Impact Evaluation

Some Highlights from The Toolkit For The Evaluation of Financial Capability Programs in LMIC

World Bank Dissemination Workshop, New Delhi
March 2013

What is the Toolkit?

- Russia Trust Fund (RTF) aims to foster monitoring and evaluation (M&E) for financial capability programs in low and middle income countries (LMIC)

  - Develop a comprehensive definition of financial capability

  - Review and expand existing research base via a range of pilot projects

  - Build and make accessible corresponding methodological tools

Aim is to create a print/online resource for practitioners, policymakers and evaluators that provides practical and specific M&E guidance, drawing on RTF experiences from education to edutainment to product design

M&E In Practice

1: Identify and engage stakeholders — Concepts and programmatic approaches to financial capability

2: Understand the program — Setting the stage for M&E

3: Develop the M&E plan — Concepts and methodological approaches to M&E; planning and implementing M&E; practical and ethical considerations

4: Collect data and generate evidence — Methods of data collection and analysis

5: Put results to work — Reporting and dissemination

Monitoring — Are program operations on track?

Process Evaluation — How well did the program perform and why?

Impact Evaluation — What effects can be causally attributed to the program?

Cost-effectiveness Analysis — How much value does the program deliver and to whom?

Conceptualizing Impact

Program Logic

	Definition
Inputs	All resources used
Activities	Processes undertaken
Outputs	Deliverables
Outcomes / ST impacts	Immediate or intermediate effects
Final Outcomes / LT impacts	Long term effects

Program Logic

|	Definition	Community Bank Program
Inputs	All resources used	Finances, Labor, Equipment
Activities	Processes undertaken	Financial education training
Outputs	Deliverables	Workshops held, individuals trained
Outcomes / ST impacts	Immediate or intermediate effects	KAP changes
Final Outcomes / LT impacts	Long term effects	Behavioral changes Financial status changes

Program Logic

	Definition	Community Bank Program
Inputs	All resources used	Finances, Labor, Equipment
Activities	Processes undertaken	Financial education training
Outputs	Deliverables	Workshops held, individuals trained
Outcomes / ST impacts	Immediate or intermediate effects	KAP changes
Final Outcomes / LT impacts	Long term effects	Behavioral changes Fnancial status changes

Pre-post Comparison

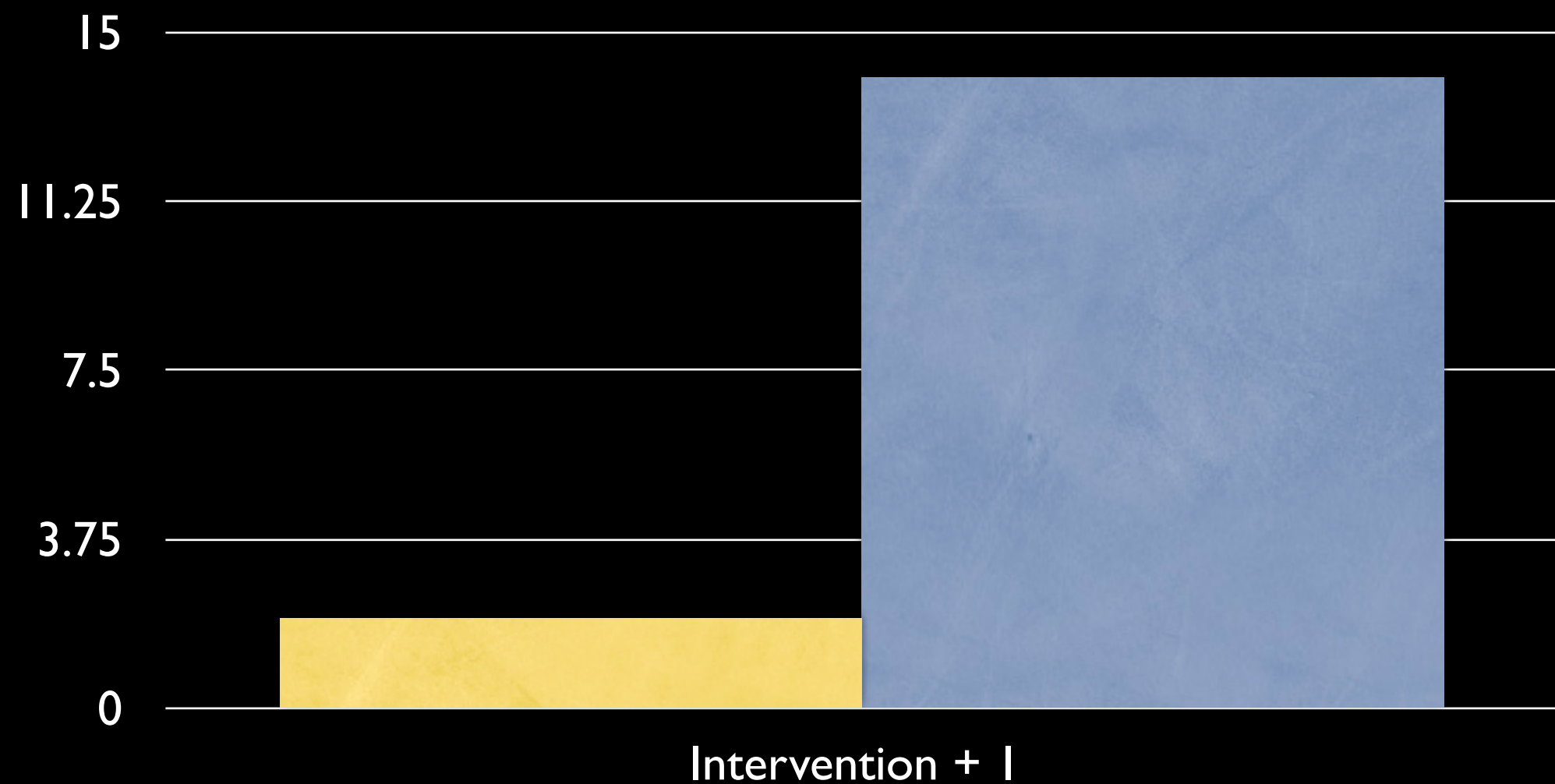- $ balance for customers who take up financial education

Pre-post Comparison

$ balance for customers who take up financial education

15

15

14

14

14

13

13

Intervention   Intervention +1
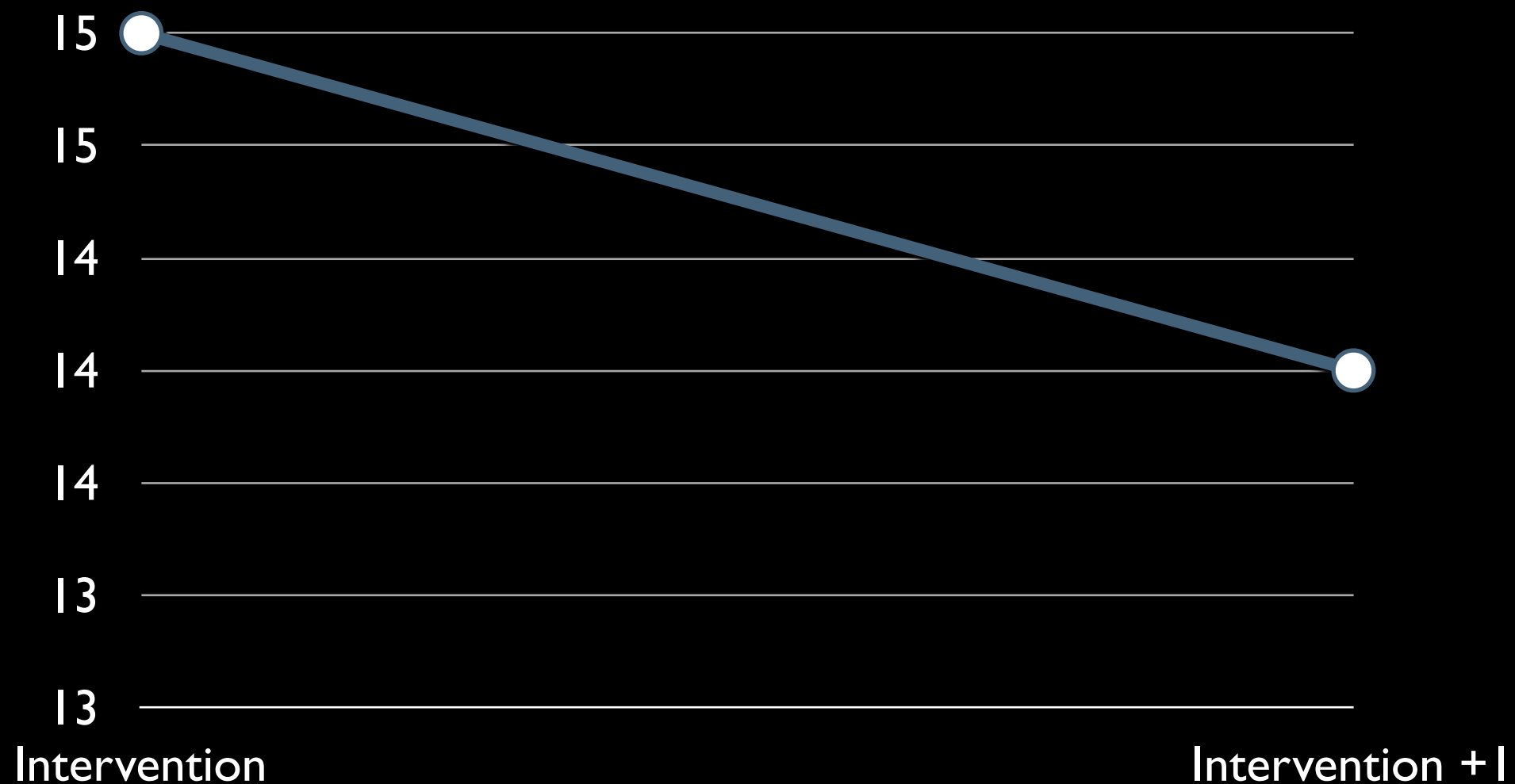
With/Without Comparison

$ balance for customers who did not take up financial education
$ balance for customers who take up financial education

Identifying Impact

- Ideally, we would observe participants both in reality and in a hypothetical world without the financial capability program (counterfactual)

- Causal effect = difference between observed outcomes and outcomes under the true counterfactual of no program

Identifying Impact

- Ideally, we would observe participants both in reality and in a hypothetical world without the financial capability program(counterfactual)

- Causal effect = difference between observed outcomes and outcomes under the true counterfactual of no program

- We can only infer causality by establishing as well as we can that an effect would not have happened without the program

  - Given our knowledge of the program and environment, identify a comparison group

  - Strengthen comparison using pre-post data on both groups

  - Provide qualitative support where possible

- The validity of an impact evaluation relies on how credible this is

Estimating Impacts

How do we know if the program actually had an effect and how large it was?

Finding Comparisons

Our *ideal* proxy for the counterfactual is a group is similar to the treatment group on all relevant traits, reacts in the same way to given treatments and experiences the same environment changes other than the program

Experimental control group: randomize program assignment (treatment) or the likelihood of assignment

Non-experimental / quasi-experimental: use statistical methods + assumptions to generate a comparison group

Randomized Controlled Trial (RCT)

- Assign program randomly to some eligible participants and not others (controls).

- Example: Community Bank randomizes the delivery of financial education to some customers and not others

- Estimate program effects by comparing outcomes between treatment and control groups
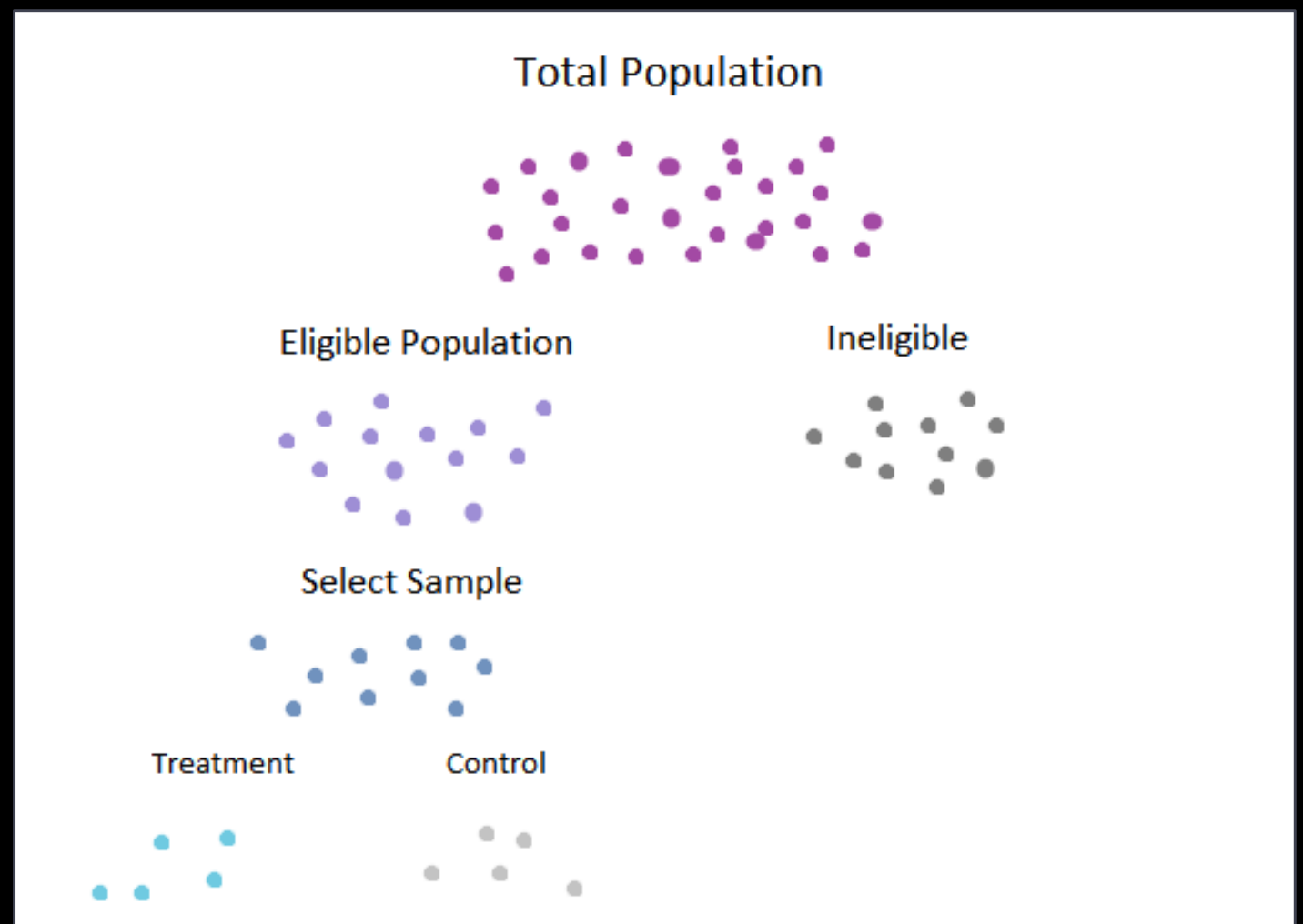
When Is This Suitable?

- Potential to randomize over a large sample at a low level of implementation

- Long-term commitment and ability to maintain control over delivery

- Conditions that naturally facilitate RCTs such as piloting, expansion/phase-in and program oversubscription

- Acceptable mechanisms for allocation, communication and incentives such as public lotteries and future benefits

- May not be feasible if program is very small or if randomization can only be done at a high level for a small number of units

- May fail if logistics are are weak: noncompliance, attrition or contamination

- Randomization of some interventions may be politically or ethically sensitive, or may not fit with program cycle

- Programs with lack of buy-in from stakeholders (including participants) can lead to destabilization and reactivity

Encouragement Design

- Promote program randomly to some eligible participants (treatment) and not others (controls)

- Example: Incentives are randomly provided to one group of potential Community Bank customers

- Estimate program effects by comparing outcomes between treatment and control groups, adjusting for differences in participation rates
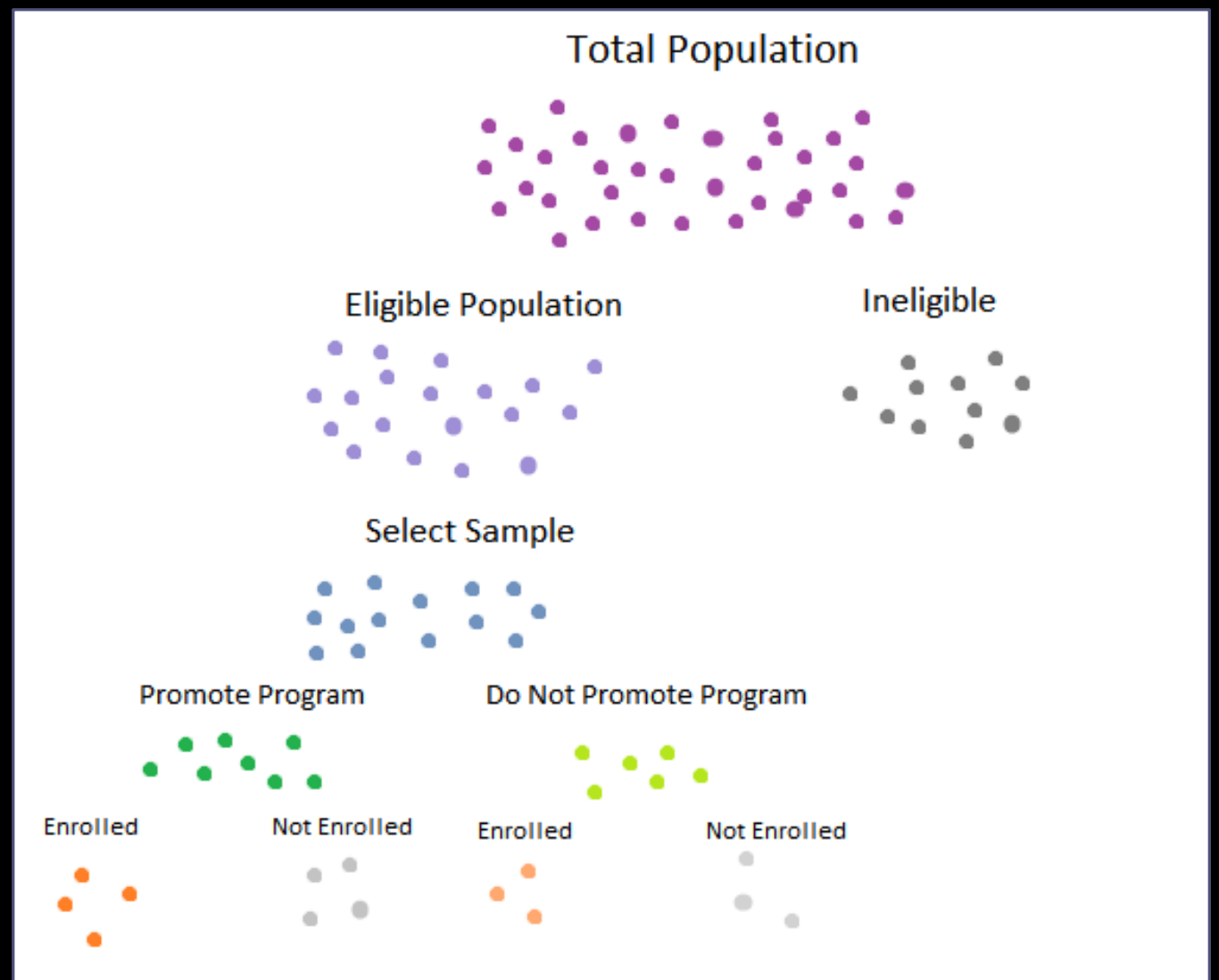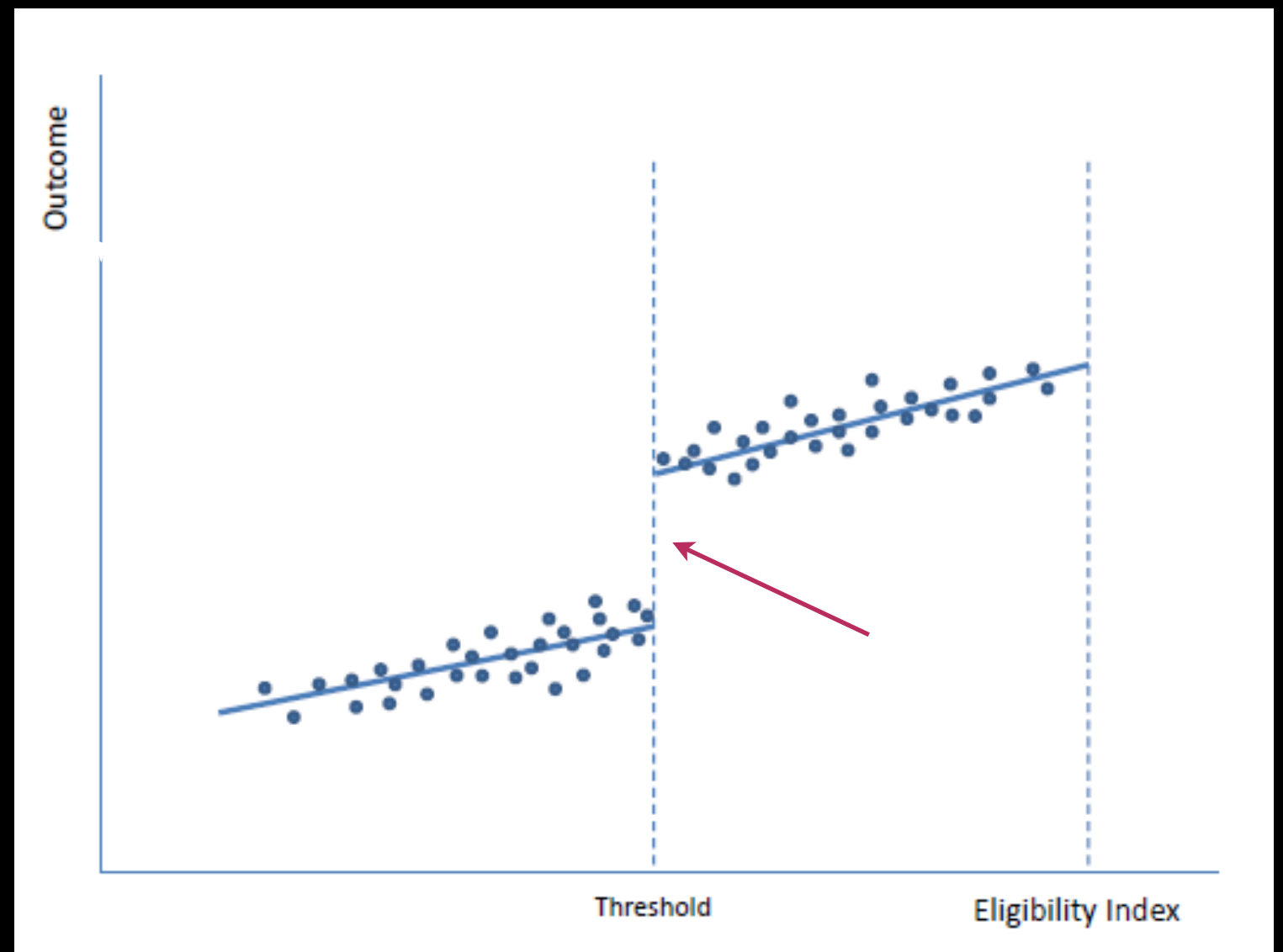
When Is This Suitable?

- Programs where access to services cannot be restricted (e.g. mass media) but promotion can be effectively randomized

- Programs with a very large sample size

- Strong incentives are available that are not correlated with the program outcomes

- May not be feasible if promotion itself is disallowed or if the logistics/administration of promotion is prohibitive

- Relative to random program assignment, larger sample is needed to account for more nonparticipants

- May fail if the offered incentives do not increase participation OR if the incentives lead to confounding results

Discontinuity Design

- Assign the program prospectively based on an eligibility index or review program rules retrospectively

- Example : Suppose in the past, Community Bank provided financial education only to priority clients i.e. individuals above a certain income level.

- Estimate program effects by comparing individuals "just below" the (arbitrary) cutoff point to individuals who are "just above" it
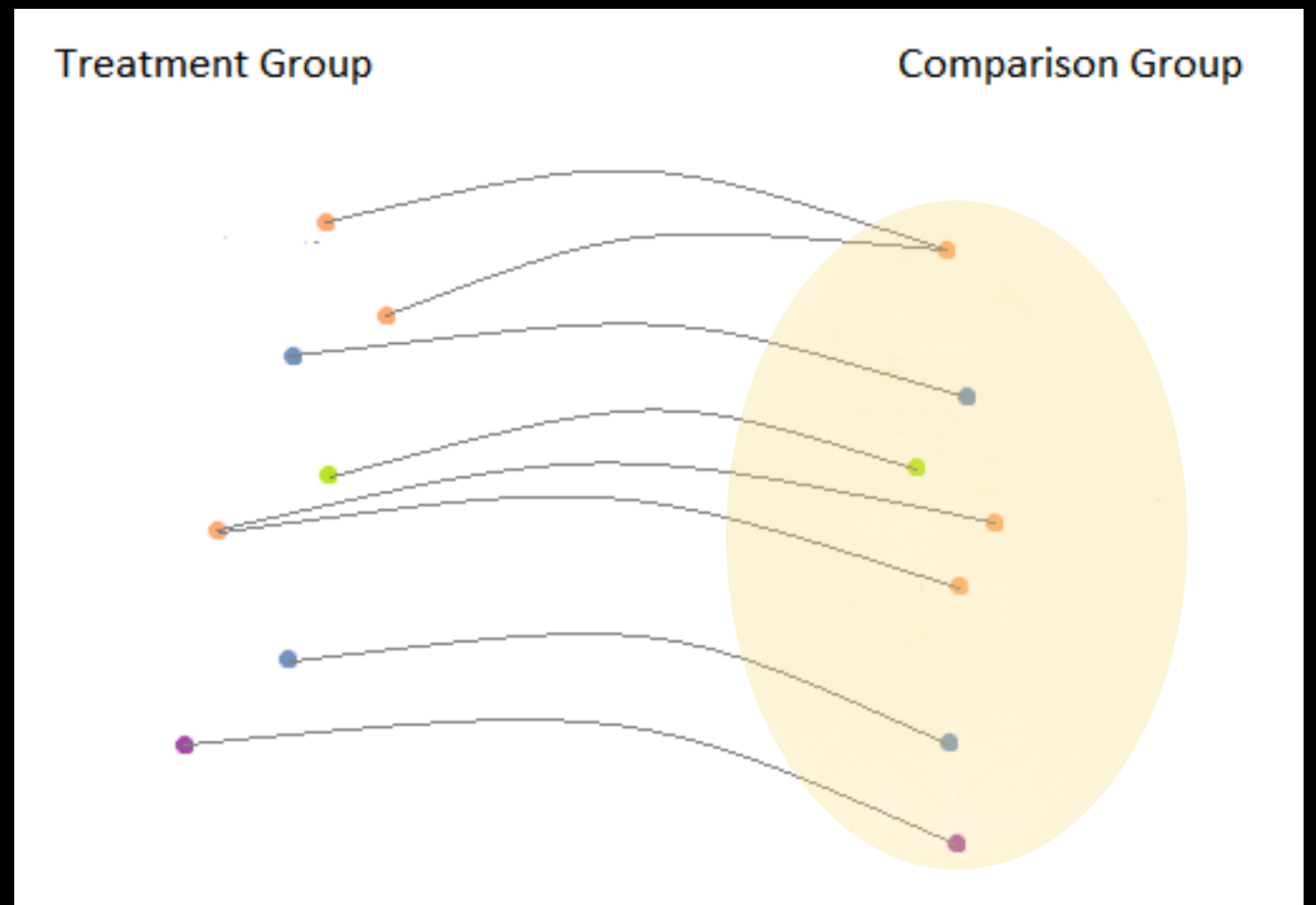
When Is This Suitable?

- Programs with clear, well-defined program eligibility rules

- A history of rules that have been consistently and strongly enforced

- Programs where the eligibility rules are relatively arbitrary

- Programs where the general target population is clustered around the cutoff point.

- Cannot use if the rule is not based on a quantifiable common measure eg subjective assessments

- No actual difference in assignment will take place if the rules are generally disregarded

- Common rules may apply to other programs e.g. assignment based on poverty lines may result in confounding with other social welfare programs

- Limited generalizability of effect size to individuals that are farther away

Matching

- Create a comparison group by matching treated individuals to untreated individuals based on observable characteristics

- Example : Use the entire bank database of customers to identify close matches for customers who take up financial education

- Estimate program effects by comparing the treated individuals to their matches

Treatment Group   Comparison Group

When Is This Suitable?

- Programs where selection bias is driven largely by observable characteristics such as age

- Programs where administrative data is available on a large set of possible characteristics for matching

- Program has a large group of untreated individuals from which to draw matches

- If selection bias is based on unobservables, matching will not resolve the issue

- Cannot be carried out if suitable background data does not exist

- If pool of potential matches is too small, may not actually be able to extract a good comparison group

- If matching is not possible but selection is largely based on observables, use other (more restrictive) methods to control for observables e.g multivariate regression

Using Pre-Post Data

Customers who take up financial education
Randomized / Matched Comparison Group
Customers who do not take up financial education

- Difference-in-Differences: estimate program effects by comparing the changes in the treatment group to changes in the comparison group

- Can be combined with other methods, even RCT

- Validity depends on the assumption of a common time-trend

Using Pre-Post Data

- Difference-in-Differences: estimate program effects by comparing the changes in the treatment group to changes in the comparison group

- Can be combined with other methods, even RCT

- Validity depends on the assumption of a common time-trend

Legend:
- Customers who take up financial education
- Randomized / Matched Comparison Group
- Customers who do not take up financial education

Y-axis: 30, 23, 15, 8, 0
X-axis: Intervention -1, Intervention, Intervention +1
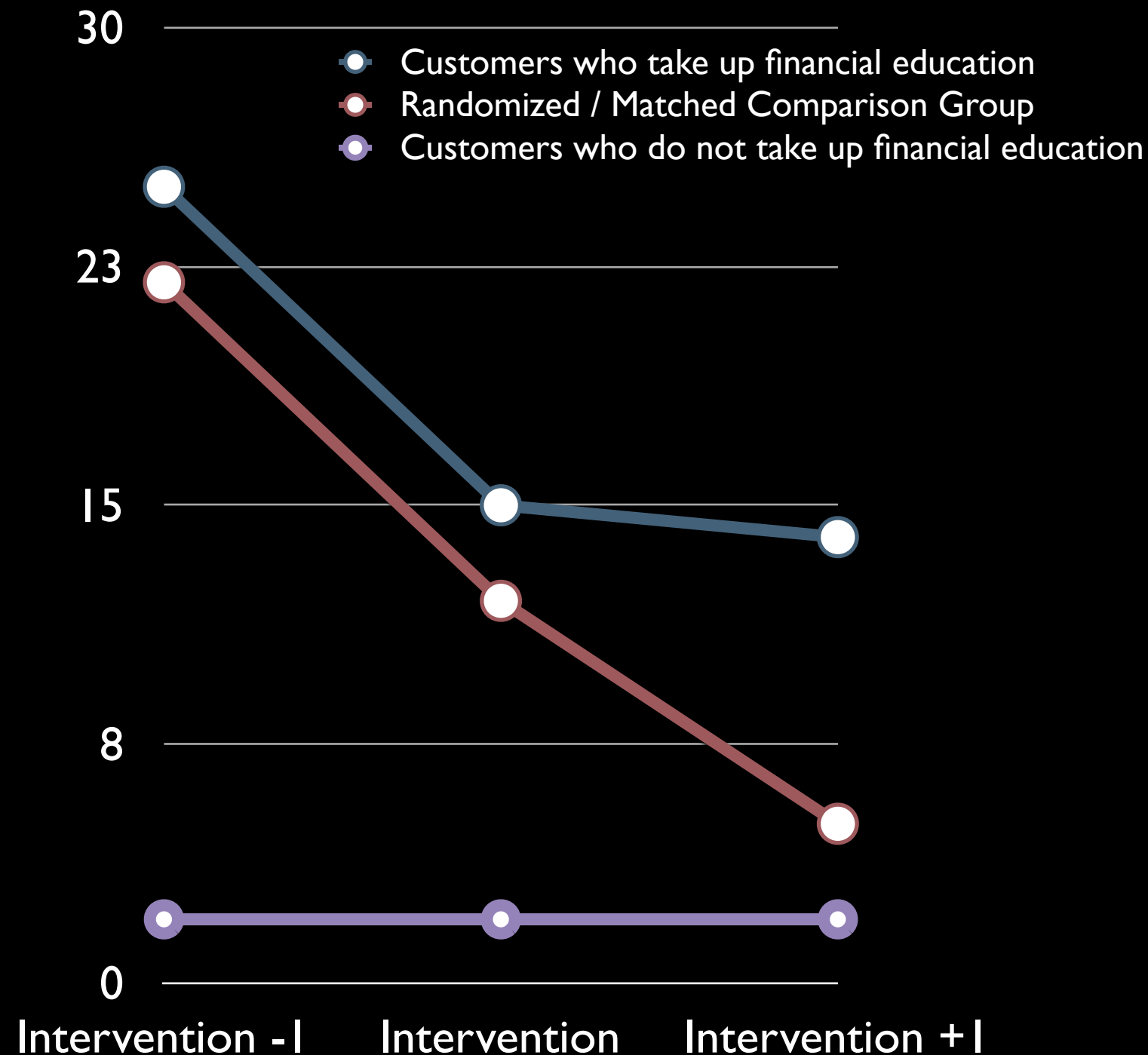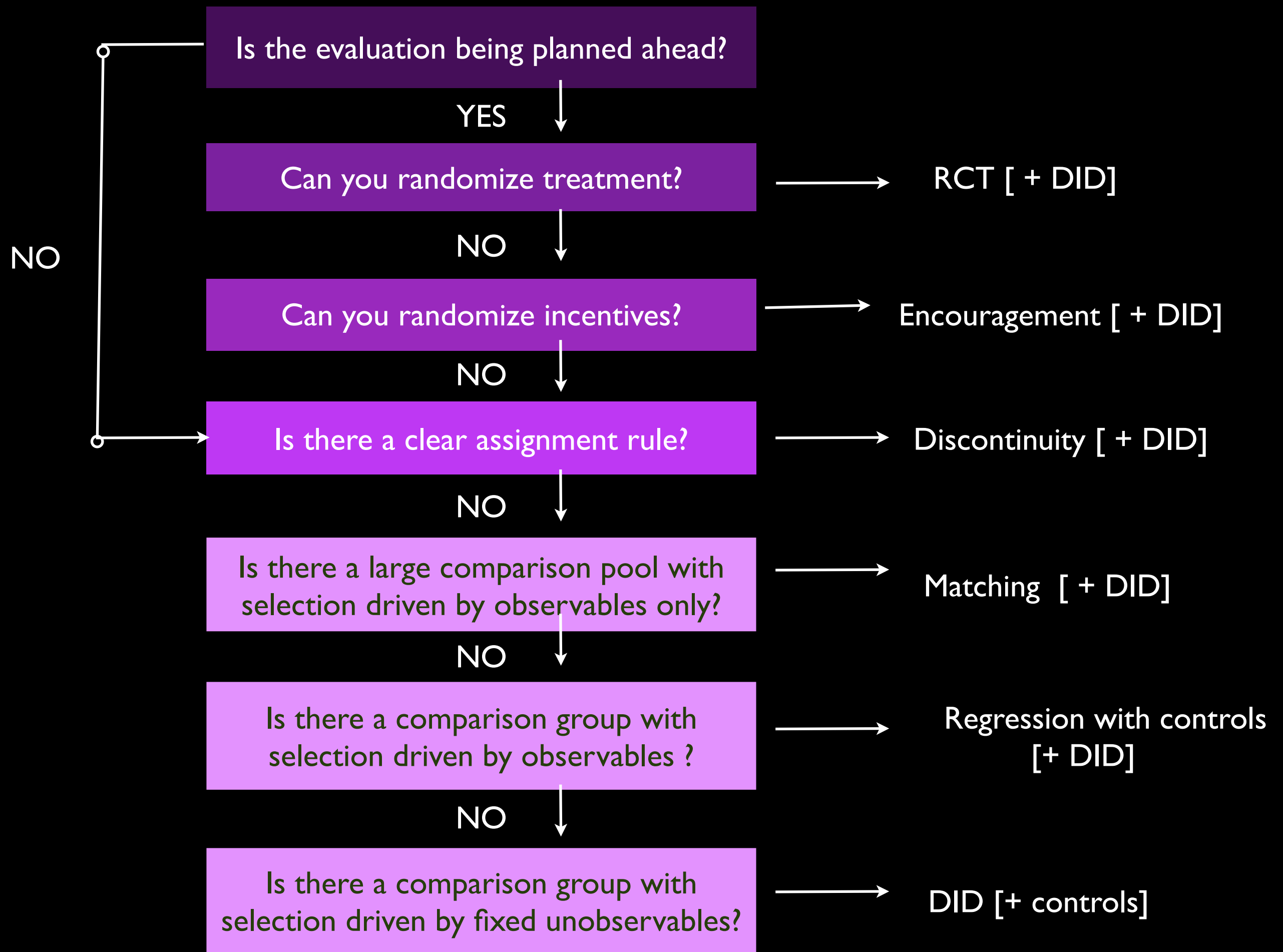
Qualitative Support

- Qualitative data gathered from sources such as focus groups or interviews trade off small sample sizes and subjectivity against breadth and depth of information not generally found in large scale surveys or administrative data

- Findings can help by

  - Explaining why impacts are observed and how realized

  - Confirming hypotheses or bringing up unexpected effects

  - Characterizing effects that cannot be feasibly explored quantitatively, including the experiences of outliers, individuals from special subpopulations, outcomes that are difficult to capture

- Example: In addition to an RCT focused on savings balances, focus groups of Community Bank customers discuss changes perceived in their experiences and attitudes towards saving in general and Community Bank in particular, and key employees are interviewed as well.

Murphy's Law

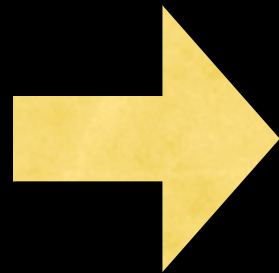Anything that can go wrong will go wrong

The Fieldwork Corollary

Even things that you didn't even know could go wrong
will also go wrong

Addressing Threats

Minimize the likelihood of failure

Measure the problem(s) that arise

Manage the situation responsibly

- **Engagement**: Explain research significance to all stakeholders

- **Design**: Anticipate problems and (i) make design adjustments (ii) ensure the appropriate process data is collected

- **Implementation**: Actively monitor operations

- **Analysis**: Make appropriate adjustments and sign remaining bias

- **Reporting**: Acknowledge limitations

- **Dissemination**: Ensure implications are understood

Three Usual Suspects

Noncompliance	Contamination	Attrition
Some Community Bank customers who were not randomized to treatment still participated in classes	Customers who did not participate get information from participants in their social networks	Customers refuse to be recontacted at the end of the study, especially customers who did not think the education was useful
Report both the unadjusted comparison between groups (Intent-to-Treat) and the comparison adjusted for participation differences (Treatment on the Treated)	Randomize in a way that minimizes spillovers e.g. branch level rather than individual Measure networks and estimate spillover effects directly	Track dropouts and check for non-random attrition between treatment and comparison Conduct sensitivity analysis

Ethical Considerations

- Financial capability programs have specific ethical issues to consider

  - Exposure to unreasonable risks - especially financial risk but also risks of physical, emotional and reputational or social harm.

  - "Informed" consent - especially when financial capability is very low

  - Equity - especially with arbitrary program assignment

- Process and analytical steps to take

  - Ensure appropriate communication, especially with respect to consent

  - Safeguard confidentiality and privacy of $ flows and information

  - Identify potential conflicts of interest and be transparent, rigorous and objective

  - Pay attention to distributional effects and benefit incidence

Understanding Impacts

Process Evaluation

- In a process evaluation, we assess the overall quality of operations relative to the program's targets using qualitative and quantitative methods.

- Example: Community Bank finds no impact of the program in one major bank branch. Process evaluation reveals poor managerial oversight and irregularly-held sessions and consequently fall in customer attendance and satisfaction.

- Process evaluation makes impact evaluation more useful and actionable, so that we can answer questions that draw on the entire causal chain from inputs to impacts.

  - If the program is performing well, is it operationally ready to scale up? In what environments is the program likely to be replicable?

  - If the program failed to deliver impacts, was this likely due to lack of resources, operational failure or inappropriate objectives/logic?

  - If the program performed well but not up to its full potential or not efficiently, what could be strengthened ?

Cost Evaluation

- Cost-benefit and cost-effectiveness analysis put impact estimates into context for decision makers considering how best to allocate resources

- In cost-benefit analysis, we weigh the total monetary benefits of the program against the total costs incurred to arrive at net present value, or the return on investment

  - Example: Community Bank wants to continue the program if it is profitable to do so. Using estimates of impact on savings balances and new accounts, Community Bank compares the increase in total revenue attributable to the program to the fees for training and costs of its employee time.

- In cost-effectiveness analysis, we compare the cost per outcome achieved by the program to another program to see whether the program provides relative value

  - Example: Community Bank wants to know if this is the best way of generating new accounts. Using estimates of impact on new accounts opened, Community Bank computes the cost per new account from this program and compares this to the cost per new account incurred by offering financial incentives.

At the End of the Day

- Not all programs can or need to undertake impact evaluation: poor or inappropriate evalution can be costly and a waste of scarce resources

Program Review	Evaluability Assessment
Stakeholders	Current/future interests and commitment
Program goals/objectives and logic	Strategic clarity
Capacity and operational environment	Technical and resource constraints
Stage of development and future plans	Suitability and timing

- A strong case for impact evaluation includes

  - Program features: innovativeness, scalability, replicability, target population

  - Feasibility:: program conditions and resources for study design/execution

  - Potential influence: input into decisionmaking and discourse

Please visit us at www.finlitedu.org !

	Our main concern is that...	Differences in outcomes could be affected by...	which leads potentially to...
Pre-and-post participants	Things other than the program change.	Other background trends	Omitted variable bias
Voluntary participants versus non-participants	Participants may systematically differ from non-participants (by choice, circumstance or conditions)	Underlying observable and unobservable differences	Selection bias