



Human Development *Network*

Options for impact evaluation of youth programs

By David Newhouse, Labor Economist

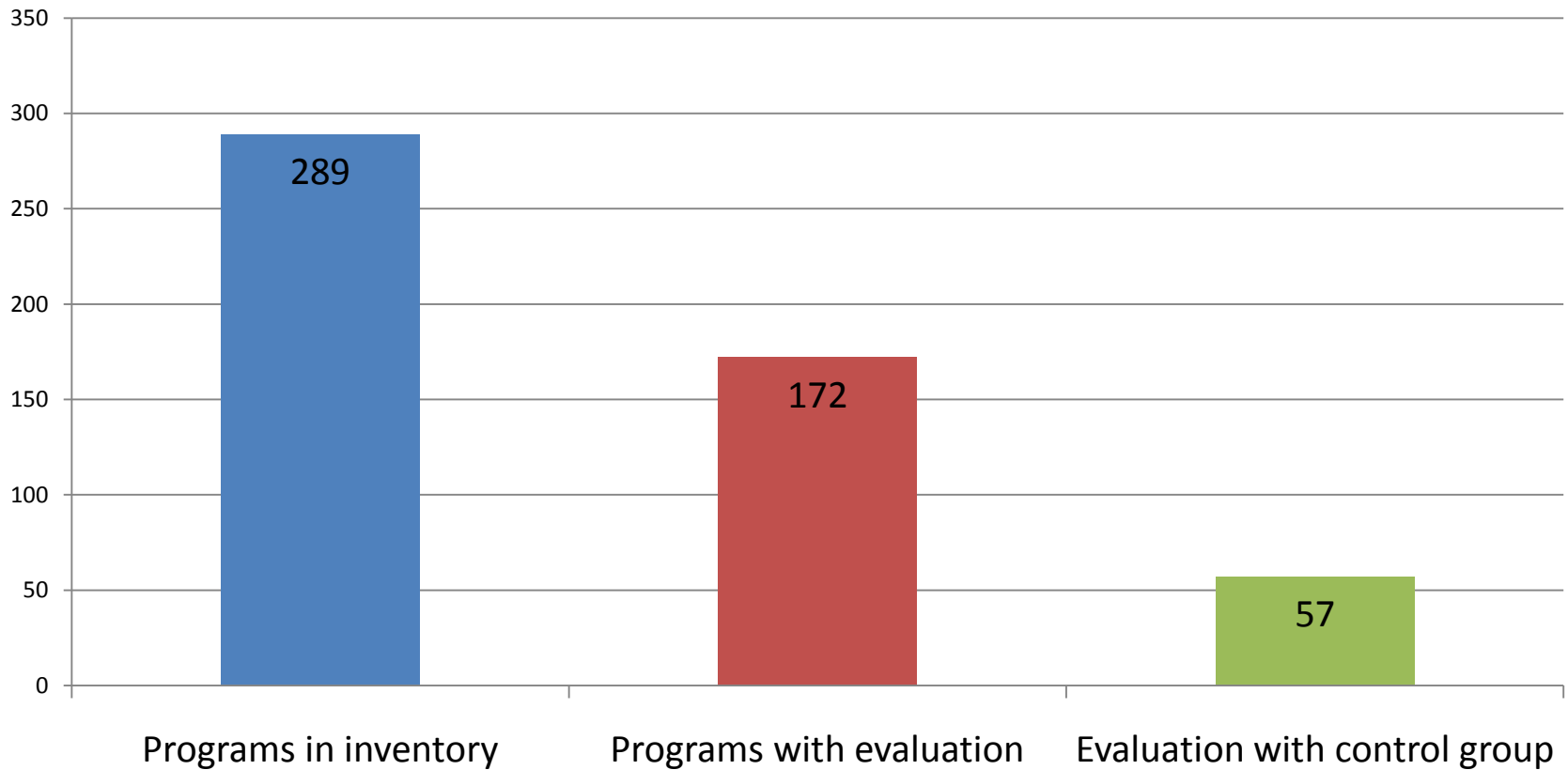
August 22, 2011

Two-part presentation

1. Review methods for constructing a control group for impact evaluation
2. Discuss an example evaluation strategy for an upcoming youth public works and training program in Papua New Guinea

Impact evaluation: Growing but still rare

Few evaluations can show impact



Source: Betcherman, et al, 2007, "A review of interventions to support young workers"

How to find the control group needed to show impact?

- Three main criteria for methods
 - **Attributability** (internal validity)
 - How much can the estimated impact be attributed to the program rather than to other differences between participants and non-participants?
 - **Comprehensiveness**
 - Is the impact an average over all beneficiaries, or just a subset?
 - Can the method estimate long-term impacts?
 - **Political acceptability**
 - Context-specific
- Three main tiers
 - Gold: Attributable and comprehensive
 - Silver: Attributable but not comprehensive
 - Bronze: Extent of attributability usually unknown.

The gold standard

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Lottery	A sample of eligible people is randomly assigned to receive the intervention.	Yes	Yes	<ul style="list-style-type: none"> • Post-intervention data for treatment and control groups • Baseline data are desirable 	<ul style="list-style-type: none"> • If study can be designed before the program begins • If demand exceeds the number of available spaces • If it is possible to exclude the comparison group for the length of the evaluation

The silver standard

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Random phase-in	Eligible units are assigned to treatment tranches and receive the program sequentially	Yes	No. Method can't measure long-term effects because controls "expire"	<ul style="list-style-type: none"> • Post-intervention data for treatment and comparison group • Baseline data are desirable 	<ul style="list-style-type: none"> • If study can be designed before the program begins • If program is rolled out over time • If randomization is seen as a fair way to determine initial access.

The silver standard (cont.)

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Random promotion	A random set of individuals or groups is encouraged to enroll in the program.	Yes	No. Estimates impact only for those swayed by promotion.	<ul style="list-style-type: none"> • Post-intervention data for treatment and comparison group • Baseline data are desirable 	<ul style="list-style-type: none"> • If study can be designed before the program begins. • If nobody can be excluded from the program • If participation is voluntary and more people will participate in the program if the program is promoted to them

The silver standard (cont.)

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Discontinuity	Individuals are ranked based on specific, measurable criteria. There is a cutoff that determines who is eligible to participate.	Yes	No. Estimates impacts only for those near threshold	<ul style="list-style-type: none"> • Post-intervention data for those nearest the cutoff • Baseline data are desirable 	<ul style="list-style-type: none"> • If randomization is not possible or the evaluation starts after the beginning of the program • If the selection is based on a transparent ranking with cutoff

The bronze standard

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Difference-in-differences	Outcomes of program participants and nonparticipants are compared before and after the intervention, controlling for observable characteristics	No	Yes	<ul style="list-style-type: none"> • Baseline and follow-up data on the same people • Multiple baselines highly desirable to establish common trends 	<ul style="list-style-type: none"> • If the study starts after the beginning of the program • If selection into program is reasonably transparent • If data is rich and includes most criteria related to selection into program

The bronze standard (cont.)

Method	Description	Attributable	Comprehensive	Data Needed	When to Use?
Matching	Individuals in the treatment group are matched with nonparticipants that have similar observable characteristics.	No	Yes	<ul style="list-style-type: none"> Large survey (census, DHS, etc.), ideally in combination with project-based household survey 	<ul style="list-style-type: none"> If the study starts after the beginning of the program and there is no control group If selection into program reasonably transparent If data is rich and includes most criteria related to selection into program

Displacement effects: An important unsolved problem

- What if program displaces others from jobs?
 - Difficult to measure
 - Depends on extent to which beneficiaries succeed because of increased productivity rather than information provision.
- What can be done?
 - Try innovative strategies like randomizing across regions or surveying firms
 - Or randomizing size of beneficiaries, as was recently done for French employment firms
 - Examine a wide variety of indicators
 - For example, could examine impact among workers who get a formal job. Arguably, wage differences among formal sector workers are more likely to reflect productivity than information.

A sample evaluation design

- Urban Youth Employment program
 - Currently under preparation in Port Moresby, Papua New Guinea
 - Objective: To provide urban youth with income from temporary employment opportunities and to increase their employability.
 - Aims to serve roughly 18,000 youth over 5 years
- Program contains public works and job training components
 - Public works: 1 week basic training + 2 months public works cleaning roads
 - Job training: 1 week basic training + 1 month in-depth training + 5 months on-the-job internship
 - Beneficiaries are assigned to public works or job training based on literacy and numeracy test scores after 1 week of basic training
- Main question for evaluation
 - Are components effective in promoting improved labor market outcomes?
 - How much more effective is on-the-job training than public works?
 - Secondary question
 - Does program reduce criminal behavior?

Evaluation strategy

- Lottery
 - Originally planned to use discontinuity method because of government's desire to target poor youth
 - Rejected because of technical difficulty of implementation
 - Including difficulty of auditing poverty status
- Discontinuity
 - Use to measure impact of public works vs. job training
 - But only estimates impact for those with test scores near threshold for acceptance into job training component

Data Collection Strategy

- Four surveys planned
 - Short baseline survey at community centers
 - Roughly 30 minutes
 - Community leader survey
 - Will ask community leaders about specific treatment and control youth
 - Try to get at perceptions of criminal behavior and other subjective indicators
 - Do community leaders know the youth well enough to give accurate reports?
 - Medium-term follow-up
 - Of two cohort after two years (year 3 and year 5)
 - Total 8,000 people
 - Long-term follow-up
 - Of second cohort after four years (year 5)
 - Total 8,000 people
 - Follow-up surveys will try phone interviews (program will provide mobiles)

Current status of evaluation

- Program planned to start in early 2012
 - But facing delays, primarily due to recruitment of key local personnel
- Plan to begin evaluating in summer or fall 2012
 - So program can get going before evaluating impacts
- Next steps:
 - Recruit local M&E coordinator
 - Advertise for and recruit survey firm
 - Finalize baseline
 - Training and field test of baseline questionnaire