Randomized Evaluations in Practice: Opportunities and Challenges

Kyle Murphy

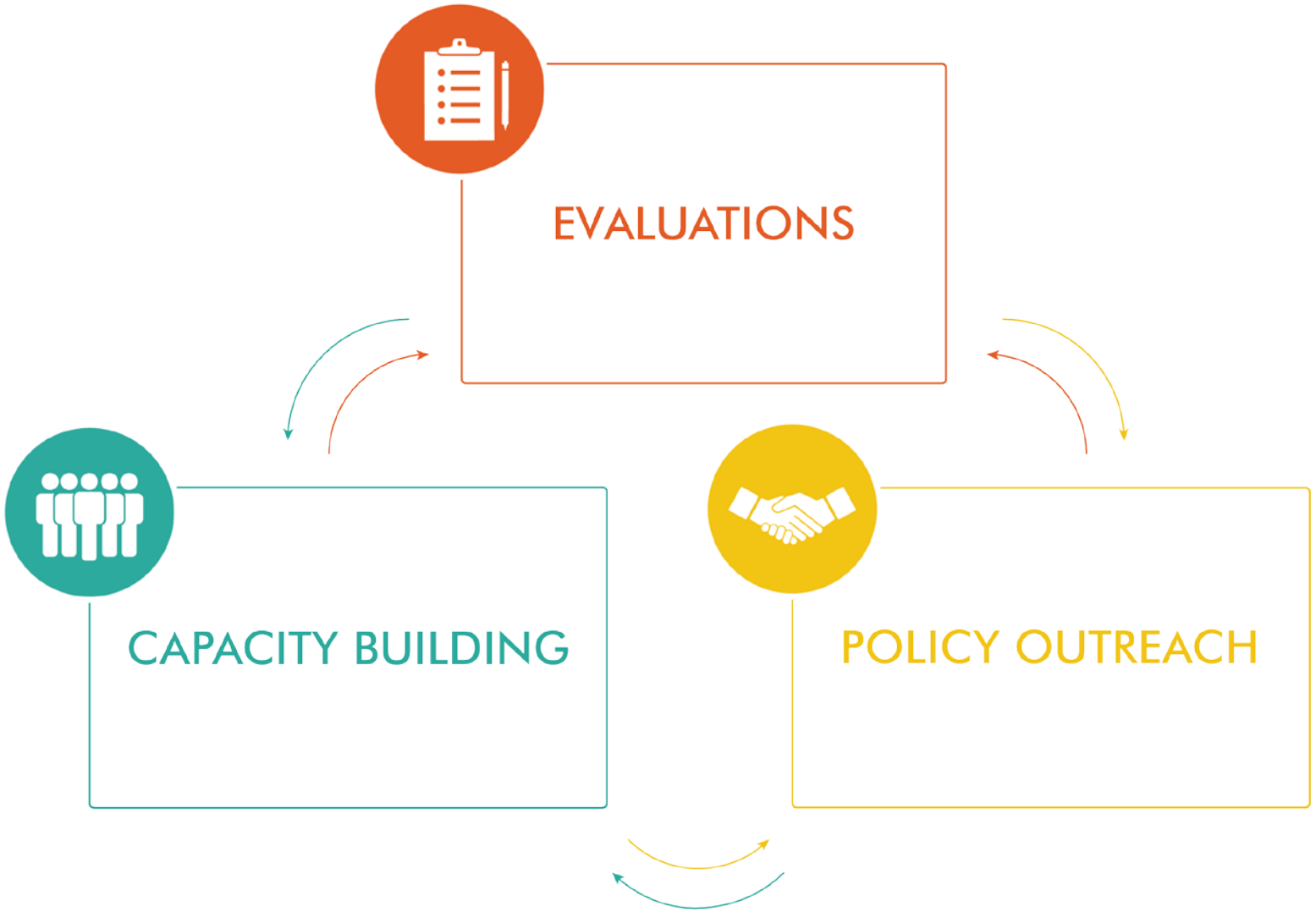
Policy Manager, J-PAL

January 30th, 2017

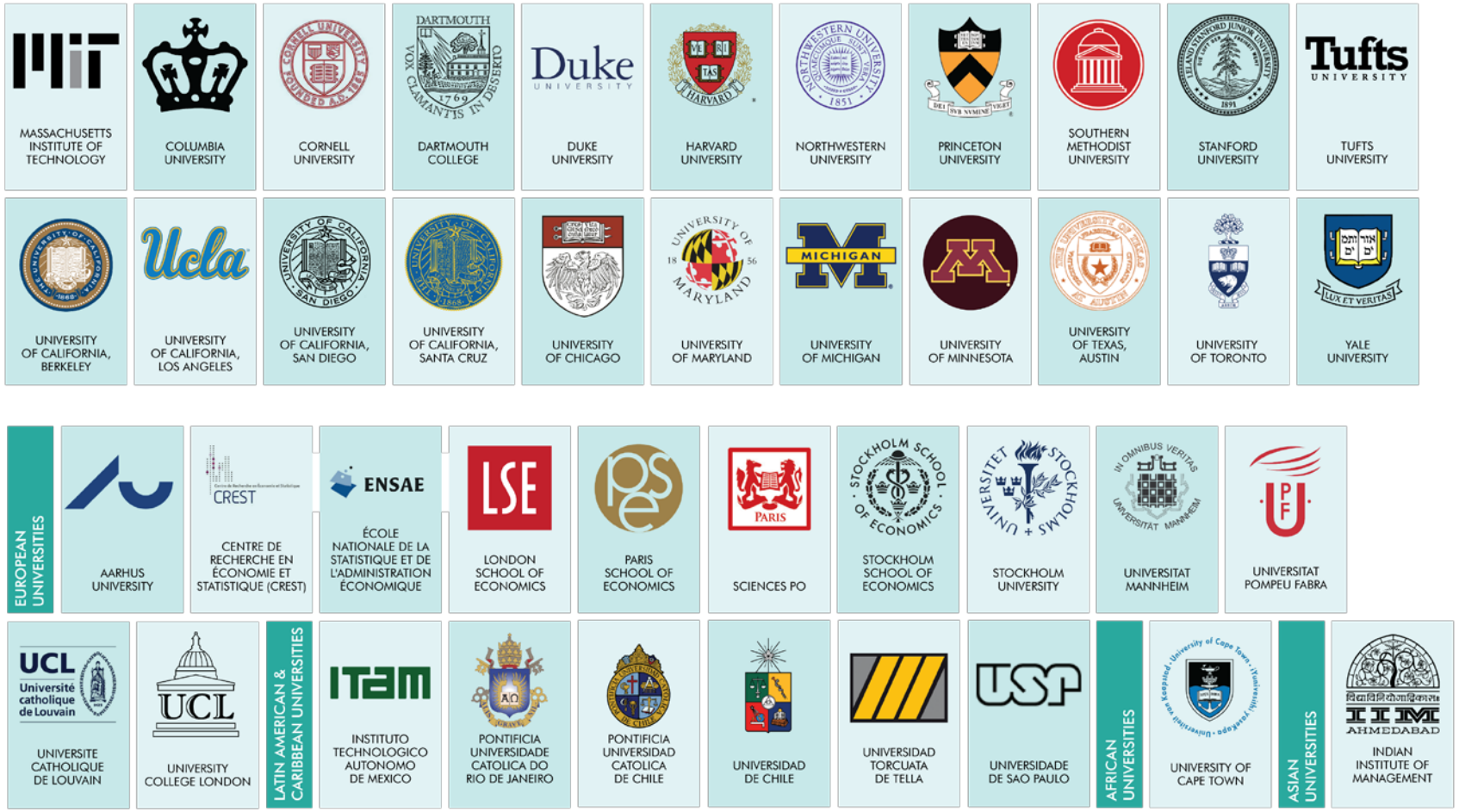
Overview

- Background
- What is a randomized evaluation?
- Why randomize?
- Advantages and drawbacks of randomized evaluations
- Conclusions

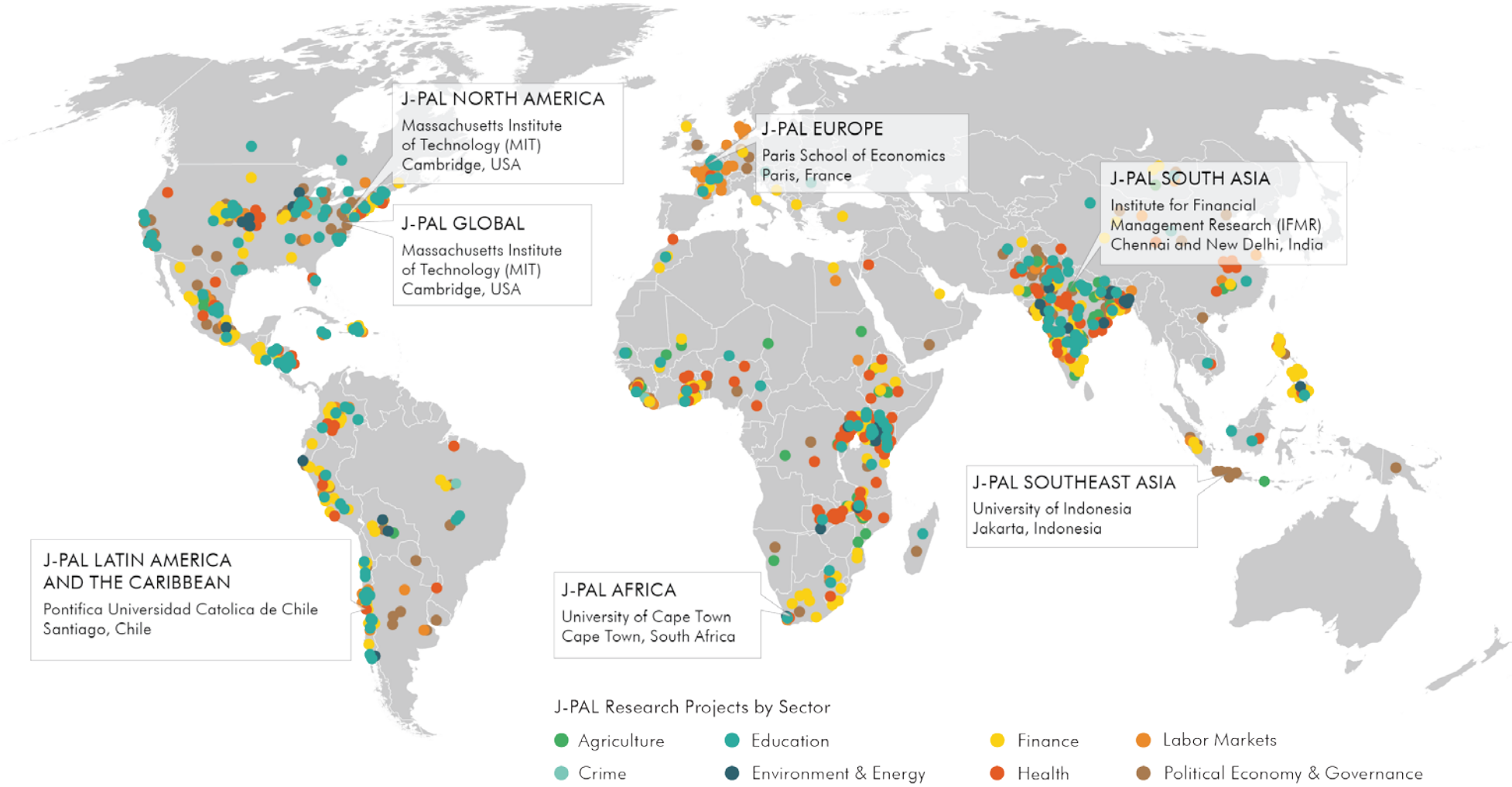
J-PAL's mission is to ensure that policy is informed by evidence and research is translated into action



J-PAL's network of 142 professors use randomized evaluations to inform policy



We have more than 770 ongoing and completed projects across 8 sectors in 69 countries



Our work focuses on 8 sectors

Agriculture

Crime

Education

Enivronment and Energy

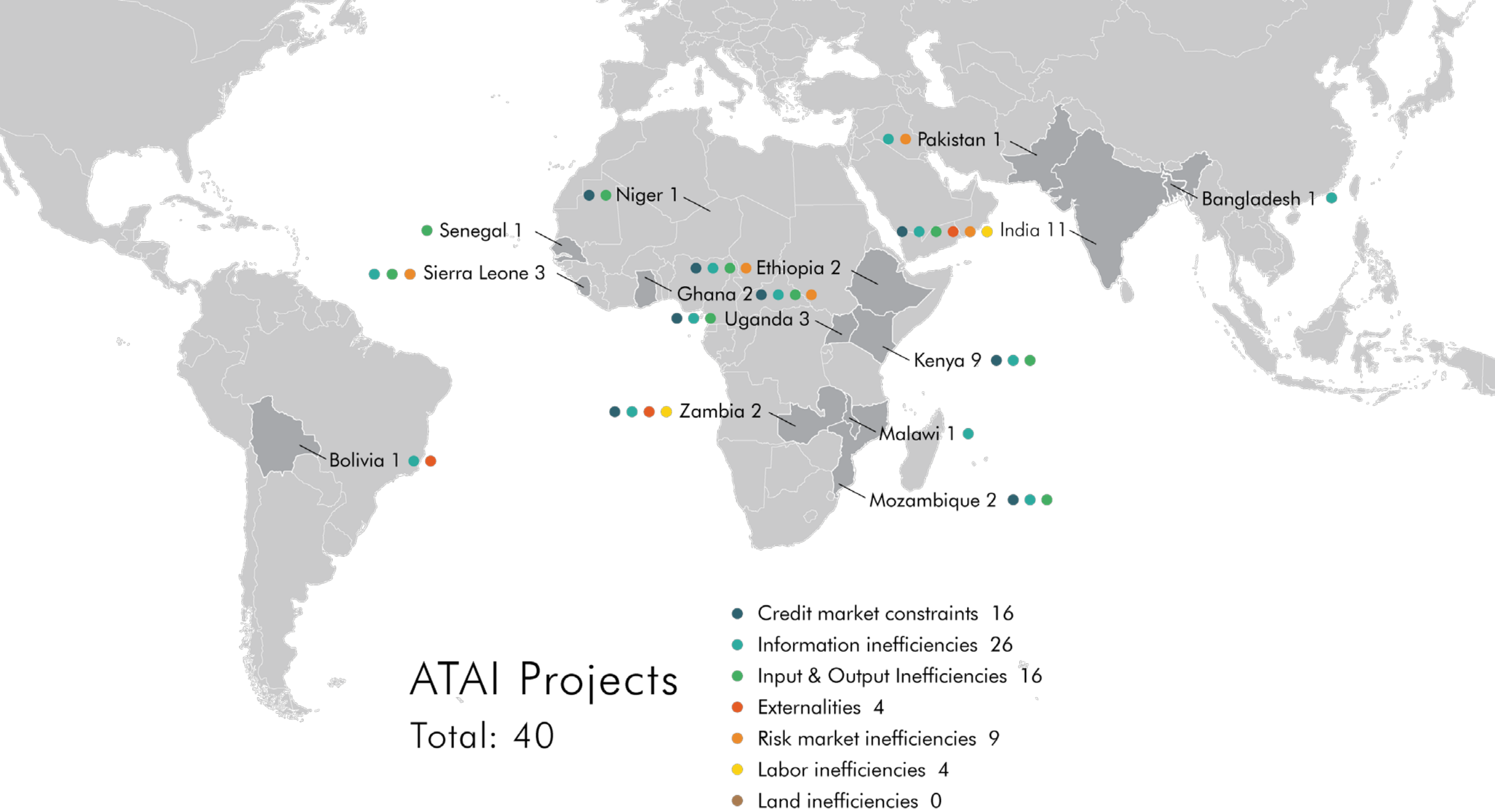
Finance and Microfinance

Health

Labor Markets

Political Economy and Governance

Agriculture project map



ATAI Projects

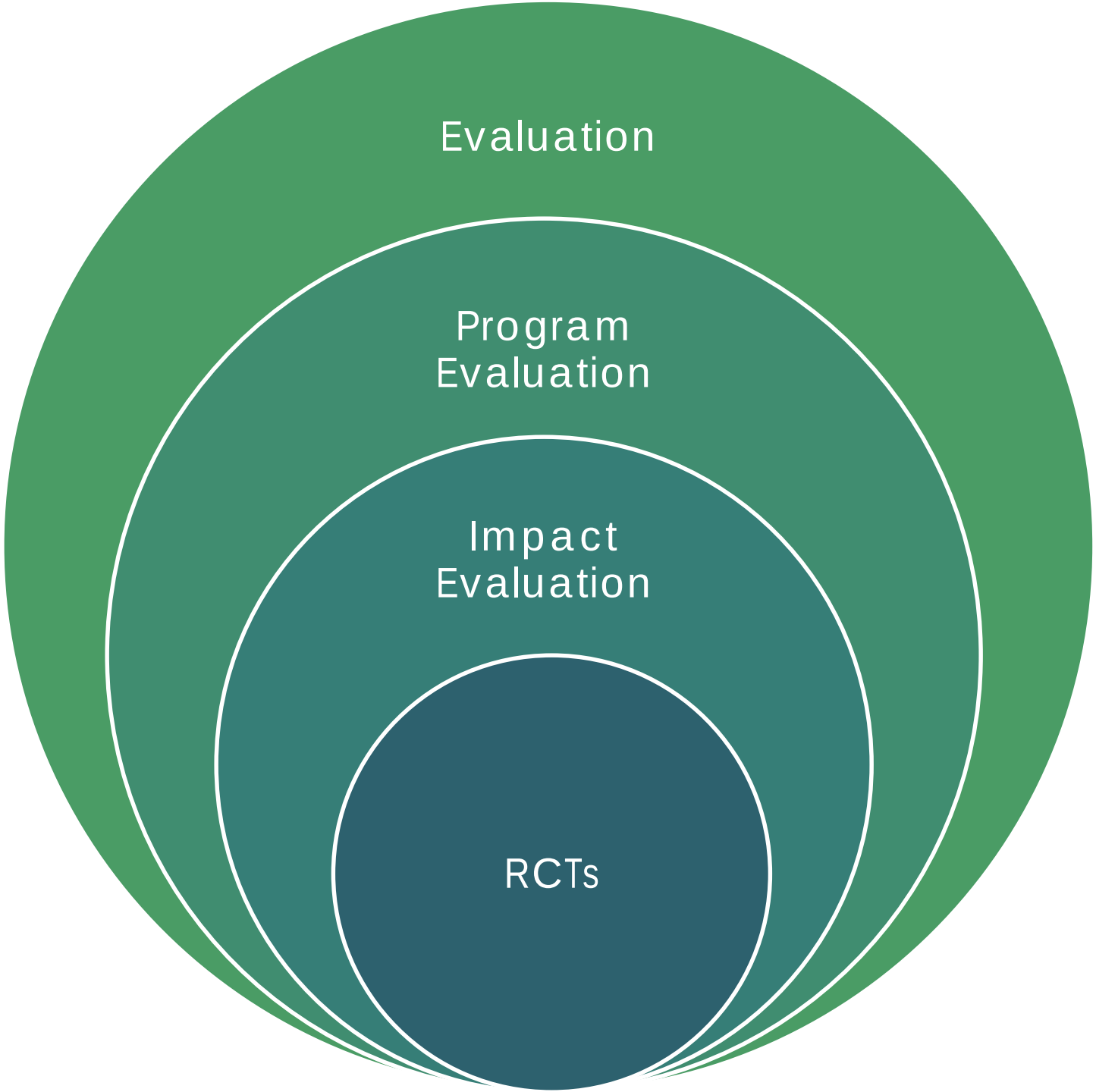
Total: 40

- Credit market constraints 16
- Information inefficiencies 26
- Input & Output Inefficiencies 16
- Externalities 4
- Risk market inefficiencies 9
- Labor inefficiencies 4
- Land inefficiencies 0

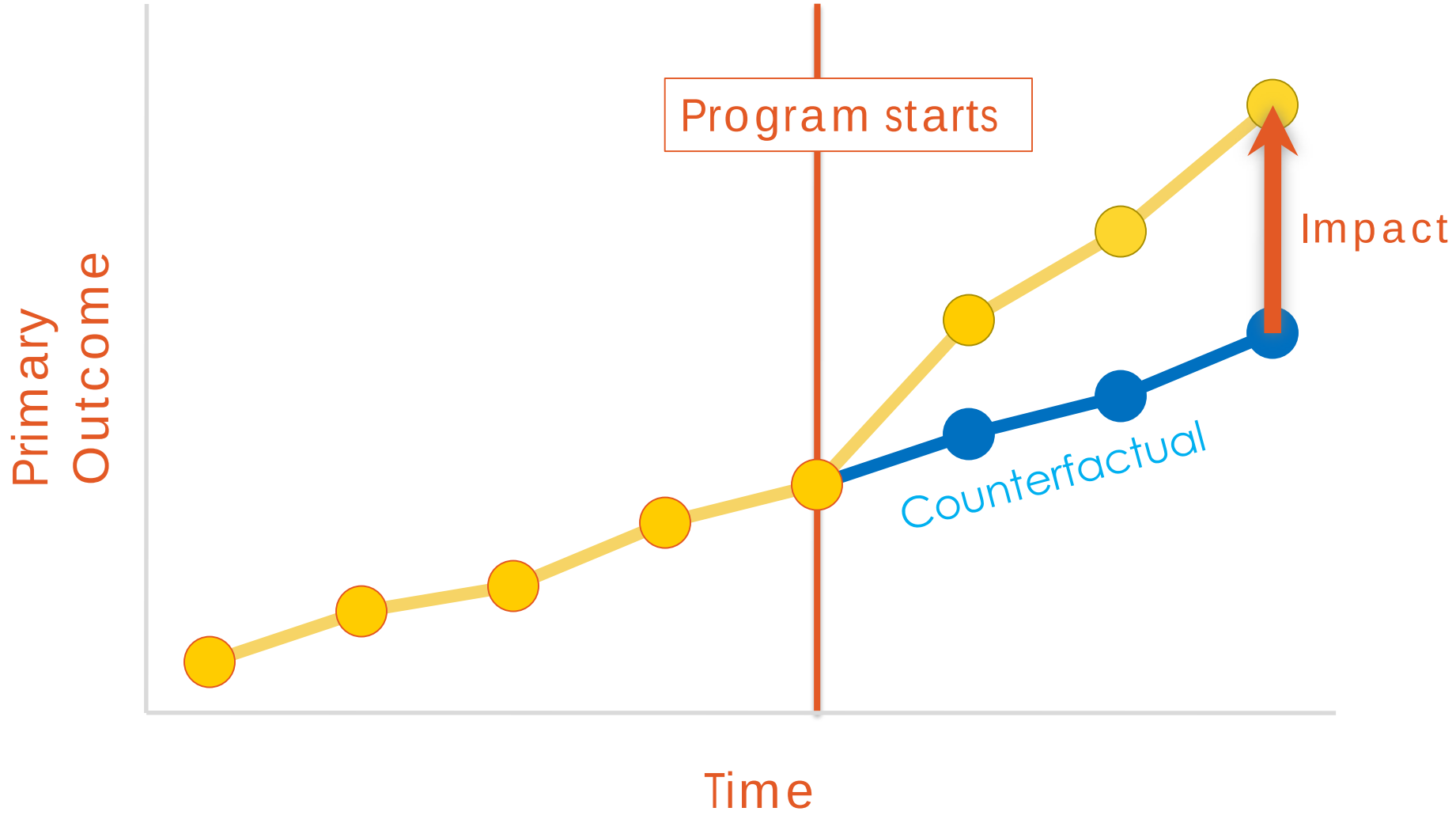
Since the start of ATAI

Category	Total
Farmers surveyed	111,351
Female farmers surveyed	47,845
Farmers whose behavior has changed	17,932
ATAI Awards	55
Unique ATAI projects	42
Countries with ATAI projects	14
Researchers on ATAI projects	89

Background


Evaluation
Program Evaluation
Impact Evaluation
RCTs

What is the impact of this program?

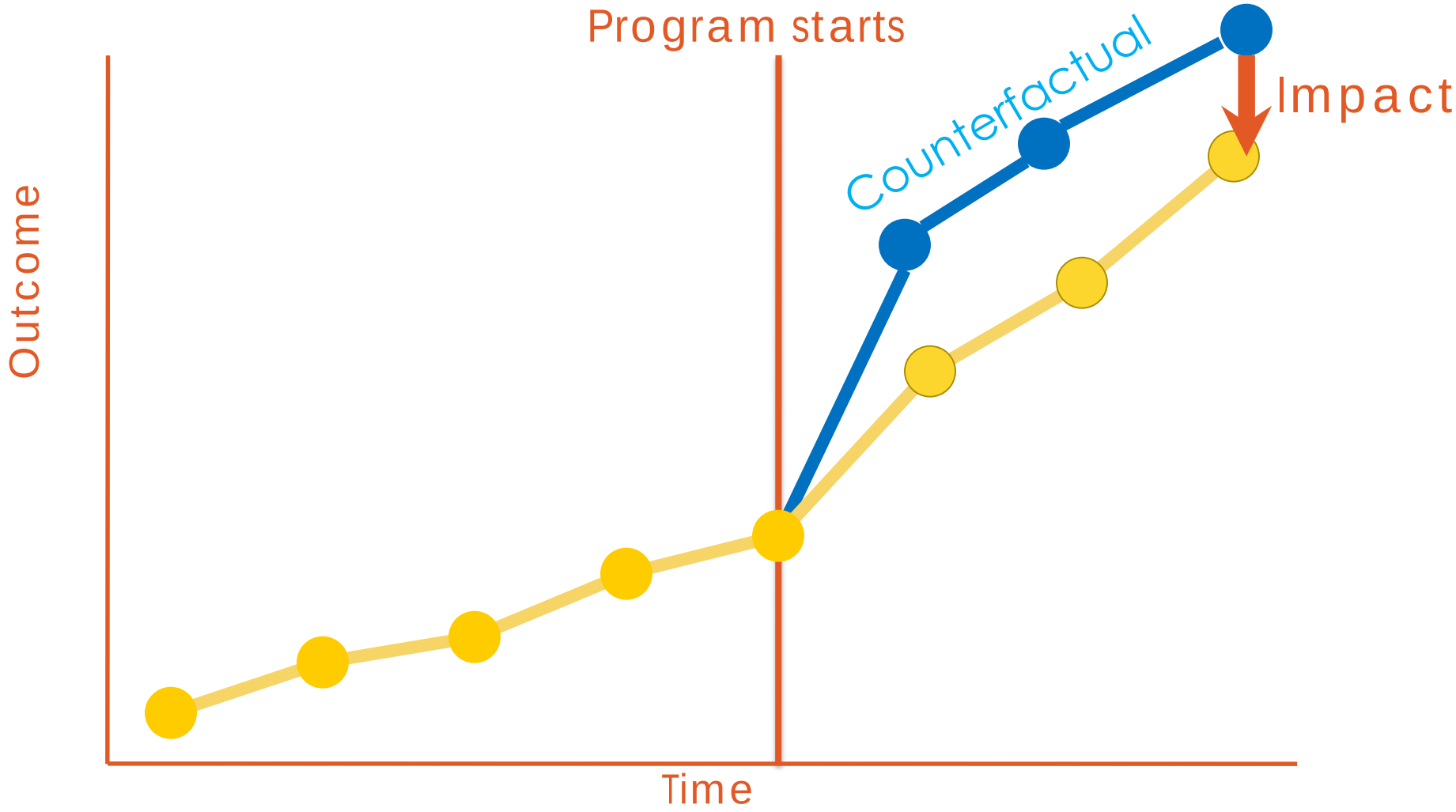


How to measure impact?

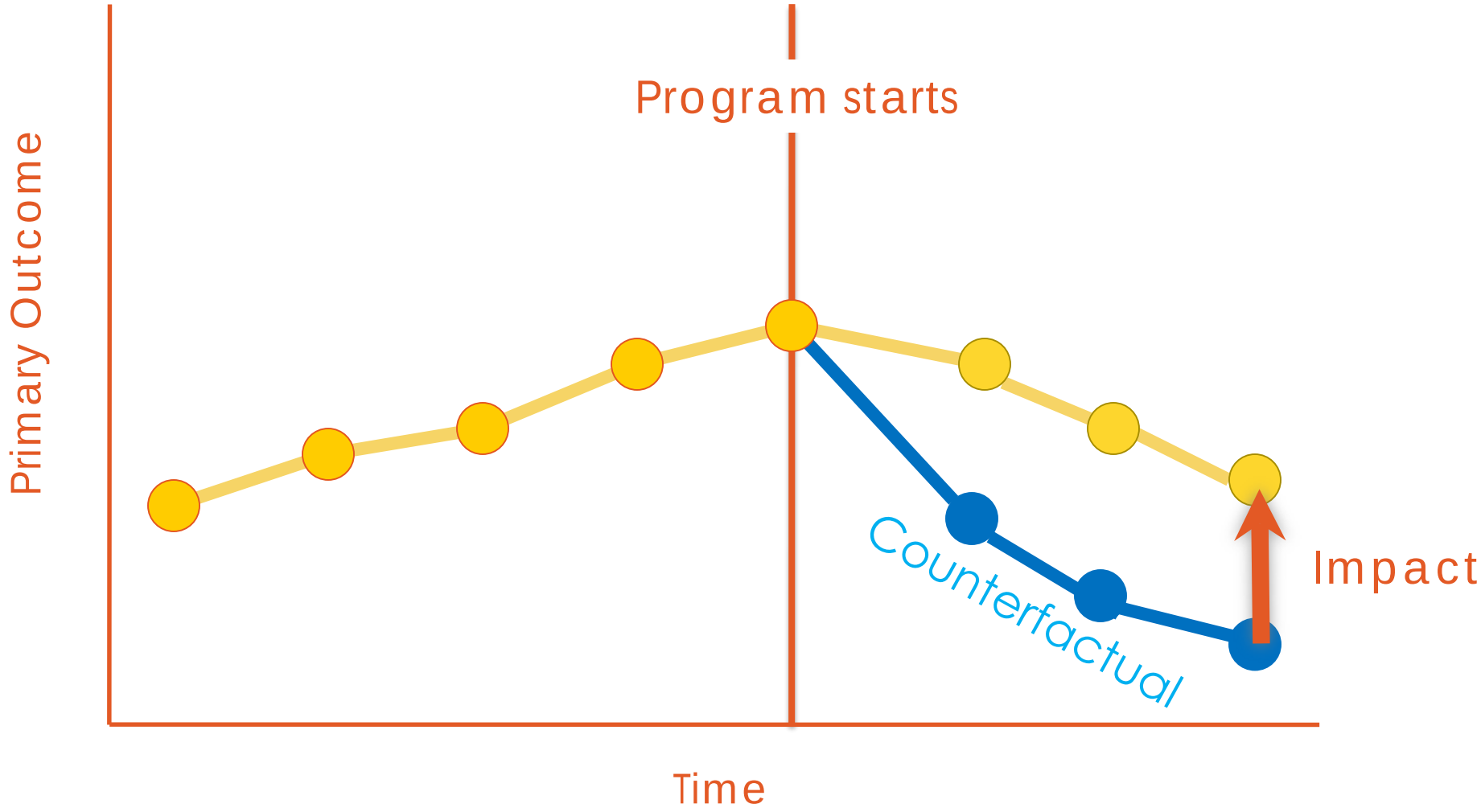
Impact is defined as a comparison between:

1. the outcome some time after the program has been introduced
2. the outcome at that same point in time had the program not been introduced (the *"counterfactual"*)

Impact: What is it?



Impact: What is it?



Counterfactual

The ***counterfactual*** represents the state of the world that program participants would have experienced in the absence of the program

Problem: Counterfactual cannot be observed

Solution: We need to "mimic" or construct the counterfactual

Constructing the counterfactual

- Usually done by selecting a group of individuals that ***did not*** participate in the program
- This group is usually referred to as the ***control group*** or ***comparison group***
- How this group is selected is a **key decision** in the design of any impact evaluation

Selecting the comparison group

- **Idea**: Select a group that is exactly like the group of participants in all ways except one: their exposure to the program being evaluated
- **Goal**: To be able to attribute differences in outcomes between the group of participants and the comparison group to the program (and not to other factors)

The problem of selection bias

- Individuals who participate in a program and those who do not are often different

- Comparing outcomes of these groups results in
  - **Impact of the program + pre-existing differences**

Impact evaluation methods

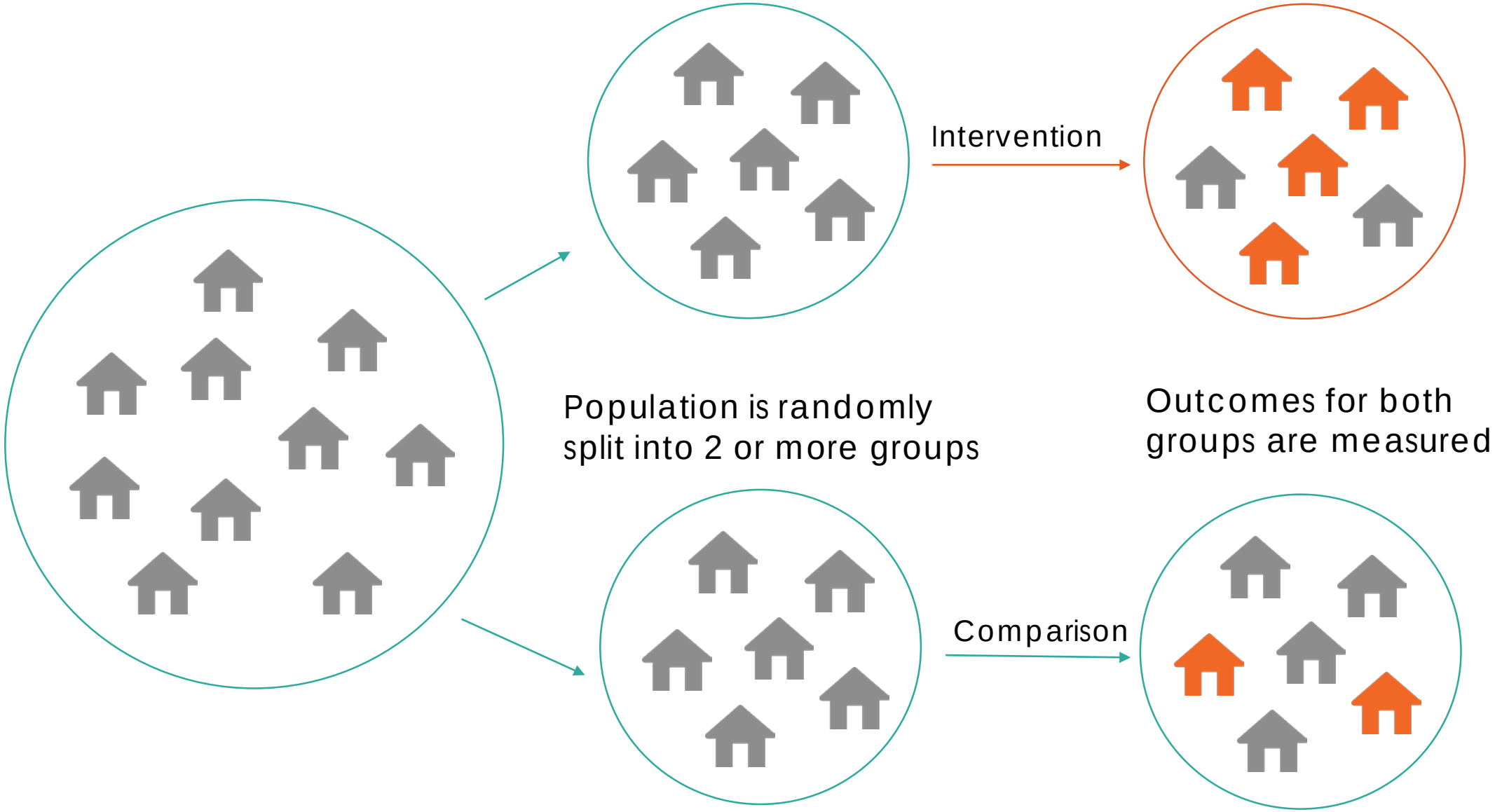
Randomized Experiments

- Random Assignment Studies
- Randomized Field Trials
- Social Experiments
- Randomized Controlled Trials (RCTs)
- Randomized Controlled Experiments

Non- or Quasi-Experimental

- Pre-Post
- Simple Difference
- Differences-in-Differences
- Multivariate Regression
- Statistical Matching
- Interrupted Time Series
- Instrumental Variables
- Regression Discontinuity

What is a randomized evaluation?


Intervention
Population is randomly split into 2 or more groups
Outcomes for both groups are measured
Comparison

Key steps in conducting a randomized evaluation

1. Design the study carefully
2. Collect baseline data
3. Randomly assign people to treatment or control
4. Verify that assignment is random
5. Monitor process so that integrity of experiments is not compromised

Key advantage of experiments

Because members of the groups (treatment and control) do not differ systematically at the outset of the experiment,

Any difference that subsequently arises between them can be attributed to the program rather than to other factors.

Fewer assumptions, clearly explainable results

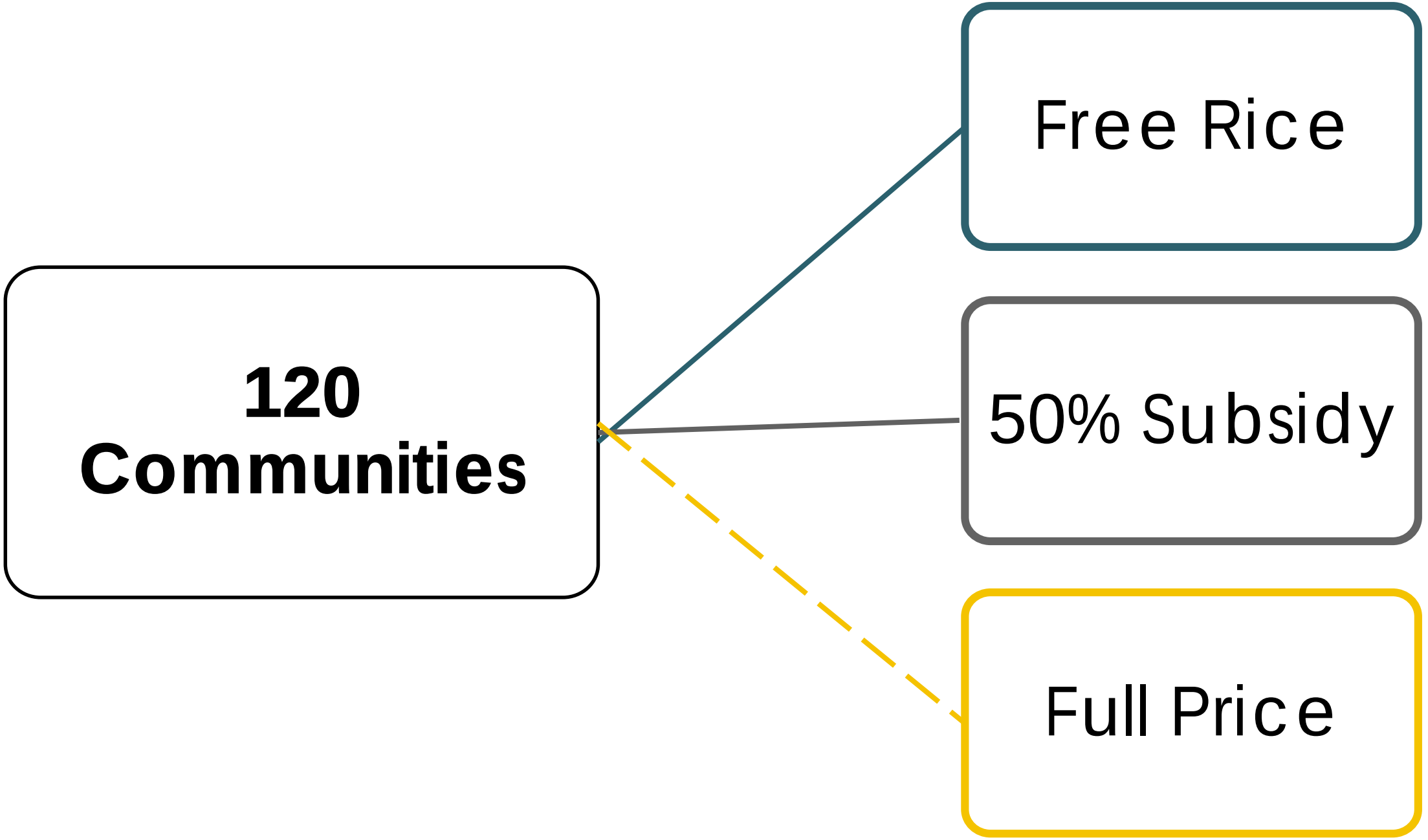
What can we learn from randomized evaluations?

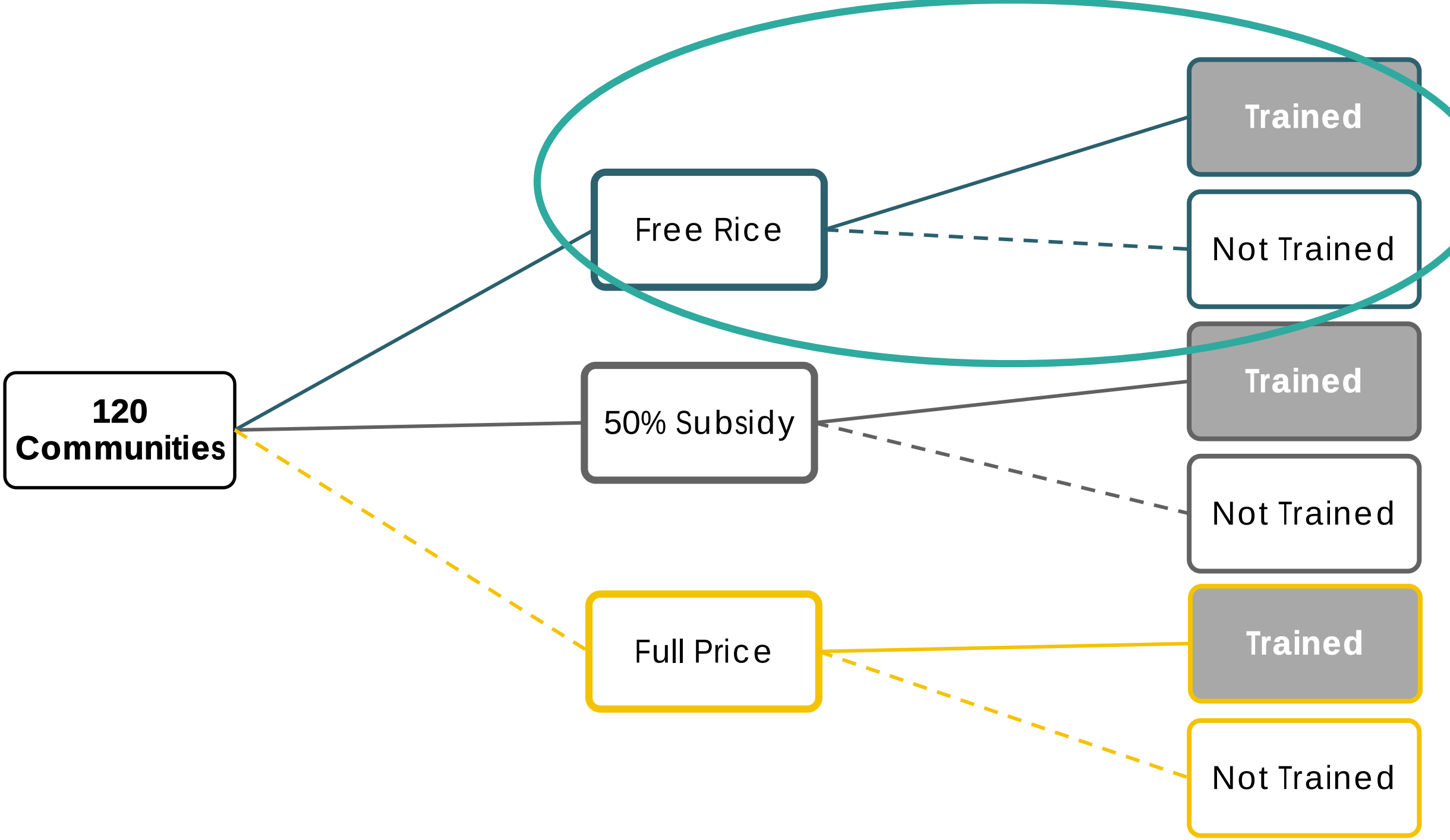
NERICA in Sierra Leone

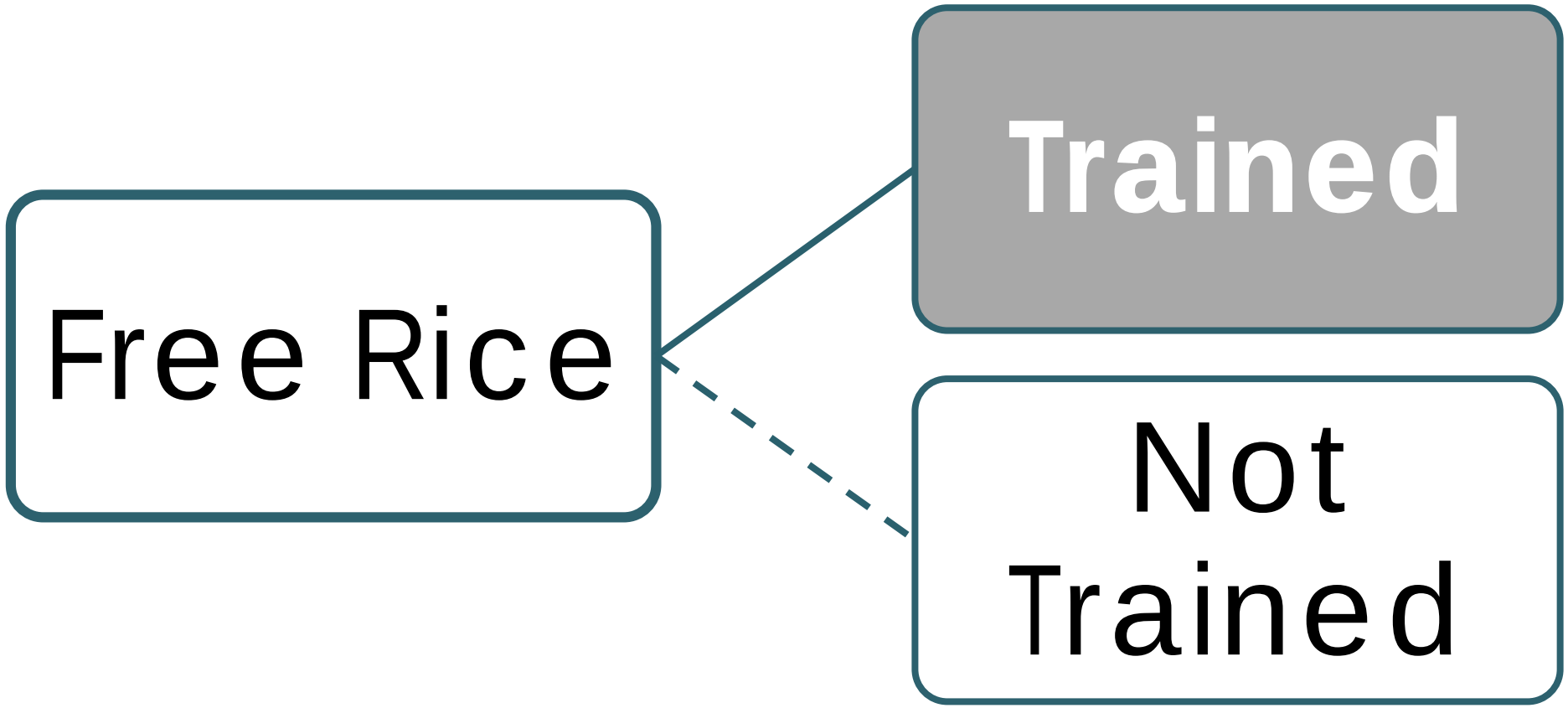
- Problem: Adoption of high-yielding crop varieties has been low
- Potential solution: Offer subsidies and trainings to increase take-up and yields

- What levels of subsidies are most effective? Does agronomic training help increase yields?

120 Communities


120 Communities
Free Rice
50% Subsidy
Full Price


120 Communities
Free Rice
50% Subsidy
Full Price
Trained
Not Trained
Trained
Not Trained
Trained
Not Trained


Free Rice
Trained
Not Trained

Yields only increased with training

- Yield increased 16% for trained farmers
- No increase without training

- Randomized design disentangled intervention components
- Revealed cost to ignoring extension

Benefits and Limitations of Randomized Evaluations

Benefits of Randomized Evaluations

- Tailoring the evaluation to the question
- Prospective evaluation
- Few assumptions, transparent findings

When to do a randomized evaluation?

- When there is an important question you want/need to answer
- When budgets are limited
- Timing--not too early and not too late
- Program is representative not gold plated
  - Or tests an basic concept you need tested
- Time, expertise, and money to do it right
- Develop an evaluation plan to prioritize

When NOT to do an RE

- When the program is premature and still requires considerable "tinkering" to work well
- When the project is on too small a scale to randomize into two "representative groups"
- If a positive impact has been proven using rigorous methodology and resources are sufficient to cover everyone
- After the program has already begun and you are not expanding elsewhere

Common Real World Constraints

Constraints: Political Advantages

- Not as severe as often claimed
- Lotteries are simple, common and transparent
- Randomly chosen from applicant pool
- Participants know the “winners” and “losers”
- Simple lottery is useful when there is no a priori reason to discriminate
- Perceived as fair
- Transparent

Constraints: Resources

- Most programs have limited resources
  - Vouchers, Farmer Training Programs
- Results in more eligible recipients than resources will allow services for
- Limited resources can be an evaluation opportunity

Constraints: contamination
Spillovers/Crossovers

- Remember the counterfactual!
- If control group is different from the counterfactual, our results can be biased
- Can occur due to
  - Spillovers
  - Crossovers

Constraints: logistics

- Need to recognize logistical constraints in research designs.
- E.g. individual de-worming treatment by health workers
  - Many responsibilities. Not just de-worming.
  - Serve members from both T/C groups
  - Different procedures for different groups?

Constraints: fairness, politics

- Randomizing at the child-level within classes
- Randomizing at the class-level within schools
- Randomizing at the community-level

Constraints: sample size

- The program is only large enough to serve a handful of communities
- Primarily an issue of statistical power

Conclusion

Impact evaluations are hard to do well

- Badly done impact evaluation can be very misleading
  - Can suggest that ineffective programs work
  - Create noise of competing claims that drown out good evidence
- Good impact evaluations require:
  - Good outcome measures
  - Enough sample size to measure impact precisely
- Do an impact evaluation when we can answer an important question well
  - Otherwise do a process evaluation and don't make impact claims
- Complement impact evaluations with other methods

Integrating RE into an Evaluation Strategy

- Good descriptive work important for diagnosing problem and selecting possible solutions
  - If children get one vaccine but don't complete course, probably not cultural barrier.
- Business case assessment
  - What would the impact need to be for this program to be cost-effective?
- Literature reviews tell you existing evidence, don't reinvent the wheel

Integrating RE into an Evaluation Strategy

- Process evaluation is always needed, can be dramatically improved
  - What % of eligible people are taking up the product?
  - Do people know more at the end of the training than at the beginning?
- If program shown to be effective in many contexts: time to scale
  - Scaling needs to be complemented with a good process evaluation
- Randomized evaluations can be most useful where there is an important question from both programming and academia

Thank you!

Kyle Murphy

kmurphy@povertyactionlab.org

What if you have 500 applicants for 500 slots?

- Consider non-standard lottery designs
- Could increase outreach activities
- Is this ethical?

Sometimes screening matters

- Suppose there are 2000 applicants
- Screening of applications produces 500 “worthy” candidates
- There are 500 slots
- A simple lottery will not work

- What are our options?

Consider the screening rules

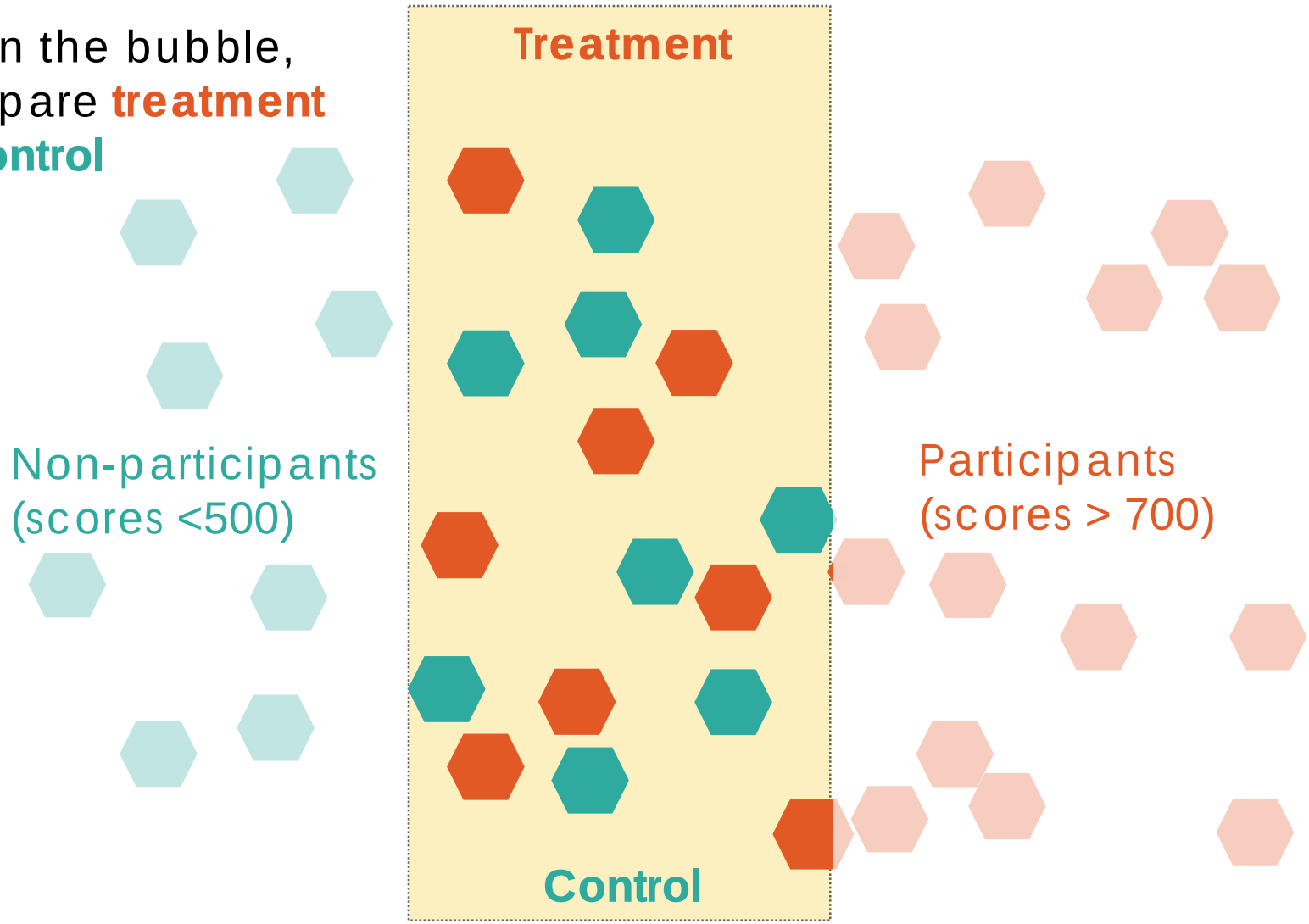
- What are they screening for?
- Which elements are essential?
- Selection procedures may exist only to reduce eligible candidates in order to meet a capacity constraint
- If certain filtering mechanisms appear "arbitrary" (although not random), randomization can serve the purpose of filtering *and* help us evaluate

Randomization in “the bubble”

- Sometimes a partner may not be willing to randomize among eligible people.
- Partner might be willing to randomize in “the bubble.”
- People “in the bubble” are people who are borderline in terms of eligibility
  - Just above the threshold → not eligible, but almost
- What treatment effect do we measure? What does it mean for external validity?

Randomization in "the bubble"

Within the bubble, compare **treatment** to **control**



When screening matters: Partial Lottery

- Program officers can maintain discretion
- Example: Training program
- Example: Expansion of consumer credit in South Africa

Phase-in: takes advantage of expansion

- Everyone gets program eventually
- Natural approach when expanding program faces resource constraints
- What determines which schools, branches, etc. will be covered in which year?

Phase-in design

Round 1
Treatment: 1/3
Control: 2/3

Round 2
Treatment: 2/3
Control: 1/3

Randomization evaluation ends

Round 3
Treatment: 3/3
Control: 0

Phase-in designs

Advantages

- Everyone gets something eventually
- Provides incentives to maintain contact

Concerns

- Can complicate estimating long-run effects
- Care required with phase-in windows
- Do expectations change actions today?

Rotation design

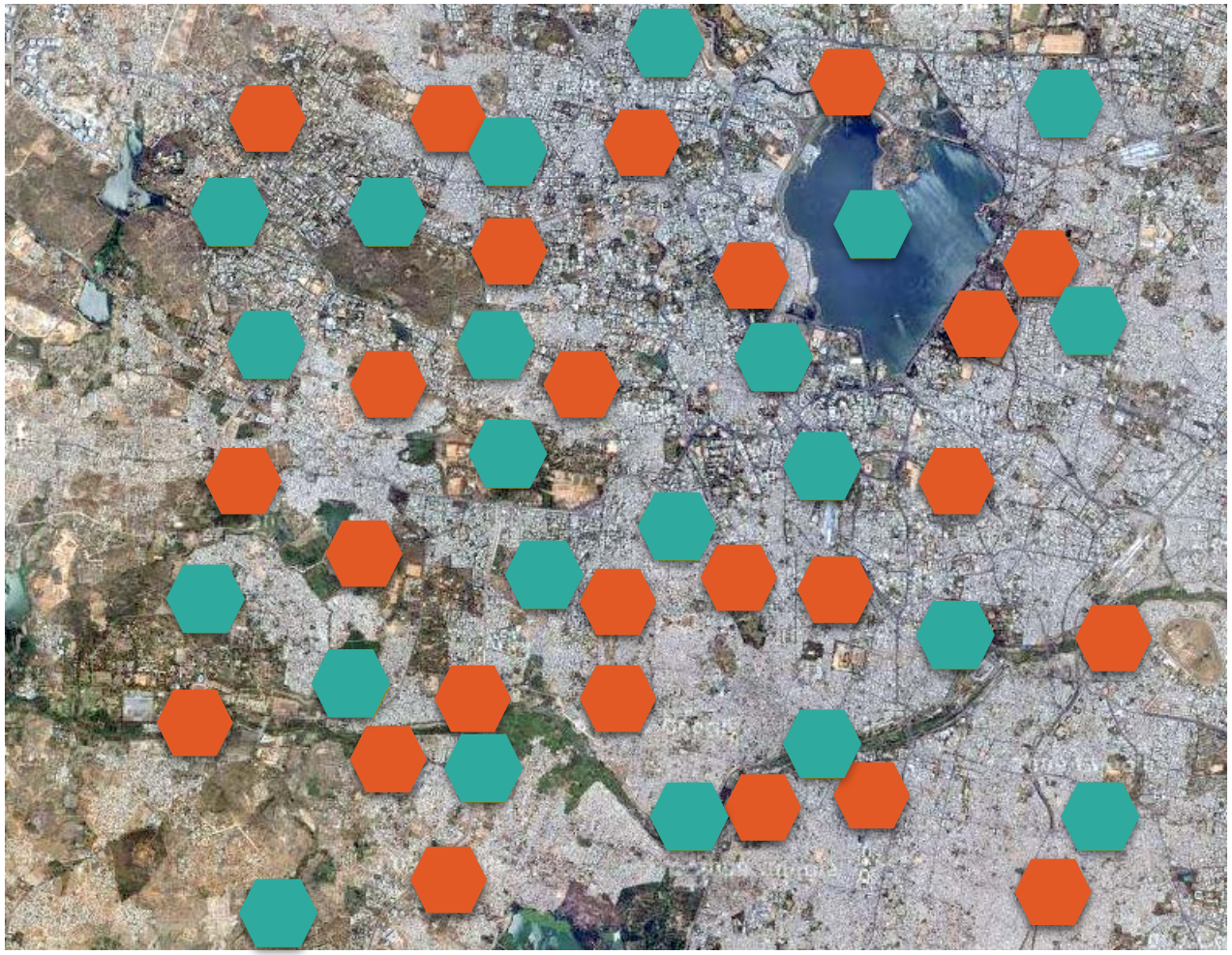
- Groups get treatment in turns
- Advantages?
- Concerns?

Rotation design

Round 1
Treatment: 1/2
Control: 1/2

Round 2
Treatment from Round 1 → Control

Control from Round 1 → Treatment



Encouragement design: What to do when you can't randomize access

- Sometimes it's practically or ethically impossible to randomize program access
- But most programs have less than 100% take-up
- Randomize encouragement to receive treatment

Encouragement design

Encourage

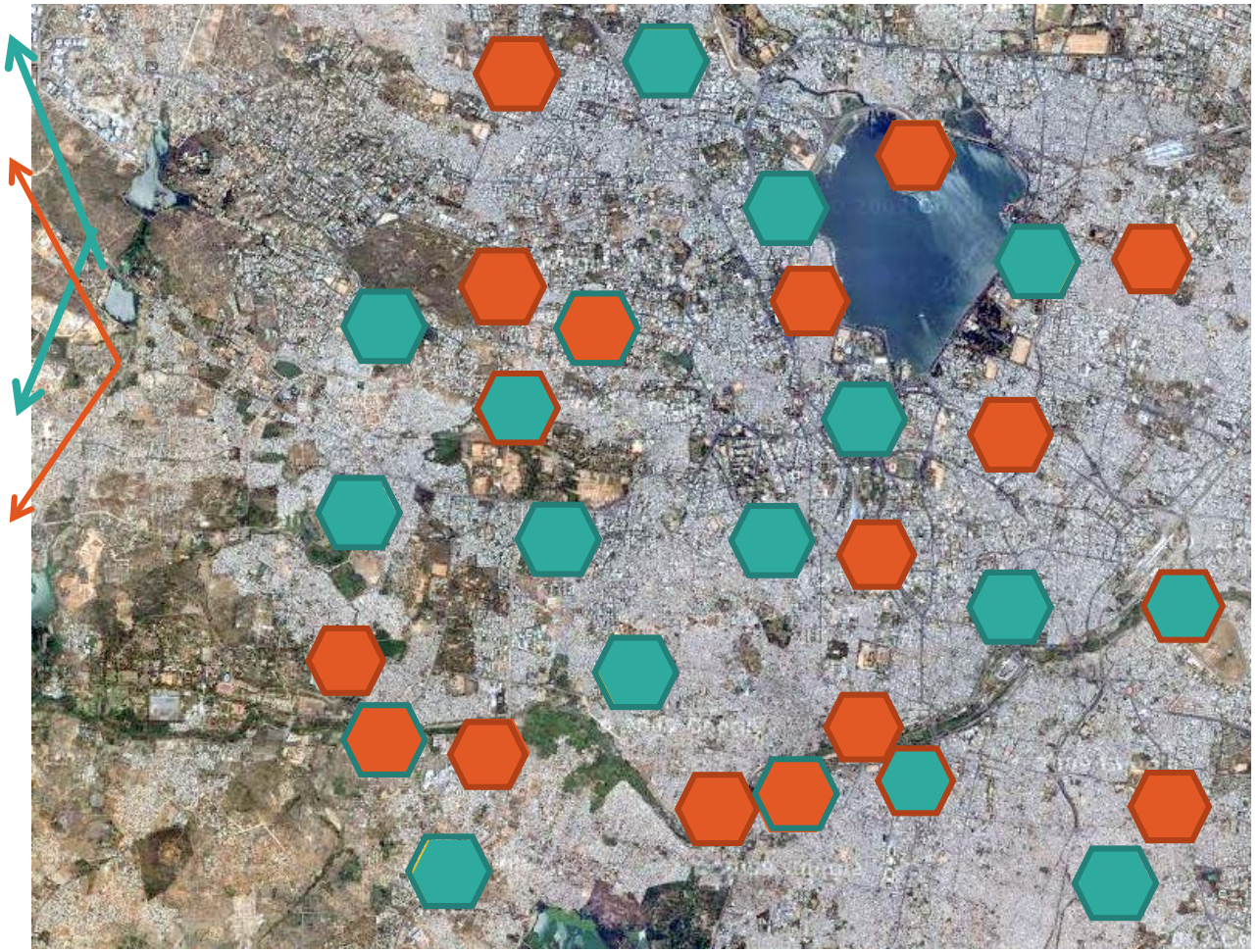
Do not encourage

Participated

Did not participate

Complying

Not complying



Encouragement design

Encourage

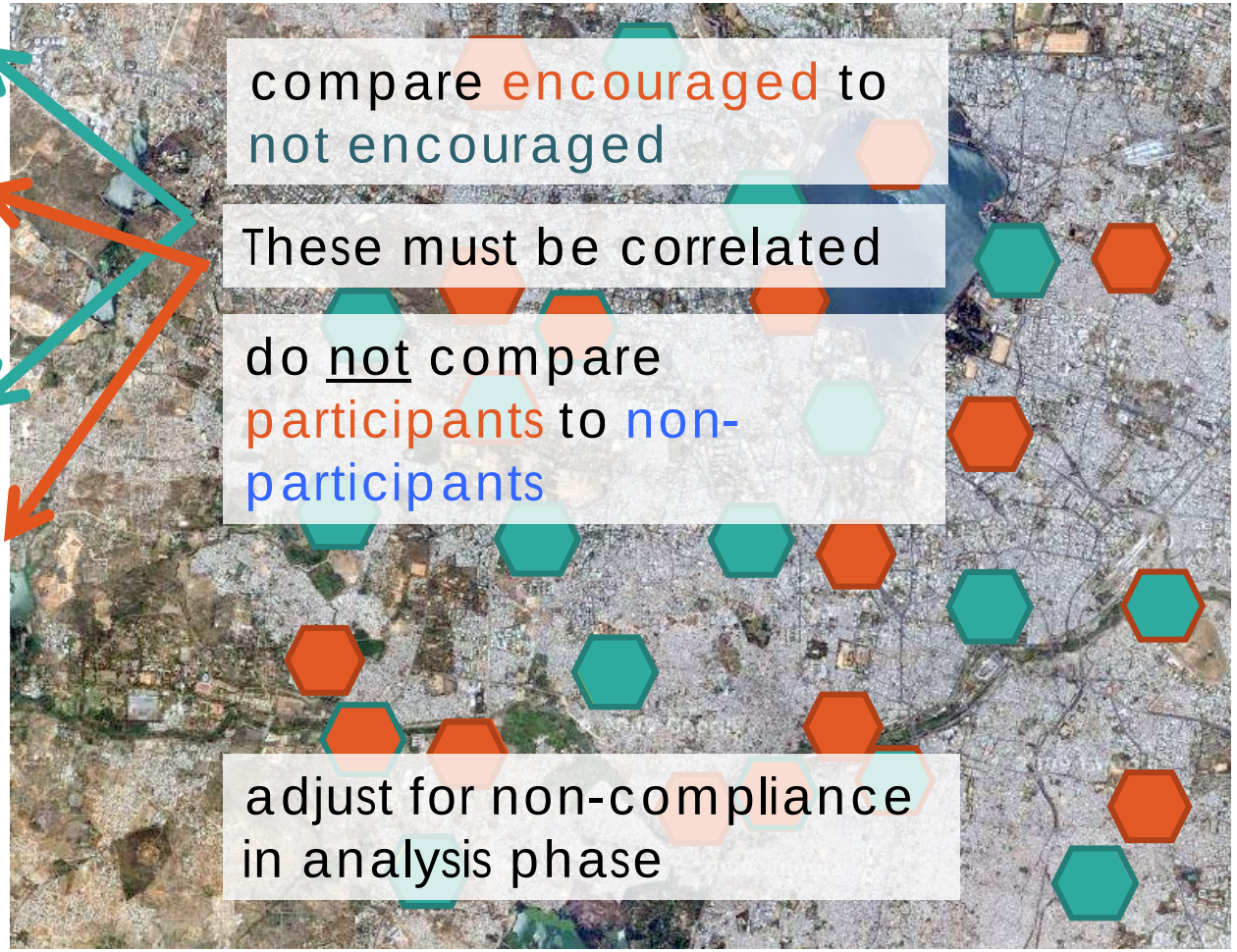
Do not encourage

Participated

Did not participate

Complying

Not complying



What is “encouragement”?

- Something that makes some folks more likely to use program than others
- Not itself a “treatment”
- For whom are we estimating the treatment effect?
- Think about who responds to encouragement

To summarize: Possible designs

- Simple lottery
- Randomization in the “bubble”
- Randomized phase-in
- Rotation
- Encouragement design

  – Note: These are not mutually exclusive.

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Basic Lottery	•Program oversubscribed	•Familiar •Easy to understand •Easy to implement •Can be implemented in public	•Control group may not cooperate •Differential attrition

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Phase-In	•Expanding over time •Everyone must receive treatment eventually	•Easy to understand •Constraint is easy to explain •Control group complies because they expect to benefit later	•Anticipation of treatment may impact short-run behavior •Difficult to measure long-term impact

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Rotation	•Everyone must receive something at some point •Not enough resources per given time period for all	•More data points than phase-in	•Difficult to measure long-term impact

Methods of randomization - recap

Design	Most useful when…	Advantages	Disadvantages
Encouragement	•Program has to be open to all comers •When take-up is low, but can be easily improved with an incentive	•Can randomize at individual level even when the program is not administered at that level	•Measures impact of those who respond to the incentive •Need large enough inducement to improve take-up •Encouragement itself may have direct effect