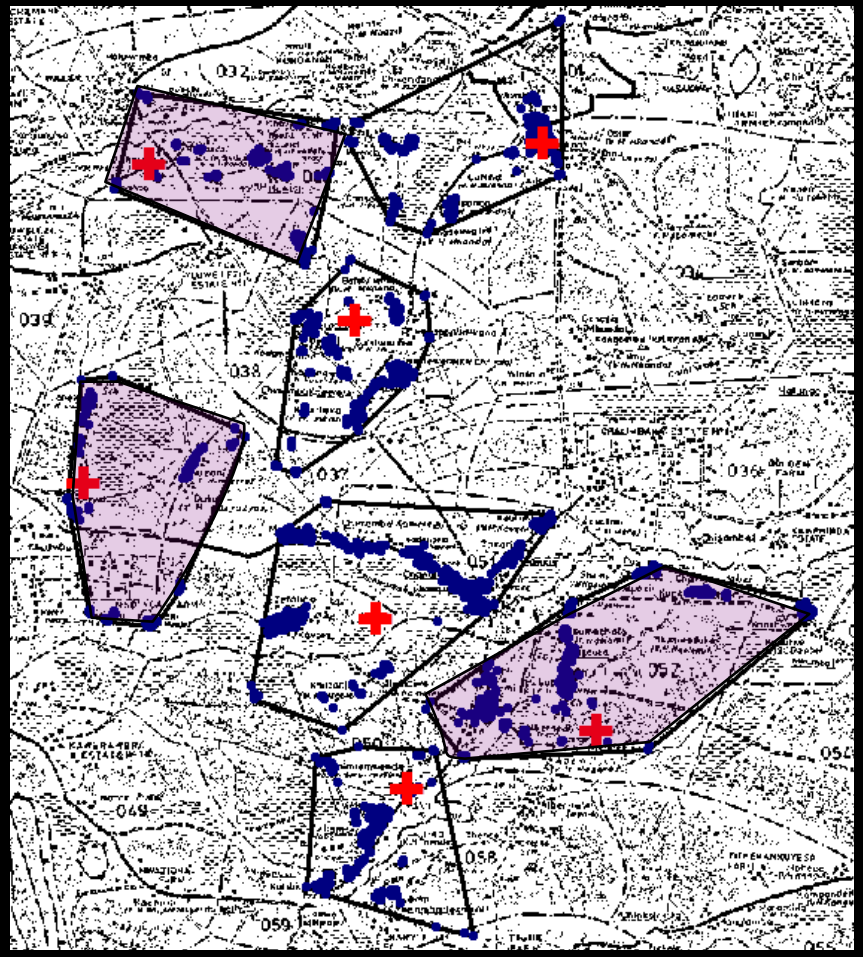
Random Sample again?

Unit of Intervention/ Randomization

Unit of Analysis

Study Designs

Phased Roll-out



Use when a simple lottery is impossible (because no one is excluded) but you have control over timing.

Phased Roll-Out

Clusters

	A
1	
2	
3	

Phased Roll-Out

Clusters	A	B
1		Program
2		
3		

Phased Roll-Out

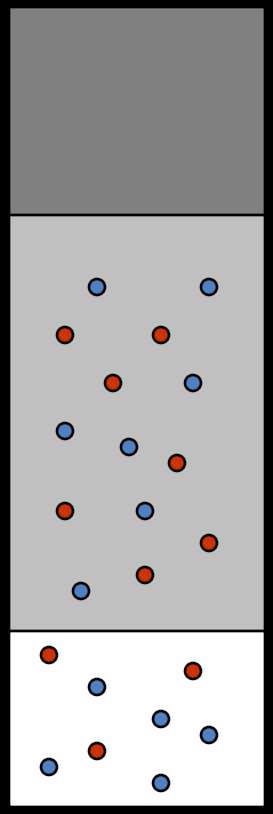
Clusters	A	B	C
1		Program	Program
2			Program
3			

Phased Roll-Out

Clusters	A	B	C	D
1		Program	Program	Program
2			Program	Program
3				Program

Study Designs

Randomized Encouragement



Already have the program

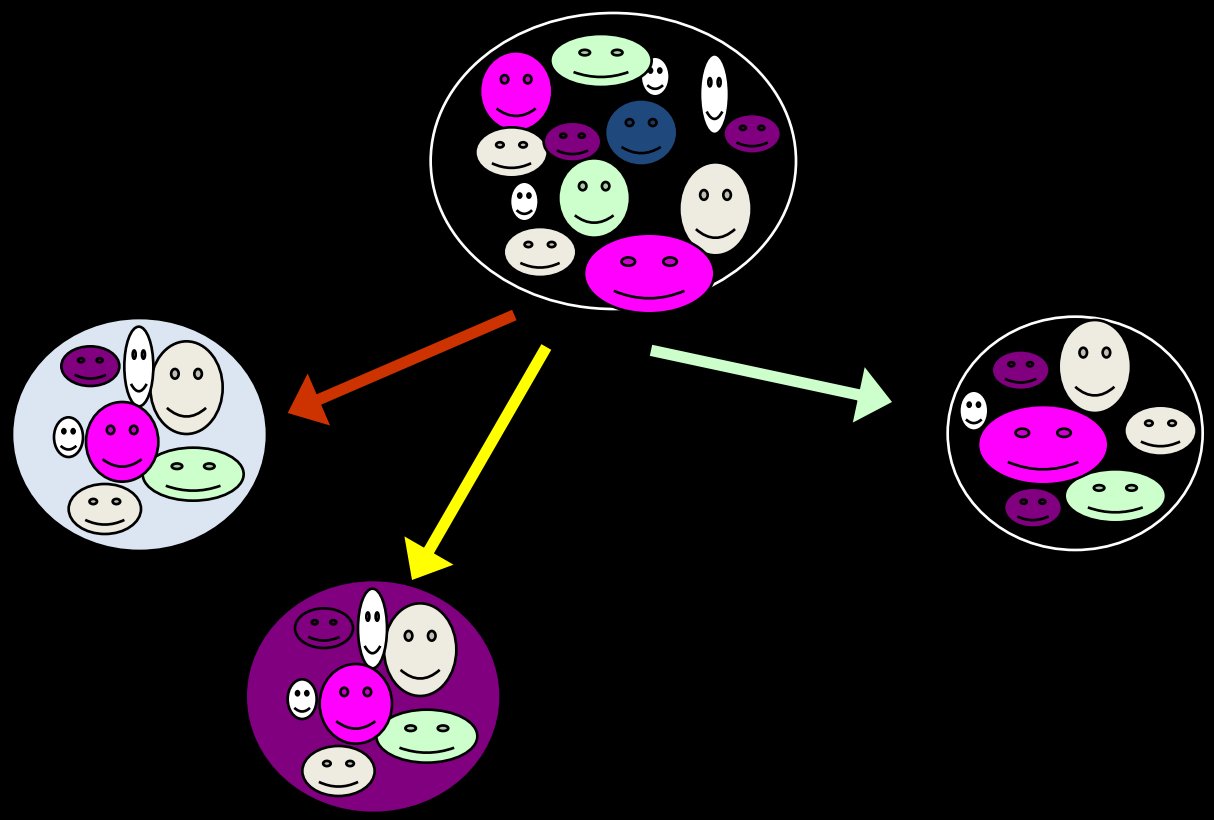
Likely to join the program if pushed

Unlikely to join the program regardless

Randomize who gets an extra push to participate in the program

Evaluation Toolbox

Variation in Treatment



Advantages of “experiments”

- Clear and precise causal impact
- Relative to other methods
  - Much easier to analyze
  - Can be cheaper (smaller sample sizes)
  - Easier to explain
  - More convincing to policymakers
  - Methodologically uncontroversial

When to think impact evaluation?

- EARLY! Plan evaluation into your program design and roll-out phases
- When you introduce a CHANGE into the program
  - New targeting strategy
  - Variation on a theme
  - Creating an eligibility criterion for the program

- Scale-up is a good time for impact evaluation!
  - Sample size is larger

When is prospective impact evaluation impossible?

- Treatment was selectively assigned and announced
  and no possibility for expansion
- The program is over (retrospective)
- Universal take up already
- Program is national and non excludable
  - Freedom of the press, exchange rate policy
  (sometimes some components can be randomized)
- Sample size is too small to make it worth it