

Lecture 4a RCTs in Practice

from Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial by Attanasio, Kugler and Meghir

Costas Meghir

December 2, 2009

- Lack of qualifications and skills is a key issue in developing countries

Introduction

- Lack of qualifications and skills is a key issue in developing countries
- Youth often lack employment opportunities because they lack skills

- Lack of qualifications and skills is a key issue in developing countries
- Youth often lack employment opportunities because they lack skills
- They tend to work in informal market settings with little insurance and health coverage

- Lack of qualifications and skills is a key issue in developing countries
- Youth often lack employment opportunities because they lack skills
- They tend to work in informal market settings with little insurance and health coverage
- Training these individuals may be a solution to improving their skills

- Lack of qualifications and skills is a key issue in developing countries
- Youth often lack employment opportunities because they lack skills
- They tend to work in informal market settings with little insurance and health coverage
- Training these individuals may be a solution to improving their skills
- Evidence from developed countries does not provide cause for optimism

- Lack of qualifications and skills is a key issue in developing countries
- Youth often lack employment opportunities because they lack skills
- They tend to work in informal market settings with little insurance and health coverage
- Training these individuals may be a solution to improving their skills
- Evidence from developed countries does not provide cause for optimism
- However there is almost no evidence from developing countries

The Randomisation

- The training centres selected qualified applicants.

The Randomisation

- The training centres selected qualified applicants.
- They made sure they selected 50% more than they had places.

The Randomisation

- The training centres selected qualified applicants.
- They made sure they selected 50% more than they had places.
- We chose randomly who would participate in the programme.

The Randomisation

- If the individuals who were initially assigned to the program did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the class lists randomly generated by the information system.

The Randomisation

- If the individuals who were initially assigned to the program did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the class lists randomly generated by the information system.
- In addition, individuals who were not initially offered a slot could request to be released from the waiting list in a particular class and to apply to other classes.

The Randomisation

- If the individuals who were initially assigned to the program did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the class lists randomly generated by the information system.
- In addition, individuals who were not initially offered a slot could request to be released from the waiting list in a particular class and to apply to other classes.
- In practice, there were only 56 individuals in our sample who did this.

The Randomisation

- If the individuals who were initially assigned to the program did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the class lists randomly generated by the information system.
- In addition, individuals who were not initially offered a slot could request to be released from the waiting list in a particular class and to apply to other classes.
- In practice, there were only 56 individuals in our sample who did this.
- We have information on the *initial* random assignment, so that our analysis is based on the initial offer of training

The Randomisation

- If the individuals who were initially assigned to the program did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the class lists randomly generated by the information system.
- In addition, individuals who were not initially offered a slot could request to be released from the waiting list in a particular class and to apply to other classes.
- In practice, there were only 56 individuals in our sample who did this.
- We have information on the *initial* random assignment, so that our analysis is based on the initial offer of training
- The number of non-compliers was so small that, in practice, it does not matter for the results we obtain.

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme
- We chose the number of interviews to be able to detect effects similar to those found in other programs based on a 10% level of significance.

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme
- We chose the number of interviews to be able to detect effects similar to those found in other programs based on a 10% level of significance.
- This yielded a sample of 3,300 with 1,650 in each group.

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme
- We chose the number of interviews to be able to detect effects similar to those found in other programs based on a 10% level of significance.
- This yielded a sample of 3,300 with 1,650 in each group.
- Taking an ex-ante pessimistic view on attrition, we increased the sample to 2,040 and 2,310 for the treatment and control groups, respectively.

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme
- We chose the number of interviews to be able to detect effects similar to those found in other programs based on a 10% level of significance.
- This yielded a sample of 3,300 with 1,650 in each group.
- Taking an ex-ante pessimistic view on attrition, we increased the sample to 2,040 and 2,310 for the treatment and control groups, respectively.
- The baseline sample includes 2,066 individuals in the treatment group and 2,287 controls.

The Data Collection

- Data was collected at baseline (January 2005) before the randomisation and the start of the programme
- We chose the number of interviews to be able to detect effects similar to those found in other programs based on a 10% level of significance.
- This yielded a sample of 3,300 with 1,650 in each group.
- Taking an ex-ante pessimistic view on attrition, we increased the sample to 2,040 and 2,310 for the treatment and control groups, respectively.
- The baseline sample includes 2,066 individuals in the treatment group and 2,287 controls.
- The sample was stratified by city and sex, with equal numbers of women and men in each city.

Some Evidence on the Quality of the Randomisation

Table 2: Baseline Differences between Treatment and Control Groups

	Pre-Training 2004 No Course Fixed Effects	Pre-Training 2004 Course Fixed Effects
Employment	0.001 (0.021)	-0.010 (0.021)
Paid Employment	0.024 (0.020)	0.018 (0.020)
Contract (0 if no work)	-0.000 (0.011)	0.001 (0.011)
Formal (0 if no work)	0.005 (0.011)	0.006 (0.011)
Wage and Salary	850 (5,922)	-6,806 (6,000)
Earnings (0 if no work)	-3,414 (3,394)	-3,002 (5,460)
Self-employment		
Earnings (0 if missing)		
Tenure (0 if no work)	0.942** (0.410)	0.768* (0.440)
Days Worked per Month (0 if no work)	0.188 (0.531)	-0.312 (0.535)
Hours Worked per Week (0 if no work)	-0.074 (1.185)	-1.436 (1.190)
Women	0.046** (0.018)	0.262 (0.174)
Education	0.257*** (0.083)	0.280*** (0.072)
Age	-0.113 (0.081)	-0.090 (0.085)
Married	-0.009	-0.015

Methods and Approach

- Randomisation took place at the course level

Methods and Approach

- Randomisation took place at the course level
- Individuals self select into different courses. Hence the populations may not be comparable between courses.

Methods and Approach

- Randomisation took place at the course level
- Individuals self select into different courses. Hence the populations may not be comparable between courses.
- If we were sure that the probability of being allocated to treatment was the same across courses this does not matter (exercise to prove this)

Methods and Approach

- Randomisation took place at the course level
- Individuals self select into different courses. Hence the populations may not be comparable between courses.
- If we were sure that the probability of being allocated to treatment was the same across courses this does not matter (exercise to prove this)
- It is safest to perform *within* course comparisons and then average across courses

- The within course comparison is

$$\begin{aligned}\delta_T &= E\{Y_i^1 - Y_i^0 | D_i = 1, C\} \\ &= E\{Y_i | D_i = 1, C, R_i = 1\} - E\{Y_i | D_i = 1, C, R_i = 0\}.\end{aligned}$$

- The within course comparison is

$$\begin{aligned}\delta_T &= E\{Y_i^1 - Y_i^0 | D_i = 1, C\} \\ &= E\{Y_i | D_i = 1, C, R_i = 1\} - E\{Y_i | D_i = 1, C, R_i = 0\}.\end{aligned}$$

- Our estimator of the intention to treat is obtained as a weighted average of the within course effects, i.e.

$$\delta = \frac{\sum_c P_c(1 - P_c)(\bar{Y}_{1c} - \bar{Y}_{0c})}{\sum_c P_c(1 - P_c)}$$

- The within course comparison is

$$\begin{aligned}\delta_T &= E\{Y_i^1 - Y_i^0 | D_i = 1, C\} \\ &= E\{Y_i | D_i = 1, C, R_i = 1\} - E\{Y_i | D_i = 1, C, R_i = 0\}.\end{aligned}$$

- Our estimator of the intention to treat is obtained as a weighted average of the within course effects, i.e.

$$\delta = \frac{\sum_c P_c(1 - P_c)(\bar{Y}_{1c} - \bar{Y}_{0c})}{\sum_c P_c(1 - P_c)}$$

- This is the standard fixed effects estimator

Outcome Variables

- As outcome variables we use

Outcome Variables

- As outcome variables we use
- Employment

Outcome Variables

- As outcome variables we use
- Employment
- Earnings

Outcome Variables

- As outcome variables we use
- Employment
- Earnings
- Formal sector employment

Interpretation

- As far as employment is concerned the interpretation is straightforward

Interpretation

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

Interpretation

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

- Earnings increase because of:

Interpretation

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

- Earnings increase because of:
 - 1 Increased employment (the last term)

Interpretation

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

- Earnings increase because of:
 - 1 Increased employment (the last term)
 - 2 Increased earnings of those employed. This can happen because

Interpretation

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

- Earnings increase because of:
 - 1 Increased employment (the last term)
 - 2 Increased earnings of those employed. This can happen because
 - 1 increased productivity due to improved skills

- As far as employment is concerned the interpretation is straightforward
- Now consider measuring the effects of the programme on earnings:

$$E(S_i | R_i = 1) - E(S_i | R_i = 0) = [E(S_i | P_i = 1, R_i = 1) - E(S_i | P_i = 1, R_i = 0)] \times \Pr(P_i = 1 | R_i = 0) \\ + [\Pr(P_i = 1 | R_i = 1) - \Pr(P_i = 1 | R_i = 0)] \times E(S_i | P_i = 1, R_i = 1)$$

- Earnings increase because of:
 - 1 Increased employment (the last term)
 - 2 Increased earnings of those employed. This can happen because
 - 1 increased productivity due to improved skills
 - 2 changes the composition in terms of unobservables of those working, *even* if employment did not change.

Table 3: Effect of Training Offer on Probability of Being Trained

	(1)	(2)
Training Offer	0.966*** (0.005)	0.965*** (0.005)
N	2,857	2,850
City Effects	Yes	Yes
Pre-treatment Characteristics	No	Yes
Course Fixed Effects	Yes	Yes

**Table 4: Intention-to-Treat Estimates of Effects of Training
on Employment and Earnings**

	Labor Supply					Earnings	
	<u>Employment</u>	<u>Paid Employment</u>	<u>Days/Month</u>	<u>Hours/Week</u>	<u>Tenure</u>	<u>Wage & Salary Earnings</u>	<u>Self-employment Earnings</u>
Control Means	0.723	0.603	16.903	37.472	8.781	209,017	23,260
A. Course Fixed Effects							
	0.023	0.048**	0.789	1.077	-1.775***	32,554***	-2,317
	(0.019)	(0.020)	(0.504)	(1.181)	(0.570)	(8,622)	(3,683)
N	2,348	2,348	2,348	2,348	2,335	2,348	2,348

**Table 5: Intention-to-Treat Estimates of Effects of Training
on Formal Employment and Earnings**

	<u>Contract</u>	<u>Formal Employment</u>	<u>Formal Salary</u>	<u>Informal Salary</u>
Control Means	0.205	0.235	98381	110636
A. Course Fixed Effects				
	0.078*** (0.019)	0.063*** (0.019)	29,553*** (8,469)	59.607 (7,565)
N	2,348	2,348	2,348	2,348

- What do these results tell us about the programme?

Interpretation and Rough Cost/Benefit analysis

- What do these results tell us about the programme?
- It seems to have been effective, particularly for women

Interpretation and Rough Cost/Benefit analysis

- What do these results tell us about the programme?
- It seems to have been effective, particularly for women
- A substantial part of the effect seems to have been due to a shift to the formal sector

Interpretation and Rough Cost/Benefit analysis

- What do these results tell us about the programme?
- It seems to have been effective, particularly for women
- A substantial part of the effect seems to have been due to a shift to the formal sector
- Key to evaluating the effect of such a programme is to know how long the effects last

Interpretation and Rough Cost/Benefit analysis

- What do these results tell us about the programme?
- It seems to have been effective, particularly for women
- A substantial part of the effect seems to have been due to a shift to the formal sector
- Key to evaluating the effect of such a programme is to know how long the effects last
- This would require longer term follow up.

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000
- Assume a 10% depreciation rate of these gains and a 5% annual discount rate

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000
- Assume a 10% depreciation rate of these gains and a 5% annual discount rate
- Assume same growth rate of earnings for treatment and control

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000
- Assume a 10% depreciation rate of these gains and a 5% annual discount rate
- Assume same growth rate of earnings for treatment and control
- the gains are of 2,137,490 Colombian pesos or \$899.41 (2.5 times this if the gains are permanent)

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000
- Assume a 10% depreciation rate of these gains and a 5% annual discount rate
- Assume same growth rate of earnings for treatment and control
- the gains are of 2,137,490 Colombian pesos or \$899.41 (2.5 times this if the gains are permanent)
- The direct cost of operation of the training program was US\$875 per person

Interpretation and Rough Cost/Benefit analysis

- The results imply a gain of about 25,500, so there are yearly gains of 306,000
- Assume a 10% depreciation rate of these gains and a 5% annual discount rate
- Assume same growth rate of earnings for treatment and control
- the gains are of 2,137,490 Colombian pesos or \$899.41 (2.5 times this if the gains are permanent)
- The direct cost of operation of the training program was US\$875 per person
- If we focused just on women the gains are much larger