Measuring Impact
Impact Evaluation Methods

Nathan Fiala | DIW Berlin

March 2014

Why Evaluate?

- Refine program design
- Design new programs
- Improve program targeting (effectiveness)
- Identify inefficient programs (cost-benefit analysis)

Usually done through the evaluation of pilots.

Role of Impact Evaluations

Monitoring is not Impact Evaluation

- "Traditional" M&E ➔ NO inherent Causality
  - Is the program being implemented as designed?
  - Could the operations be more efficient?
  - Are the benefits getting to those intended?

- Impact Evaluation ➔ Causality
  - What was the effect of the program on outcomes?
  - Because of the program, are people better off?
  - What would happen if we changed the program?

Main Objective of Program Evaluation

Estimate the *CAUSAL* effect (impact) of program ***X*** on labor market outcome ***Y***

- E.g., What is the effect of a *job training* ***(X)*** on *employability and labor earnings* ***(Y)***?

How to Quantify the Impact?

- Causal effect:

  Impact = Outcome if in program – Outcome if not in program

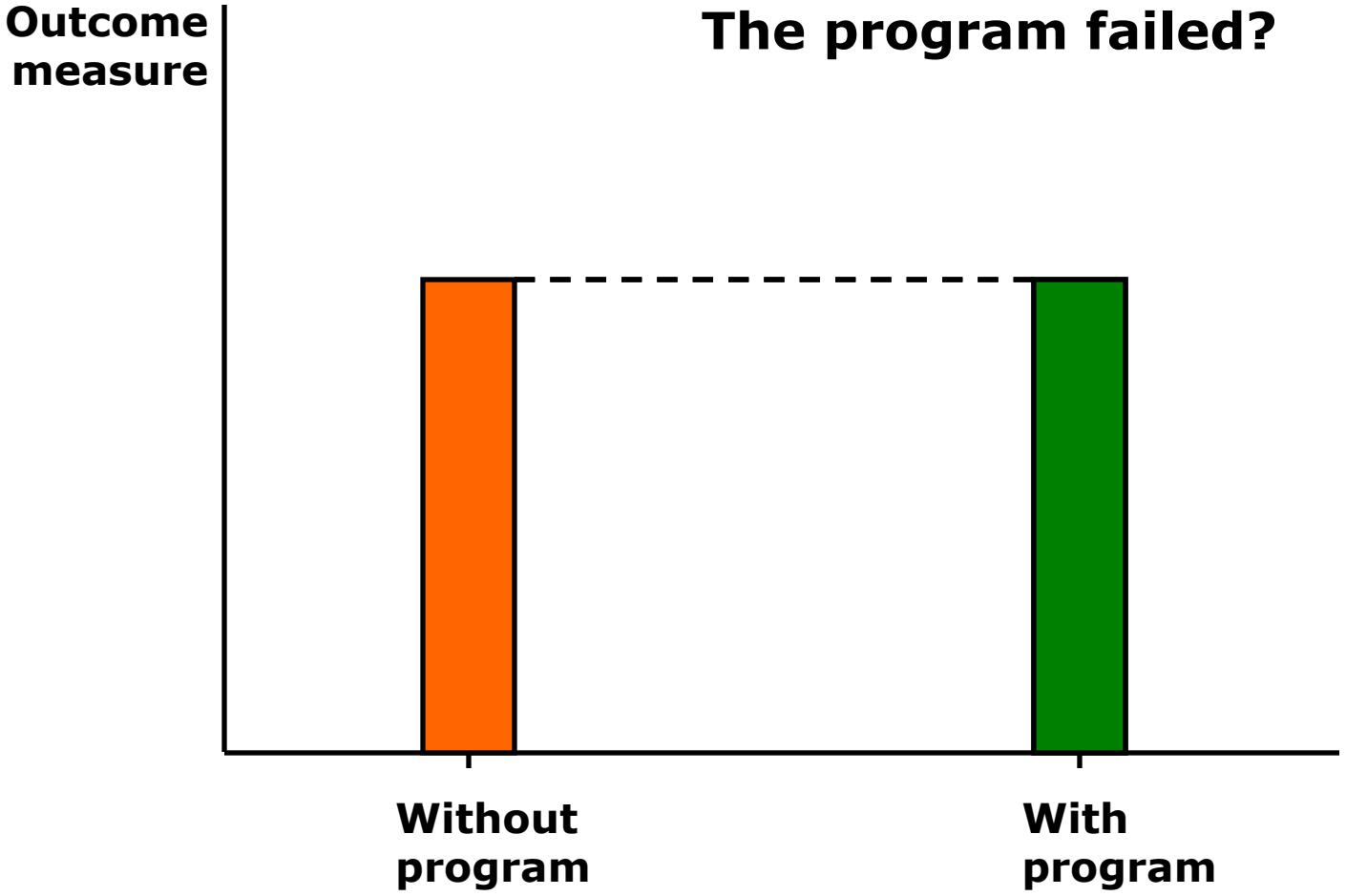
- The counterfactual quantifies what would have happened to program participant had they not participated in the program.

- But we never observe these both for the same person.

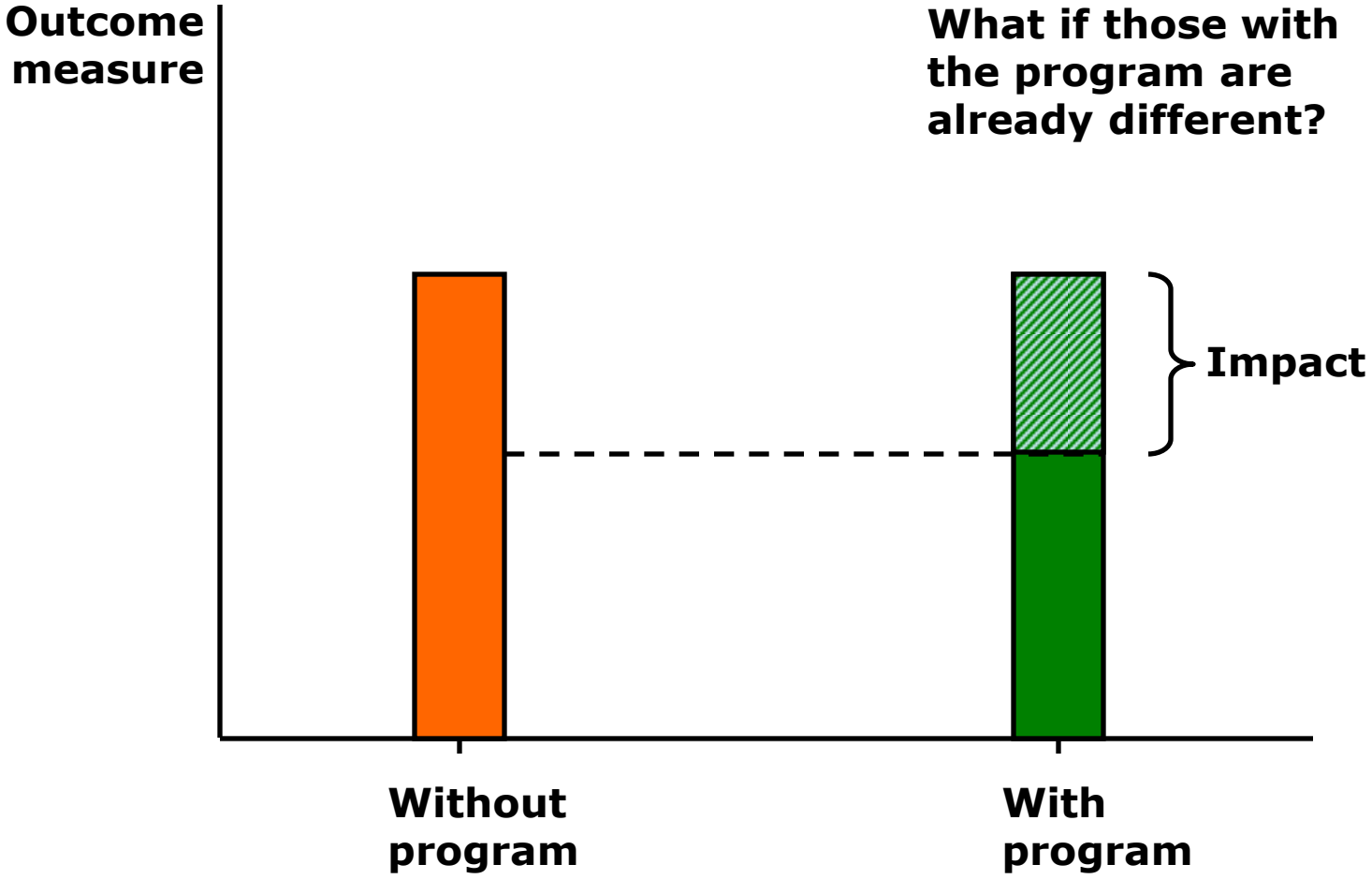
How to Identify Counterfactuals?

- Could we simply compare:
  - Same individuals before and after the program?
    - **No**. Hard to disentangle effect of program from other shocks.
  - Individuals with and without the program?
    - **No**. Often individuals could systematically differ in participation and outcomes (self-selection)

Those with the program and those without



Those with the program and those without



Problem: Unobserved Counterfactual

- Treated and counterfactual people should have identical characteristics, except for benefiting from the intervention.

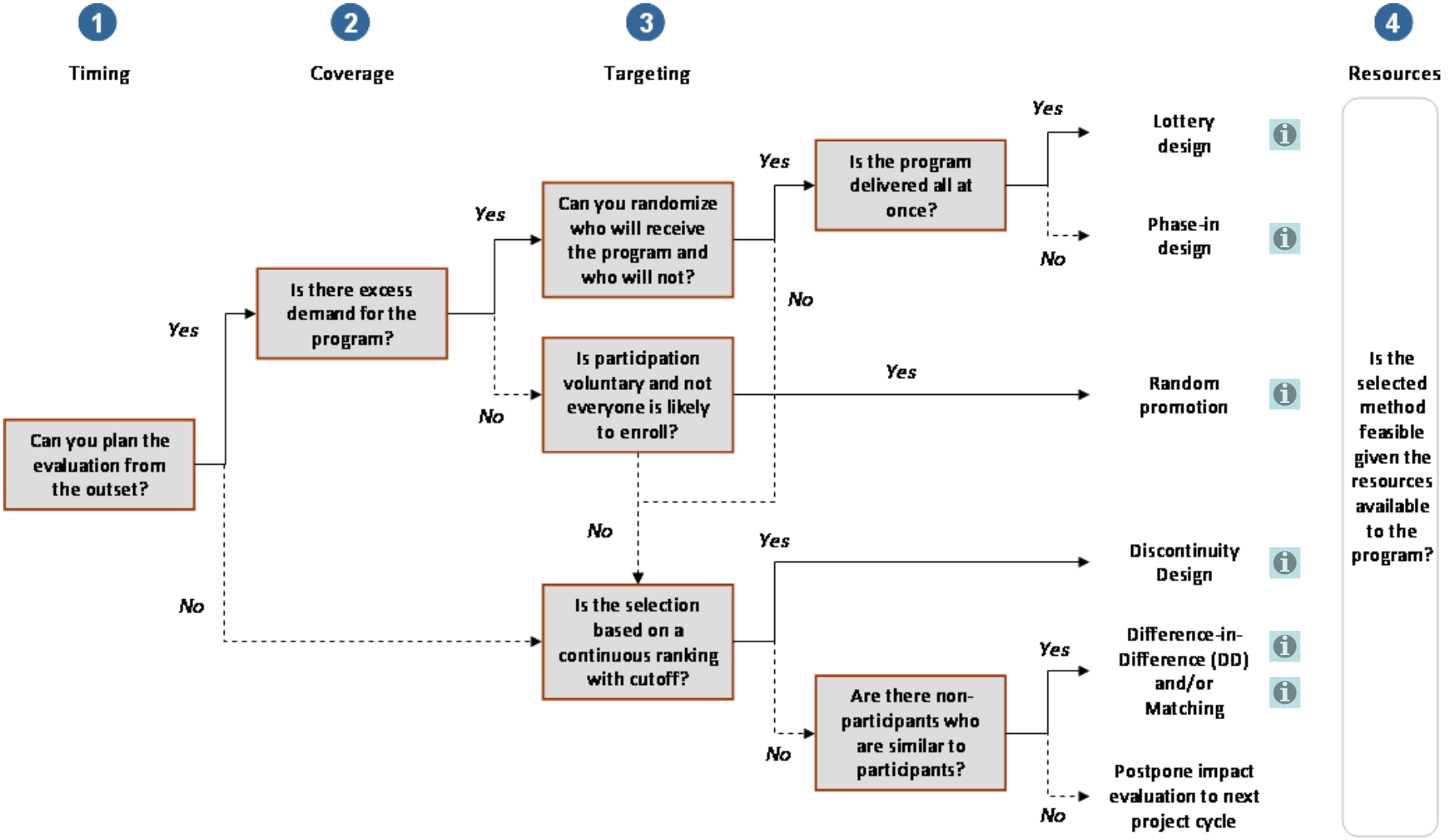
- No other explanations for differences in outcomes between the treated observation and counterfactual.

Choosing the Methodology

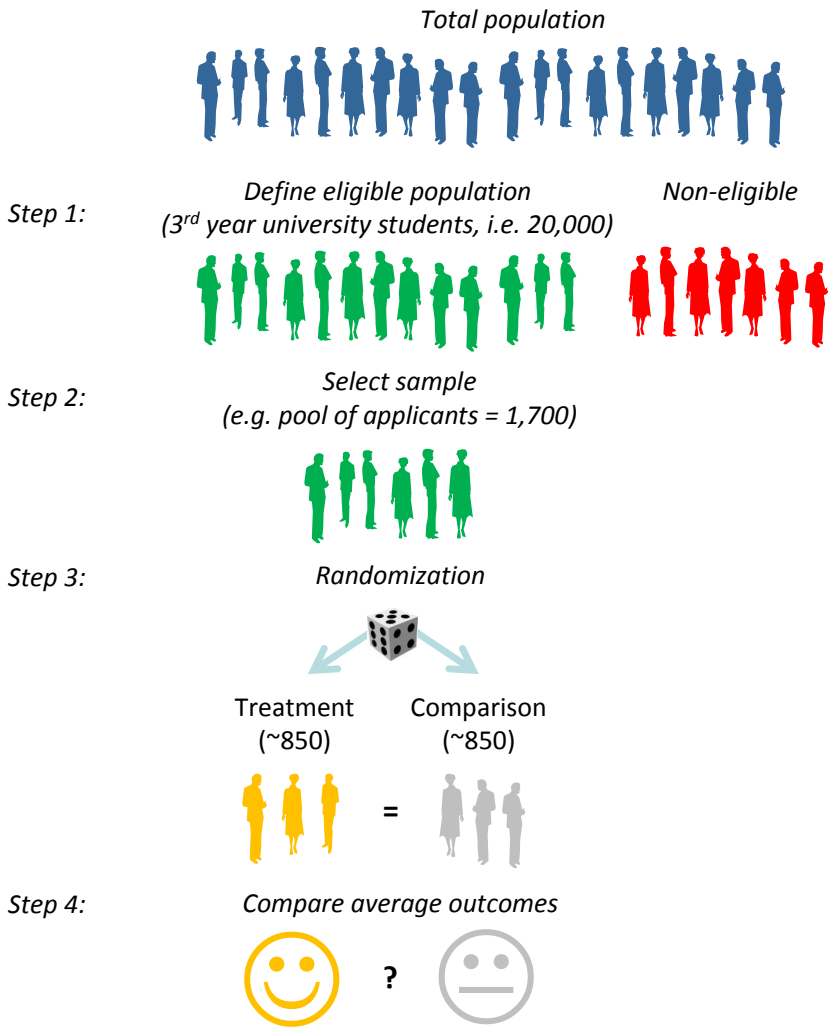
- Choose the most robust strategy (i.e. addressing self-selection) that fits the operational context.

- When possible, explore program budget and capacity constraints:
  - Universe of eligible individuals possibly larger than available resources.
  - Fairest and most transparent way to assign benefit may be to give all an equal chance of participating → randomization

Impact evaluation methods should be chosen based on operational context



Lottery design – the classic Randomized Control Trial for prospective evaluations



How it works

- Gives each individual/group the same chance of receiving the program
- Compares outcomes of those randomly selected with those not selected
- Since selection is done randomly, participants are likely to have the same characteristics on average

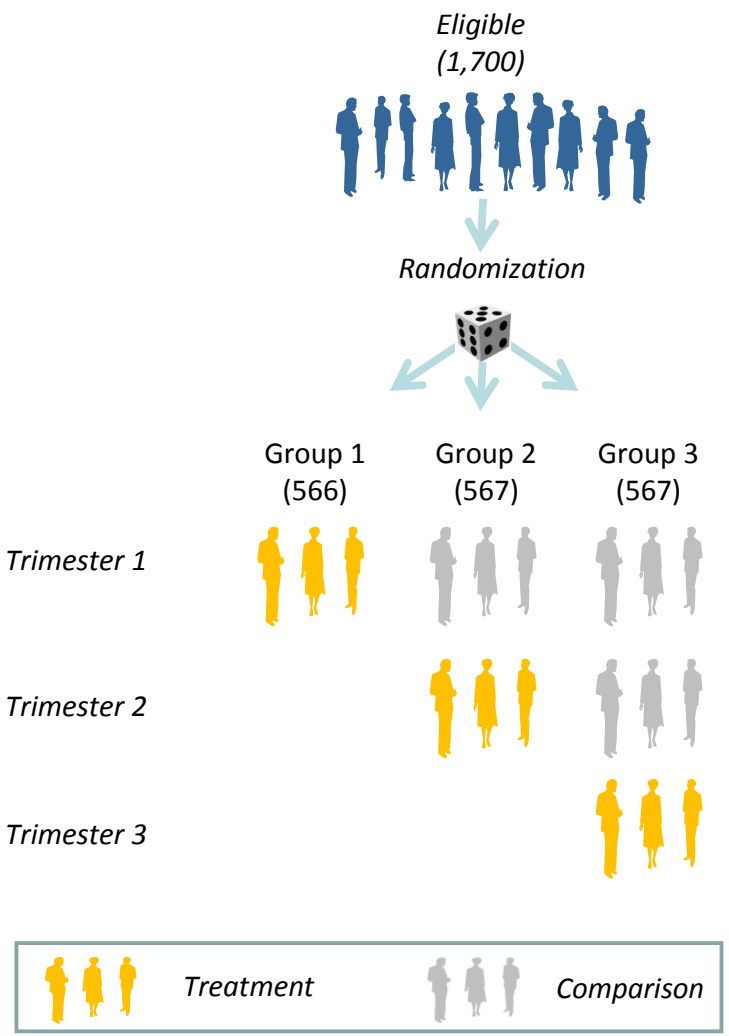
Advantages

- Most robust impact evaluation method
- Analytically simple (impact = difference in average outcomes)
- Can involve communities in selection process (fair and transparent)

Disadvantages

- Requires comparison group to be excluded from the program for duration of impact evaluation
- May be politically more difficult
- Validity depends on the fact that randomization works and is maintained (e.g. low drop-out)

Random Phase-In design – for large programs rolled out over time



How it works

- Gives each indvidual/group the same chance of receiving the program first, second, thrid, etc.
- Those receiving the program later serve as comparison group while waitlisted
- Compares outcomes of those receiving treatment first with those receiving treatment later

Advantages

- Fair and transparent selection; robust methodology
- Suits natural roll-out of programs
- Politically easier because everyone will receive the program in the end

Disadvantages

- Challenges of randomization and maintaing treatment and comparison groups over time
- Waitlisted participants may change their behavior in anticipation of the program
- Cannot estimate long-term impact of the program

Random Promotion design – for programs with universal coverage and voluntary participation

	Not promoted group	Promoted group	Observed change
Enrollment (in % of eligible population)	30%	80%	50%
Type 1: Never enroll			
Type 2: Always enroll			
Type 3: Only enroll if promoted			

Those who actually enroll in each scenario

How it works

- Randomly promotes the program to a subset of the eligible population (increases likelihood to enroll)
- Since encouragement is done randomly, not promoted group and promoted group have same characteristics on average
- Compares average outcomes of those who received the promotion with those who didn't

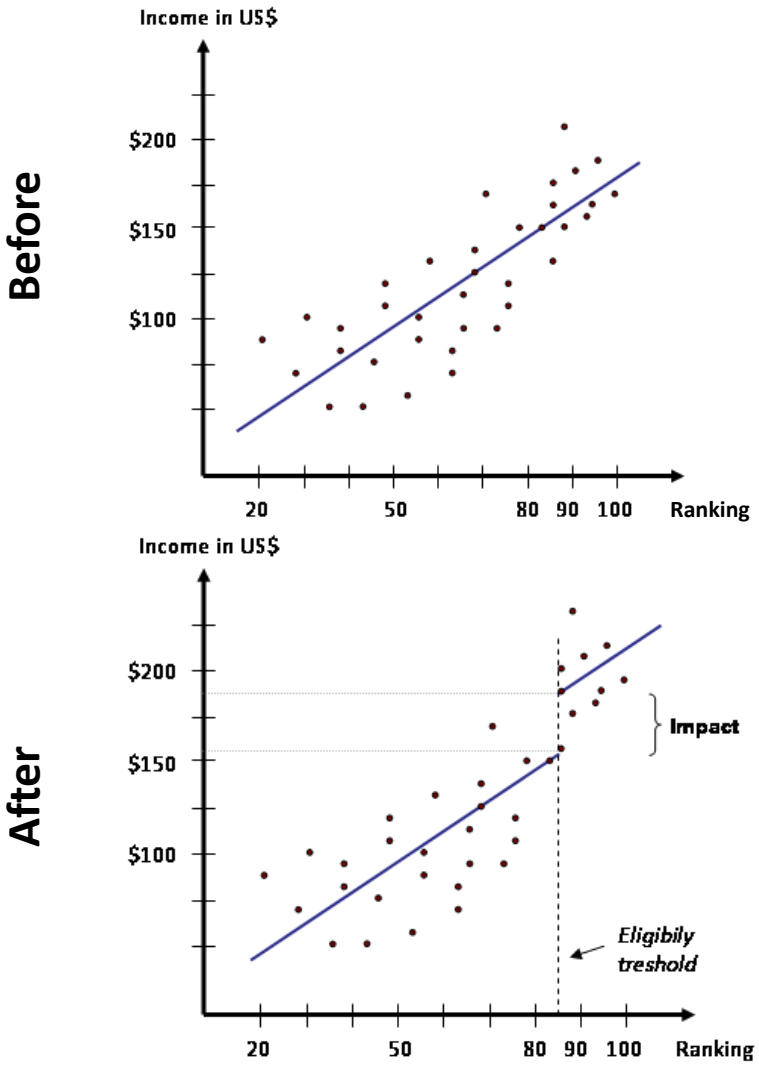
Advantages

- Robust methodology (high-quality comparison group)
- Never denies anyone to receive the program

Disadvantages

- Can only be used for specific programs (universal eligibility with volontary participation)
- Often needs larger sample size, increasing cost
- Results cannot be generalized beyond those who participated because of encouragement

Discontinuity design – when prospective beneficiaries are ranked along a continuum



How it works

- Many programs establish an eligibility threshold for acceptance into the program (e.g. based on income level, test score, credit score, age)
- Premise is that people just below and just above the cutoff line are very similar
- Compares outcomes of those who were just accepted with those whe were just rejected

Advantages

- Takes advantage of existing targeting rules
- Does not require randomization of any kind, so can be politically acceptable
- Can identify potential effects of marginal scaling

Disadvantages

- Requires very specific threshold for determining eligibility into the program
- Impact estimates are valid only for the group near the cutoff and cannot be generalized to others
- Requires large evaluation samples

Case Study: "NUSAF Youth Opportunities Program": Uganda

- Beneficiaries: Youth (16-35 years old) in a war torn region with few employment options.

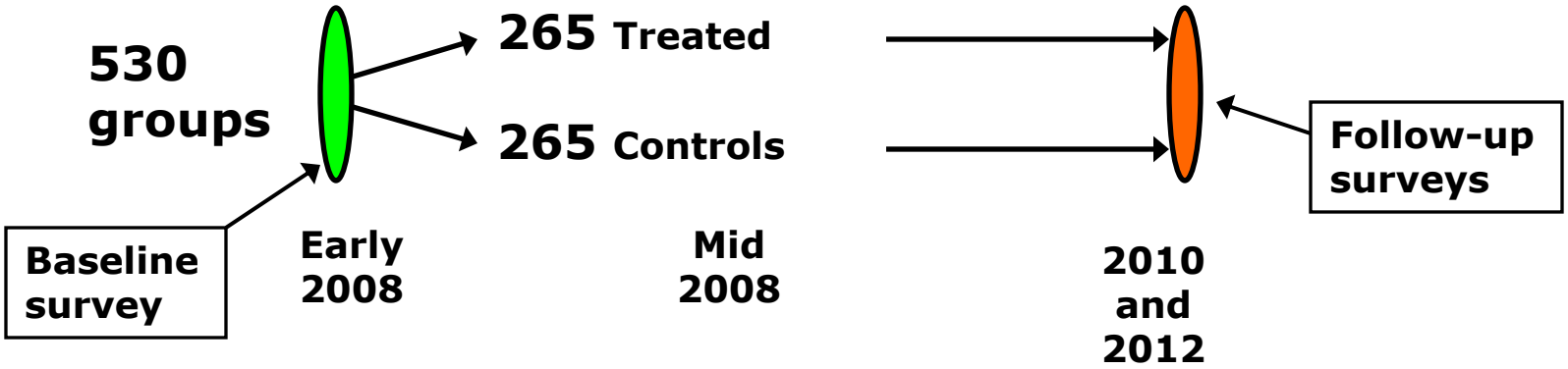
- Intervention: Program provided a cash grant to groups for:
  - 3-6 months of classroom vocational training in carpentry, metal works, sewing, etc.
  - Materials
  - Tools

- On average $8,000 given to groups of 15-30 youth (2 years of income per person).

Case Study: Evaluation Design

- Demand was high, so districts recruited more candidates than available capacity.

- Participants were randomly selected from approved groups.



Case Study: Results

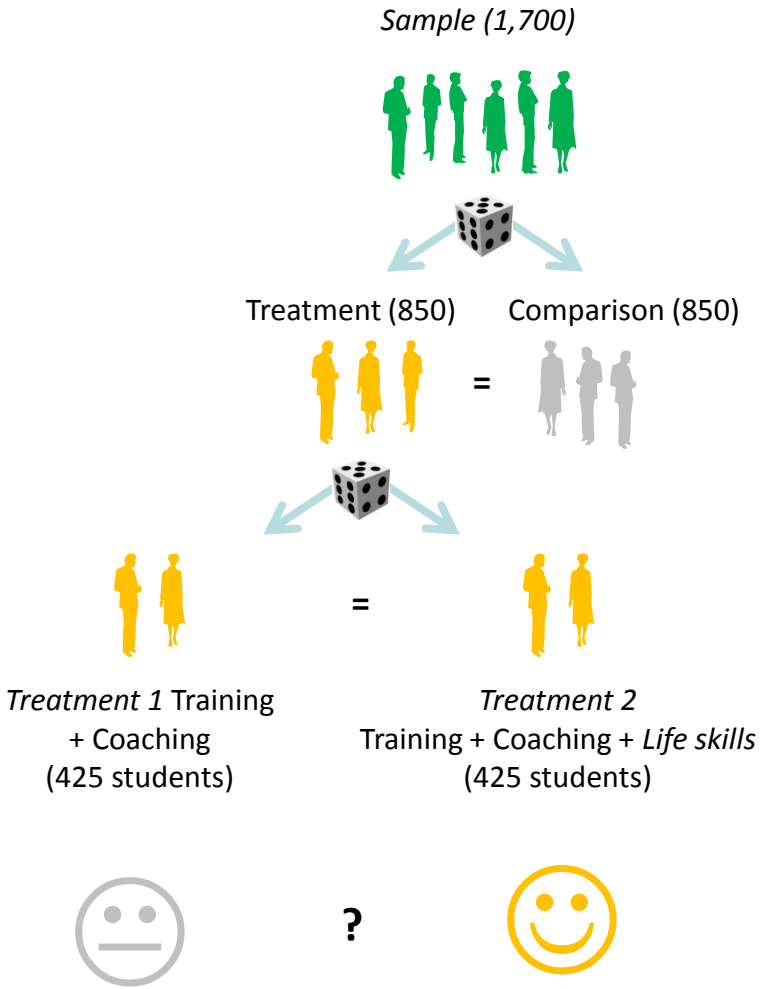
- Uptake of skills training was very high (80% vs. 25%).
- Program nearly tripled likelihood of working in a skilled labor activity.
- Incomes of those in the program increased by 18%. This grew over time for women.
- No social impacts.

Optimizing program design

- Two questions every program officer should lose sleep over:
  - What is it about a particular program that makes it work *in this context*?
  - What is the best program design to help those in need?

- Can answer both of these with cross-cutting designs.

Impact evaluations also allow for testing program alternatives



Description

- There are often many questions around the best possible program design
- You may want to test the effect of different:
  - Combination of program components (e.g. with or without life skills, access to finance, mentoring, etc.)
  - Dosages (e.g. how many hours; how much money)
  - Pedagogies
  - Delivery channels (government, private sector, NGO)
- Requires subdividing the treatment group in several subcategories
- Different treatment groups reiceive different version of the intervention
- Compare outcomes across different treatment groups and comparison group

Increase effectiveness

Increase efficiency (i.e. reduce costs)

CCD Example: NUSAF YOP

- Program questions:
  - Once money is released, should youth get support?
  - What kind of support?

- Divided treated groups into three groups:
  - Normal treatment
  - Support funded by district
  - Support funded by group

- Results forthcoming

Do the program characteristics justify an impact evaluation?

Impact evaluations usually require:

- **More time**
- **More money**
($100-500k are common)
- **More capacity**
- **A favorable political context**

Ideal program characteristics:

- **Strategically relevant and influential**
- **Innovative and untested**
- **Replicable**

Case study	
Pilot Program: $100,000 Impact Evaluation: $200,000	**Is it worth doing…?**

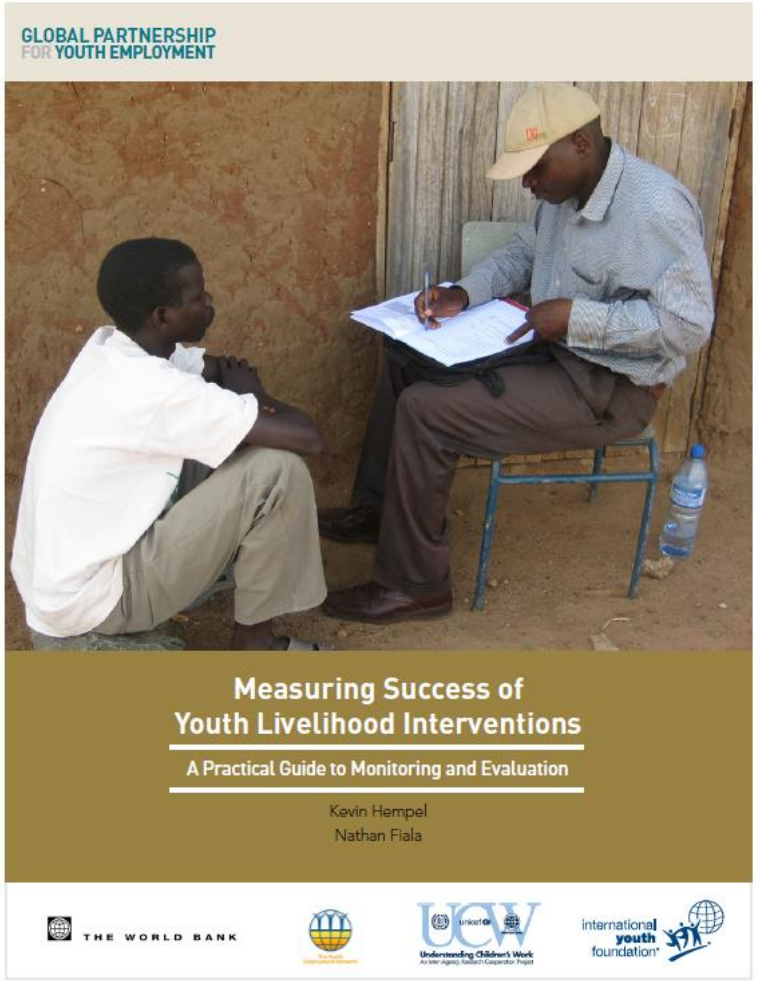
Summary

- Impact evaluation rests on the ability to provide a reliable estimate of the **counterfactual**

- Only a selected **range of IE methods** allows for obtaining a good comparison group and trustworthy impact estimates

- **No single method is best** for every program. Appropriate IE technique depends on operational context of the intervention (timing, coverage, targeting, resource constraints)

- Start the design of the evaluation **early** (yields more options for high quality IE and facilitates testing program alternatives)



Thank you

nfiala@diw.de

Visit www.gpye.org