Effects of Youth Training in Developing Countries: Evidence from a Randomized Training Program in Colombia[1]

Orazio Attanasio
University College London, NBER and CEPR

Adriana Kugler
University of Houston, NBER, CEPR,
Stanford Center for the Study of Poverty and Inequality and IZA

Costas Meghir
University College London and CEPR

May 28, 2007

Abstract

This paper evaluates the impact of a randomized training program introduced in Colombia in 2005 on the labor market outcomes of trainees. This is one of two such randomized training trials conducted in developing countries and, as such, it offers the unique opportunity to examine the causal impact of training in a developing context. We use originally collected data on individuals randomly selected and not selected to training and find that training had widespread and large effects on women, but fewer and less pronounced effects on men. In particular, women who received training have a higher probability of being employed, of having a formal job and of having a job with a written contract. Moreover, trained women earn higher wages and profits and work more days. Training also increases the probability of having a formal job and of having a job with a written contract for men, but men's earnings and profits are not affected by training. These results are robust to the use of an IV strategy which uses the initial selection into training as an instrument for having being trained to control for endogenous take-up of training. Similarly, the results are robust to IV estimates which directly control for pre-treatment characteristics and to the use of Kernel matching. Cost-benefit analysis of these results suggest that the program generates a net gain of $295.33 for women but a net loss

of $510.09 for men, suggesting that a training program of this sort would be best targeted to young women.

1. Introduction

Lack of skills in developing countries is thought to be one of the key limitations for growth in these countries. However, because the accumulation of human capital through formal schooling takes many years, the catching up process for these countries is often slow. On the other hand, vocational training may provide remedial education and allow individuals beyond school-attendance age to acquire additional skills that make them more productive in the labor market. Moreover, given the lack or little generosity of government transfers in most developing countries, increased earnings in the labor market following training may be the most effective way of helping those at the bottom of the distribution to come out of poverty. However, while there may be good reasons to advocate the use of training programs in developing countries, there is little reliable evidence on the impact of training on improving the labor market standing of the poor.

This paper evaluates the impact of a large vocational program for disadvantaged youth in Colombia. The program "Jóvenes en Acción" (which translates as Youth in Action) provided 3 months of in-classroom training and 3 months of on-the-job training to young people between the ages of 18 and 25 in the two lowest socio-economic strata of the population. Training institutions in the seven largest cities of the country received applications and were allowed to choose 45 individuals for each 30-person course offered. Subsequently, the program randomized 30 of those individuals into training and 15 out of training from those initially chosen by the training institutions. The advantage of this randomization is that one can

capture the causal effect of the program on the labor market outcomes of the participants.

The results show many positive effects of the training program on women and a few positive effects on men. Comparisons between trainees and non-trainee show that training increased the likelihood of being employed, of getting a job in the formal sector, and of getting a job with a written contract. Moreover, training increased earnings, profits, days worked and the acquisition of formal education for women. The effects for men are more limited, but also noticeable. Men who were trained were more likely to get a formal job and to get a job with a written contract. On the other hand, both women and men who received training have shorter tenures than those who were not trained, by about 3 months, which is the same as the time they spent doing in-classroom training, so this may be a mechanical effect from being withdrawn by the program from the labor force for a certain number of months.

While trainees were randomized into the program, a few of the individuals who were randomly selected did not take up the opportunity to train and very few who were not initially selected managed to get placed into training after the initial selection. Since take up may be endogenous, we rely on an instrumental variables strategy where we use the initial randomized selection as an instrument for whether the person got trained. The IV results show similar effects as the simple comparisons between trained and non-trained individuals. In particular, training increases women's probability of employment, of getting a formal job and of getting a job with a written contract by 6.2%, 8%, and 8.2% respectively. In addition, training increased women's salaries and profits by about 8.9% and 32.6%, respectively. The results for

4

men show that training increased their probability of getting a formal job and a job with a written contract by 6.2% and 11.4%, respectively. Moreover, training increases both women's and men's formal education by about a third of a year. The results are also robust to the inclusion of pre-treatment variables that differed between trained and non-trained individuals. For the most part, women who were and were not trained looked very similar before training, but men differed in terms of salaries, having a written contract and other labor market variables. We control for these pre-treatment variables either directly into the IV estimation or by using Kernel matching techniques. Controlling for pre-treatment variables in these two ways yields similar qualitative, though smaller, effects.

The most reliable training evaluations have been conducting on the basis of randomized experiments in the U.S. These studies show positive but modest effects of training on earnings and employment. Consequently, given the high cost of these type of programs, cost-benefit analyses generally suggest training programs are not worth investing in (e.g., Heckman, LaLonde and Smith, 1999; Burghardt and Schochet, 2001). However, one may expect for the returns to be higher in developing countries where the levels of skills of the population are very low to begin with. A number of training programs for disadvantaged and low skilled individuals have been introduced in recent years in Argentina, Brazil, Chile, Colombia, Dominican Republic, Peru and Uruguay, and indeed suggest high returns. However, unlike in the Colombian program, in the majority of these programs individuals were not randomized into training, so these studies have mostly been evaluated using non-experimental techniques. Consistent with the results in this paper, for the most part, the results from

these non-experimental analyses show positive effects on the earnings of women. An exception to these non-experimental evaluations in developing countries is the work by Card et al. (2007) for the Dominican Republic which finds positive though insignificant effects on earnings and the probability of getting a job with health insurance, which are attributed to small sample sizes.

The rest of the paper proceeds as follows. Section 2 provides some background on the basic design and implementation of the program Jóvenes en Acción. Section 3 describes the data. Section 4 presents OLS, IV, and Kernel estimates of the impacts of the program. Section 5 concludes.

2. Background and Description of the Program

In 1998 Colombia was hit by the strongest recession since the great depression. While the economy had an average GDP growth of 3% for the entire decade, in 1999 Colombia's GDP growth fell to -6.0%. The economy only recovered to its 3% GDP growth again in 2003.

Given the absence of safety nets in the Colombian economy, in 2001 the Colombian government introduced three emergency social programs to help those hardest hit by the recession.[2] The three programs were "Familias en Acción," "Empleo en Acción," and "Jóvenes en Acción." The first was a transfer program, similar to the Progresa program in Mexico, which provides stipends for rural families conditional on sending their children to school and providing health checks to the children. The second was a Keynesian-type program, which provided temporary

---

[2] It is worth noting that unemployment insurance did not exist in Colombia until 2003 when it was introduced by legislation.

government employment to adults. The third, "Jóvenes en Acción," which is the program evaluated in this study, provided training to young people living in urban areas.

The program "Jóvenes en Acción" was to reach 100,000 young people (or 60% of the target population) and to be given to various cohorts over a period of four years. The first cohort received training in 2002 and the last one in 2005. This analysis evaluates this last cohort.

The program was available for young people between the ages of 18 and 25, who were unemployed and who were placed in the two lowest deciles of the income distribution. The program spent US$70 million or US$700 per person. The program was provided in the seven largest cities of the country: Barranquilla, Bogotá, Bucaramanga, Cali, Cartagena, Manizales and Medellin.

Training consisted of 3 months of classroom training and 3 months of on-the-job training. Classroom training was provided by private and public training institutions, which had to participate in a bidding process to be able to participate in the program. The training institutions were selected based on the following criteria: being legally registered, economic solvency, quality of teaching, and ability to place trainees after the classroom phase into internships with registered employers. There were a total of 118 training institutions offering 441 different types of courses to 989 classes with a total of 26,615 slots for trainees, which means that the average class had 27 students. Training courses provided vocational skills ranging from cosmetology to the use of computer automated systems. The maximum number of hours of lectures was set at 350 hours for three months (or about 6 hours of lectures

during weekdays). Of the participating training institutions 43.16% were for profit and 56.84 were non-profit. Training institutions were paid according to market prices and were paid conditional on completion of training by the participants of the program.

On-the-job training was provided by legally registered companies, which provided an unpaid internship to the participants. There were a total of 1,009 companies which participated in the program. These companies operated in manufacturing (textiles, food and beverages, pharmaceuticals, and electricity), retail and trade, and services (including security, transportation, restaurants, health, childcare, and recreation). The program provided a stipend to trainees throughout the 6 months to cover for transportation and lunch of US$2.20 per day for men and for women without children and of US$3.00 per day for women with children under 7 years of age to help cover for childcare expenses.

3. Data Collection and Description

3.1. Design and Implementation of the Randomization

The key feature of this analysis is that individuals were randomly assigned to the program. This is crucial to capture a truly causal effect of training, since it is often the case that those individuals who are more likely to benefit from the training are the ones that push to get training opportunities. Moreover, given that training institutions are paid conditional on individuals finishing training, they may have an incentive to cream individuals to get those most likely to complete the courses and internships. Given this, the hardest task was to convince the training institutions to agree to

randomizing individuals into training. For this reason, the program randomized individuals conditional on an initial choice of applicants by the training institutions. In particular, the training institutions were asked to choose from their applicant pool 150% of individuals, of which two thirds of these individuals filled their slots and one third of individuals were put as back ups. The training institutions were told that the list of back ups were kept in case some of the original individuals chosen were unable to accept their positions. In practice, the reason we asked them to choose more people than slots was to be able to randomize two thirds into training and leave the other one third in a "waiting list" as controls. In practice the randomization was carried out using the special information system set up especially to register applicants into the program. Accordingly, individuals who were initially randomly chosen by the system were automatically marked as selected or not selected.

If the initially assigned individuals did not accept the training opportunity, then the training institutions were allowed to fill these slots with the next individual in the 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 training institution and to apply to other institutions. In practice, there were only 8 individuals who did this. This means that although the trainees were randomly assigned for the most part, these 8 individuals who initially did not get assigned to treatment but got trained and the 56 who turned down training may be self-selected and introduce a bias. Because we have the initial random assignment, we can eliminate this possible bias due to self-selection.

3.2. Data Collection

The initial sample proposed for the analysis was of 3,300 individuals, with 1,650 in the treatment group and 1,650 in the control group. This proposed sample was chosen on the smallest samples needed to get differences in employment probabilities and earnings significantly different from zero between treatment and control groups at the 10% level. The employment probabilities and earnings used for the two groups came from the two previous programs implemented by the government, "Familias en Acción" and "Empleo en Acción." Moreover, since the analysis requires constructing a panel and following individuals after the training program is finished, the proposed treatment and proposed samples were enlarged on the basis of expected attrition rates for the two groups. For the treatment group, the attrition rate was estimated to be of 20% and for the control group it was estimated at 40%. Consequently, the proposed samples were enlarged to 1,980 for the treatment group and to 2,310 for the control group.

The collection of information at baseline (before the provision of training) was carried out on January 2005 either before the beginning of the training program or during the first week of classes to minimize any influence of participation in the program on the interviewees' responses. The sample was stratified by city and sex, so that 50% of the individuals in the treatment and control groups would be male and 50% would be female in each city. This was done to be able to carry out separate analyses by sex, which, as evidenced below, is crucial for the evaluation. To assure the balance by sex it was necessary to carry out 60 additional interviews at baseline in the cities of Bucaramanga, Cartagena and Manizales.

Table 1 shows the expected and actual interviews conducted at baseline and in the follow-up interviews by city. In total there are 2,066 individuals in the treatment group and 2,287 in the control group in the baseline sample, or more than 100% of the expected interviews for the treatment group and 98% for the control group.

The follow up interviews were carried out between August and October of 2006 or between 19 and 21 months after the beginning of the program (since the program started at the end of January 2005) with the idea of allowing at least one year since the completion of the program to evaluate its effectiveness in terms of labor market outcomes. At the same time, since there were concerns with attrition for this highly mobile group of young people in the lowest socio-economic strata of the population, telephone updates were conducted on November 2005 or 4 months after the completion of the program. These telephone follow ups verified the basic personal information of the baseline interviewees and got up to date contact information, including address and telephone, for those who had moved or were about to move. Telephones were available for 4,298 of the 4,353 individuals initially interviewed at baseline, so that there was no phone number for only 55 or 2% of those initially interviewed. Of the ones with a phone number, 3,736 or 85.8% were reached. Of these 163 or 4.36% had moved and it was not possible to get new contact information. Of the 617 who were not reached in 71% of the cases their phone lines had been cut off or were not working. However, personal visits were then conducted to update the information of these 617 individuals.

The complete follow up in-person interviews were carried out between 9 and 11 months after the telephone update. The follow up was conducted using the initial

list of individuals in the baseline with the contact information updated by telephone in November. Table 1 presents the number of actual interviews carried out during the follow-up and the percentages out of the initial number proposed and out of the actual number of interviews in the baseline. There were 1,749 and 1,814 treatment and control individuals interviewed in the follow-up approximately one year after the training program finished. This is 85% and 79% of the treatment and control groups relative to the samples in the baseline or 81.8% of the total initial sample. At the same time, these samples represent 106% and 110% relative to the initial 1,650 required for each group to obtain a significant difference at the 10% level.

3.3. Data Description

Table 2 reports descriptive statistics of personal characteristics and labor market outcomes of individuals randomly selected and not selected to training, before and after training. Columns (1) and (2) show that women selected and not selected for training were very similar along most dimensions during the pre-treatment period, including in terms of the probability of employment, the probability of formal employment, the probability of having a written contract, days worked per month and hours worked per week, earnings, and marital status. On the other hand, selected women are older, more educated, have longer tenure in their pre-training jobs, and have marginally higher profits as self-employed before receiving training. Given that selected and non-selected women were very similar in terms of outcomes, the comparisons suggest that the process of randomization worked well for women. By contrast, the post-training comparisons in Columns (5) and (6) show many and

substantial differences between the selected women and the women not selected for training. In particular, simple comparisons of means show that women who were selected for training have a higher probability of employment, higher probability of formal employment, a higher probability of having a written contract, earn a higher salary and higher profits, are more educated and have shorter tenures than women who were not selected for training.

Columns (3) and (4) and (7) and (8) show similar statistics for selected and non-selected men during the pre- and post-training periods. The pre-training comparisons between selected and non-selected individuals show that the two groups of men differed along more dimensions than women. In particular, selected men were younger, more educated, worked fewer hours and fewer days, were less likely to be employed with a written contract, and earned lower salaries and profits in they held before receiving training. These last two differences suggest an Ashenfelter dip right before training and suggests for the importance of controlling for pre-treatment differences for men. By contrast with women, there fewer post-training differences between those selected and not selected for training. The comparisons after training suggest that selected individuals were more educated and younger and had a higher probability of having a job with a written contract, shorter tenures, and shorter hours (though only significant marginally).

While the comparisons above suggest women were indeed randomly assigned, the comparisons for men suggest differences between selected and non-selected men even before training. By the same token, the post-training comparisons show strong effects for women but fewer and weaker effects for men. These descriptive statistics

thus suggest the need to interpret the results for men with caution and to condition on pre-treatment differences especially for the sample of men.

4. The Effects of Training on Labor Market Outcomes

4.1. OLS Estimates

Given random assignment to treatment, the effect of training on various outcomes can be easily estimated by comparing the difference between trained and untrained individuals:

$$\delta = E\{Y_{1it} - Y_{0it} | D_i = 1\} = E\{Y_{1it} | D_i = 1\} - E\{Y_{0it} | D_i = 1\},$$

where Y_{1i} and Y_{0i} are the outcomes for trained and untrained individuals, $D_i = \{0,1\}$ is an indicator of participation or non-participation in the program and $E\{\cdot\}$ represents expectations. In practice, individuals were first pre-selected by the training institution:

$$\delta_S = E\{Y_{1it} - Y_{0it} | D_i = 1, PS_i = 1\} = E\{Y_{1it} | D_i = 1, PS_i = 1\} - E\{Y_{0it} | D_i = 1, PS_i = 1\},$$

where $PS_i = \{0,1\}$ is the indicator of whether the individual was pre-selected by the training institution or not. The first term represents the outcome of trainees who were pre-selected by a training institution. The second term is the outcome for trainees had they been pre-selected but not been trained. While one cannot observe this counter-factual, one can observe the outcomes for individuals who were pre-selected by a training institution but were randomized out of training. In this case, it is reasonable to assume that the outcomes for individuals pre-selected by the training institutions should be the same for those who received and did not receive training:

$$E\{Y_{0ti} | D_i = 1, PS_i = 1\} = E\{Y_{0it} | D_i = 0, PS_i = 1\} = E\{Y_{0it} | PS_i = 1\}.$$

Thus, the effects of the program conditional on pre-selection can be estimated as:

$$\delta_S = E\{Y_{1it} - Y_{0it} | D_i = 1, PS_i = 1\} = E\{Y_{1it} | D_i = 1, PS_i = 1\} - E\{Y_{0it} | PS_i = 1\},$$

the simple difference between trainees and non-trainees who had been pre-selected. Since precision can be increased by controlling for observables characteristics of individuals, X_i, below I report results from a simple OLS regression,

$$Y_{it} = \delta_S^{OLS} D_i + \beta X_{it} + u_{it},$$

(1)

where Y_{it} is the outcome variable, including the probability of employment, the probability of having a formal job, the probability of having a job with a written contract, salaries, profits, tenure, days, hours and education. As indicated above, D_i, is an indicator of participation in the program and δ represents the effect of participation in the program. X_{it} is the vector of explanatory variables, including age, a head status dummy, a marital status dummy, and city effects. u_{it} is a random error term. Standard errors are clustered at the city level to allow for correlations across individuals within cities and within cities over time.

Table 3 reports results of OLS regressions. Panel A reports results for women and Panel B for men. The results for women show that training increased the probability of employment by 0.055, the probability of having a formal job by 0.085, and the probability of having a job with a written contract by 0.091. The results also show that training increased women's salaries by 10% and profits by 32%. The results also suggest an increase of 3.9% in days worked, although this effect is only marginally significant. Moreover, the results show an increase in education of a third of a year. The results also show a decline in tenure of 3 months, which is a somewhat mechanical result because the classroom phase of the program lasted 3 months or about a third of a year, which would automatically reduce tenure by withdrawing

15

individuals from the labor force. The results for men in Panel B show similar effects on tenure and education. On the other hand, the results for men only show that training increases the probability of formal employment and of having a written contract.

4.2. IV Estimates

While the results above are suggestive, these results may be subject to a number of biases. As described above, while selection into training and no training was in principle random, take up of training was not. Out of those assigned to training, 58 individuals or 1.34% did not take up training and out of those not assigned to training by their initial training institutions, 8 individuals or 0.38% of the sample looked for training opportunities in other institutions. While the number of people who self-selected into training or no training after the initial assignment is small, these endogenous take-up may still bias the estimates upwards if those who were initially assigned but did not undergo training are those with the lowest returns to training and if those who looked for additional training opportunities have the highest returns or are more motivated or able.

To address this bias, we use an instrumental variables strategy by using an indicator of whether the individual was initially randomly selected for training or not, S_i. S_i takes the value of 1 if the person gets randomly assigned to training and 0 if the person does not get selected for training. The idea is that those initially randomly assigned to training are more likely to be trained, but getting initially selected or not should be uncorrelated with their ability, motivation or returns to training. The model

is estimated in two stages, where the first stage is a regression of the indicator of being trained on the indicator of being randomly selected to the program and other explanatory variables:

$$D_{it} = \alpha S_i + \rho X_{it} + \upsilon_{it},$$

estimated for the simple of individuals initially pre-selected by the training institutions. The second stage is then:

$$Y_{it} = \delta_s^{IV} \hat{D}_i + \beta X_{it} + u_{it}$$

where $\hat{D}_i$ is the expected value of the probability of participating in the program estimated in the first stage.

Table 4 presents the results for the first stage. Not surprisingly given the few individuals who turn down training and go on to look for other training opportunities, the first stage regressions for both men and women show that the probability of being trained is strongly correlated with the probability of being randomly selected for training.

Panels A and B of Table 5 show the IV results for men and women, respectively. The IV results for women show similar but, for the most part, smaller effects to the OLS results. For instance, Panel A shows that the probability of formal employment and the probability of having a job with a written contract are 0.08 and 0.082 higher for those who were selected into training. Similarly, the results show the earnings, profits and hours of those initially selected into training were 8.2%, 32.6%, and 3.1% higher. These are all smaller than the OLS estimates suggesting positive biases due to self-selection. On the other hand, the effects of the probability

of employment and tenure estimated with the IV are larger than the OLS effects, thought these point estimates are not significantly different from each other. Similar to the OLS estimates, the IV results suggest that training increased education and reduced tenure of women. The IV results for men also show similar effects as the OLS results. Panel B shows that the probability of having a formal job and having a written contract was 0.062 and 0.114 higher for those assigned to training. Moreover, the IV results suggest that training increased formal schooling and reduced tenure.

4.3.   Conditioning on Pre-Treatment Observables

The IV strategy in the previous section deals with the potential self-selection bias due to endogenous take-up. Aside from this potential problem, a bias may arise if the randomization failed to balance people with similar characteristics into the treatment and control groups. As shown in Table 2 women in the trainee and control groups are fairly balanced in terms of their pre-treatment characteristics. On the other hand, men in the treatment group have consistently different outcomes even in the pre-treatment period. This raises questions about whether any of the differences between the treatment and control men are simply pre-existing and not due to the program. Moreover, given the lower earnings and profits of the treatment men compared to the control men, there could be a worsening of the treatment group right prior to receiving training and this would bias results towards finding no effects of training. Thus, in the context of men, it seems important to control for the well known "Ashenfelter dip."

We take two approaches to control for these pre-existing differences in observables. First, we control directly for pre-treatment characteristics in the IV regressions. Second, we rely on Kernel matching to balance the treatment and control groups in terms of pre-treatment characteristics and then compare those selected and not selected for training in terms of their post-treatment outcomes.

4.3.1. IV Estimates with Pre-Treatment Controls

First, we directly control for the pre-treatment outcomes which differed between the groups of individuals selected and not selected into training for both men and women. In particular, we re-estimate the IV regressions controlling for the pre-treatment variables, so that the second stage regression becomes:

$$Y_{it} = \delta_s^{IV} \hat{D}_i + \beta X_{it} + \gamma X_{it-1} + u_{it}$$

where $\hat{D}_i$ is, as before, the expected value of the probability of participating in the program estimated in the first stage which now includes pre-treatment characteristics and X_{it-1} are the pre-treatment characteristics included as controls.

Table 6 reports results with the pre-treatment controls. Panel A presents the results for women, which control for age and education. Moreover, since treated and control women differed in terms of profits and tenure before training was provided, the regressions for profits and tenure also control for the pre-treatment endogenous variables. As before, the results show that training increased the probability of employment and of having a formal job for women and that training increased

women's salaries and profits. The results also show the decline in tenure and an increase in formal education for women reported in previous tables. However, the effects on the probability of having a written contract and on days worked become insignificant. The results for men reported in Panel B are also robust to the inclusion of pre-treatment controls, even though we include many other pre-treatment characteristics. In particular, we control for age, education and martial status before the treatment in all regressions and for the pre-treatment indicator of having a written contract, salary, profits, tenure, days and hours in the regressions for these endogenous variables. The results for men show that training increases the probability of having a formal job and of having a job with a written contract. Similarly, the results show the mechanical effect on reduced tenure, but now the effect on education becomes insignificant.

4.3.2. Matching Estimates

Directly controlling for pre-treatment variables allows balancing the treatment and control groups in terms of specific characteristics. In addition, we try matching methods as an alternative way to control for differences in observable characteristics in a non-parametric way. Matching methods balance the training and control groups by conditioning on the probability of being in the treatment group or the propensity score.

The propensity score summarizes the impact of the pre-treatment observables on the probability of being selected into training:

$$\Pr(S_i = 1 | X_{it-1}, PS_i = 1) = E(S_i | X_{it-1}, PS_i = 1).$$

The propensity score is used to balance treatment and control observation as much as possible. In contrast to a methodology where we directly control for pre-treatment variables, the propensity score allows one to control for many other observable pre-treatment characteristics. In particular, we control for age, age squared, education, education squared, marital status, and interactions of age, education and marital status as well as an indicator of having a written contract, salary, profits, tenure, and days in the relevant specifications. In addition, Smith and Todd (2003) have pointed to the importance of including geographic controls which may capture differences in labor markets across regions, so we include city effects in all specifications.

The idea behind matching is that given the probability of being selected, the outcomes of individuals not selected into the program will be an unbiased estimate of the outcomes for individuals who were selected had they not received training:

$$\delta_S{}^m = E\{Y_{1i} - Y_{0i} \,|\, S_i = 1, PS_i = 1\} = E\{Y_{1i} - Y_{0i} \,|\, S_i = 1, Pr(S_i = 1 | X_{it-1}, PS_i = 1)\}$$

$$= E\{E\{Y_{1i} | D_i = 1, Pr(D_i = 1 | X_{it-1}, S_i = 1)\} - E\{Y_{0i} \,|\, D_i = 0, Pr(D_i = 1 | X_{it-1}, S_i = 1)\} | D_i = 1\}$$

There are various methods to match treated and control observation. Here, we rely on Kernel matching, which matches selected and to non-selected observations by assigning weights to the various observations according to the proximity to the treatment observations.

Table 7 reports the propensity scores for being selected into training. Given that the design was randomized, we should not expect the pre-treatment observables to explain much of the probability of being selected into training. Columns (1)-(3) show propensity scores for women controlling for various pre-treatment

characteristics. These results indeed show that women's probability of being selected for training can hardly be explained by observable characteristics. Only marital status and marital status interacted with age seem to matter in terms of being selected into training in the first two columns, while only age and tenure appear to matter in the third column. In addition, the pseudo R-squared is very small in all these specifications. Columns (4)-(10) report the propensity scores for men. For men, age and age squared are significant in explaining selection into treatment in only one specification, but an indicator of having had a written contract, salary, profits, days and hours worked in the job before applying to the program do seem to matter, highlighting the need to control for these pre-treatment outcomes in the case of men. However, as for women, the overall ability of these variables to explain selection is still limited as evidenced by the low pseudo R-squares, suggesting that conditional on these observables individuals were likely randomly assigned.

Table 8 reports Kernel estimates which match selected and non-selected individuals on the basis of the propensity scores reported in Table 7. Most of the results for women are robust to the use of matching. In particular, the results show that training increased women's probability of employment, the probability of having a formal job, the probability of having a written contract, and also increased women's salaries and formal education. On the other hand, Kernel estimates for profits and days worked are insignificant. As before, there is a negative effect on women's tenure due to the withdrawal of women from the labor force for a three month period.

Panel B of Table 8 reports the results for men. As when we control directly for pre-treatment characteristics, the Kernel estimates show that training increased the

probability of having a formal job and of having a job with a written contract. Thus, training seems to affect mainly the ability to get higher quality jobs. As for women, training reduces tenure in a very mechanical way, since individuals are withdrawn from the labor force for a period of three months to take classes. On the other hand, Kernel estimates on the effects of training on formal education are insignificant.

5. Cost-Benefit Analysis

Here we suggest a simple back-of-the envelope calculation of the benefits of training to be able to do cost-benefit analysis. Taking the Kernel estimates, which are the lower bound estimates, as a benchmark, the results suggest positive effects on the probability of employment, on the probability of having a formal job, and on the salaries and formal education of women and a negative effect on women's tenure. On the other hand, the Kernel estimates suggest only positive effects on the probability of having a formal job and written contract for men and a negative effect on men's tenure.

For women, the expected gain from training will be given by an increase in the likelihood of being employed of 0.053. Conditional on being employed, training brings an increase of 0.5 of non-wage benefits in formal jobs due to the increased likelihood of getting a formal job of 0.061 after training. In addition, training increases women's salaries by 5.8%, and also brings an additional return of 15% for an additional 0.093 years of formal education. On the other hand, tenure is three months shorter in the first year of training due to withdrawal from the labor force, so

the total months worked is 6 months instead of the 9 months worked on average before training or 50% of the year. Thus, the expected gain for women will be:

E(Gain) = { (0.482+0.053)[((0.222+0.061)×1.5×(1.093×1.15)×174.6×1.058)

+ ((1-(0.222+0.061))×(1.093×1.15)×114.1×1.058)]×0.5

- (0.482)[(0.222×1.5×174.6) + ((1-0.222)×114.1)]×0.75 }

+ {(0.482+0.053)[((0.222+0.061)×1.5×(1.093×1.15)×174.6×1.058)

+ ((1-(0.222+0.061))×(1.093×1.15)×114.1×1.058)

- (0.482)[(0.222×1.5×174.6) + ((1-0.222)×114.1)]}×0.75×33

The first two lines is the expected gain of working in a formal and informal job during the first year after training, where \$174.6 is the average salary in a formal job before training and \$114.50 is the average salary in an informal job before training. Thus, the first line is the expected gain from working in a formal job and the second is the expected gain from working in an informal job. The third line is the expected return from working the person would have received without training. The fourth and fifth lines are the subsequent gains after the first year and all the way to retirement (assuming a retirement age of 65). The last line is the expected return for all subsequent years had individuals not received training. Given this expression, the return during the first year is \$2.36 or 1% of a formal salary. On the other hand, the return for subsequent years is of \$30.09, which sums up to \$992.97 until retirement for a total return of \$995.33 including the first year. Given that the amount spent per pupil was of \$700, this yields a positive benefit of the program of \$295.33 for women. Even though, we consider lower bound estimates of the effects of the

program to calculate these benefits, the results clearly show that the program generates large benefits for women.

For men, the return cannot be expected to be as large, since the effects on men were much more limited. In particular, the Kernel estimates suggest an increase in the probability of having a formal job of 0.051 and a reduction in tenure of 2 years. The expected gain of the program for men will then be:

$$
\begin{aligned}
E(\text{Gain}) = & \{(0.619)[((0.272+0.051)\times1.5\times182.36) + ((1-(0.272+0.051))\times134.78)]\times0.58 \\
& - (0.619)[(0.272\times1.5\times182.36) + ((1-0.272)\times134.78)]\times0.75 \} \\
& + \{(0.619)[((0.272+0.051)\times1.5\times182.36) + ((1-(0.272+0.051))\times134.78)] \\
& - (0.619)[(0.272\times1.5\times182.36) + ((1-0.272)\times134.78)]\}\times0.75\times33
\end{aligned}
$$

The first line is the expected gain from working in the formal and informal sector after training and the second line is the expected gain of working before training during the first year following training. The third and fourth lines are the expected gains from working in the formal and informal sectors with and without training following the first year of training and up until retirement (assuming a retirement age of 65). Given this expression, there is a loss of $13.39 per trained man during the first year after training, but a gain of $6.16 for every subsequent year for a total of $203.3 for the 33 years until retirement and a net benefit of $189.91 including the first year. However, given that the cost per pupil was of $700, this generates a loss of $510.09 per man trained. Thus, the program is not cost-effective for young men.

6. Conclusion

The program "Jóvenes en Acción" introduced in Colombia in 2005 offers a unique opportunity to evaluate the causal effect of training on young people with little education. The program offered vocational training for a total period of 6 months (3 months in classroom and 3 months on-the-job) to young unemployed women and men, who belonged to the lowest two strata in the population and who were for the most part high-school dropouts. Most importantly for the purpose of this evaluation, the program randomly selected young women and men to training or no training.

The results show that the program had widespread and large effects on women, but more limited effects on men. In particular, training increased the probability of employment, the probability of having a formal job, the probability of having a job with a written contract, earnings, profits, formal education and the average days worked in a month for men. By contrast, training only increases the probability of having a formal job, the probability of having a job with a written contracts, and formal education for men. However, both men and women experience a decline in tenure of about the same length as the classroom face of the program, indicating that these individuals are loosing working experience during that first year of training due to their withdrawal from the labor force

While individuals were randomly assigned to training, individuals could decide to turn down training or to look for other training opportunities elsewhere. Although this was not a common practice, we control for the possibility of endogenous take-up by instrumenting training with the initial selection assignment. The results are all robust to the IV analysis. Moreover, while individuals were

randomly assigned to training, men selected and not selected for training, and to much lesser extent women, differ in terms of various characteristics even before training. We tried balancing individuals in the treatment and control groups in terms of observables, by controlling for pre-treatment variables in the IV analysis and by using Kernel matching. Most results are robust to these two ways of controlling for pre-treatment differences. In particular, the results that control for pre-treatment observables show an increase in the probability of employment, an increase in the probability of holding a formal job, an increase in salaries, an increase in formal education and a decline in tenure for women. The magnitudes are somewhat smaller but large. For instance, the probability of being employed increases by more than 10%, the probability of holding a formal job by close to 30%, salaries increase by close to 6%, and formal education by 1% for women. For men, results controlling for pre-treatment characteristics continue to show positive effects on the probability of holding a formal job and on the probability of holding a job with a written contract and also a decline in tenure, though there is no effect on formal education. The results for men are more limited in terms of the outcomes affected by training but they are not trivial, as they show an increase in the probability of having a formal job of close to 20% and in the probability of having a written contract of about 50%.

We then considered estimates of the effects of the program to calculate the benefits from "Jóvenes en Acción." The results show that the program generates large benefits for women, making the program highly cost-effective. On the other hand, the program is clearly not cost-effective for young men. This suggests that future policies may want to target this type of training program to young women only.

References

Abadie, Alberto, Joshua Angrist and Guido Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1): 91-117.

Aedo, Cristian and Sergio Nunez. 2004. "The Impact of Training Policies in Latin America and the Caribbean: The Case of Program Joven," IDB Working Paper No. R-483.

Ashenfelter, Orley. 1978. "Estimating the Effects of Training Programs on Earnings," *Review of Economics and Statistics*, 60: 648-660.

Banerjee, Abhijit, Esther Duflo, Rachel Glennester and Michael Kremer. 2007. "Using Randomization in Development Economic Research: A Toolkit," forthcoming in *Handbook of Development Economics,* Vol. 4.

Burghardt, John and Peter Schochet. 2001. "National Job Corps Study: Impact by Center Characteristics," Princeton: Mathematica Policy Research.

Card, David, Pablo Ibarran, Ferdinando Regalia, David Rosas, and Yuri Soares. 2007. "The Labor Market Impact of Youth Training in the Dominican Republic: Evidence from a Randomized Evaluation," NBER Working Paper No. 12883.

Card, David and Daniel Sullivan. 1988. "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56: 497-530.

Calderon-Madrid, Angel. 2006. "Revisiting the Employability Effects of Training Programs for the Unemployed in Developing Countries," IDB Working Paper No. R-522.

Chong, Alberto and Jose Galdo. 2006. "Training Quality and Earnings: The Effects of Competition on the Provision of Public-Sponsored Training Programs," Mimeo.

Dehejia, Rajeev and Sadek Wahba. 2002 "Propensity Score Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, 84(1): 151-170.

Duflo, Esther. 2006. "Field Experiments in Development Economics," in Richard Blundell, William Newey and Torsten Persson, eds. *Advances in Economic Theory and Econometrics*. Cambridge University Press.

Elias, Victor, Fernanda Ruiz, Ricardo Cossa, and David Bravo. 2004. "An Econometric Cost-Benefit Analysis of Argentina's Youth Training Program," IDB Working Paper No. R-482.

Heckman, James, Robert LaLonde and Jeffrey Smith. 1999. "The Economics and Econometrics of Active Labor Market Programs," in Orley Ashenfelter and David Card, eds. *Handbook of Labor Economics*, Vol. 3A, pp. 1865-2097.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66(5): 1017-1098.

LaLonde, Robert. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76(4): 604-620.

Smith, Jeffrey and Petra Todd. 2001. "Reconciling Conflicting Evidence on the Performance of Propensity Score Matching Methods," American Economic Review, 91(2): 112-118.

Table 1: Proposed and Actual Sample Sizes for Pre- and Post-Treatment Periods by City

	Proposed Sample		Baseline Sample		Follow-up Sample	
	Treatment	Control	Treatment	Control	Treatment	Control
Bogotá	625	741	642	712	528	530
Medellín	378	441	386	442	333	378
Cali	340	393	344	388	292	312
Barranquilla	211	246	211	256	190	207
Bucaramanga	207	212	204	212	161	146
Manizales	99	93	99	93	81	77
Cartagena	180	184	180	184	164	164
Total	2,040	2,310	2,066	2,287	1,749	1,814

Notes: The table reports the proposed sample sizes for the treatment and control groups based on power tests of a significance difference in earnings and employment between the two groups at the 10 percent level. The Baseline sample reports the actual sample sizes before training was provided and the follow-up sample reports the actual size of the sample collected after the training program.

Table 2: Descriptive Statistics by Selection Status, Before and After the Program

| | Before Training | | | | After Training | | | |
| | Women | | Men | | Women | | Men | |
	Selected	Not Selected	Selected	Not Selected	Selected	Not Selected	Selected	Not Selected
Employment	0.482	0.471	0.619	0.606	0.697	0.636*	0.857	0.847
	(0.016)	(0.015)	(0.017)	(0.017)	(0.016)	(0.016)	(0.013)	(0.014)
Formal	0.222	0.198	0.272	0.317	0.445	0.342*	0.553	0.507
	(0.022)	(0.021)	(0.023)	(0.026)	(0.022)	(0.021)	(0.022)	(0.024)
Contract	0.206	0.208	0.226	0.305**	0.446	0.323*	0.507	0.432**
	(0.021)	(0.021)	(0.022)	(0.026)	(0.022)	(0.021)	(0.022)	(0.024)
Log Salary	12.282	12.324	12.440	12.629*	12.642	12.557*	12.823	12.790
	(0.033)	(0.030)	(0.033)	(0.024)	(0.025)	(0.025)	(0.020)	(0.031)
Log Profits	11.866	11.676§	12.157	12.382*	11.904	11.641**	12.169	12.284
	(0.078)	(0.071)	(0.064)	(0.062)	(0.098)	(0.096)	(0.102)	(0.083)
Tenure	9.073	6.703**	8.521	7.592	8.378	11.754*	10.010	13.530*
	(0.868)	(0.573)	(0.658)	(0.571)	(0.444)	(0.774)	(0.572)	(0.882)
Log Days	3.121	3.087	3.116	3.173*	3.085	3.048	3.138	3.137
	(0.018)	(0.019)	(0.017)	(0.012)	(0.017)	(0.021)	(0.016)	(0.016)
Log Hours	3.785	3.760	3.809	3.882*	3.818	3.779	3.899	3.938§
	(0.024)	(0.026)	(0.023)	(0.018)	(0.020)	(0.023)	(0.017)	(0.016)
Education	9.998	9.712*	10.114	9.775*	10.293	9.985*	10.310	10.041*
	(0.058)	(0.066)	(0.064)	(0.080)	(0.059)	(0.066)	(0.067)	(0.086)
Age	21.749	21.951**	21.525	21.817*	23.304	23.463§	23.107	23.396*
	(0.062)	(0.058)	(0.074)	(0.067)	(0.068)	(0.065)	(0.090)	(0.076)
Married	0.290	0.306	0.125	0.158§	0.343	0.351	0.212	0.228
	(0.015)	(0.014)	(0.013)	(0.014)	(0.017)	(0.016)	(0.017)	(0.018)
Max N	1,072	1,230	994	1,057	939	987	810	827

Notes: The table reports descriptive statistics for workers randomly selected and not selected into the program, as well as for workers trained and not trained under the program Youth in Action. Panel A reports descriptive statistics for women and Panel B reports descriptive statistics for men. The last row of each panel reports the maximum total of observations in each category. * indicates significance at the 1% level, ** indicates significance at the 5% level, and § indicates significance at the 10% level for differences between the selected and non-selected groups of men and women before and after treatment.

Table 3: OLS Estimates of Effects of Training on Labor Market Outcomes

	Employed	Formal	Contract	Salary	Profits	Tenure	Days	Hours	Education
					A. Women				
Trained	0.055**	0.085*	0.091§	0.100*	0.320**	-3.178**	0.039§	0.036	0.322*
	(0.022)	(0.019)	(0.044)	(0.030)	(0.119)	1.058	(0.019)	(0.032)	(0.078)
R²	1,917	1,104	1,104	1,103	159	1,268	1,280	1,279	0.082
N	0.450	0.051	0.067	0.074	0.059	0.046	0.046	0.022	1,908
					B. Men				
Trained	0.009	0.048§	0.094*	0.031	-0.114	-3.444*	-0.013	-0.023	0.250**
	(0.012)	(0.024)	(0.023)	(0.042)	(0.109)	(0.963)	(0.028)	(0.022)	(0.095)
R²	0.361	1,176	1,176	0.041	169	1,348	1,353	1,353	1,622
N	1,630	0.032	0.040	1,103	0.054	0.040	0.014	0.008	0.060

Notes: The table reports the effect of being trained on the probabilities of employment, being a formal worker, and having a written contract and on log salaries, log profits, tenure, log days worked per month, log hours worked per week, and years of education. Standard errors are reported in parenthesis. Standard errors are clustered at the city level. The regressions control for age, a head status dummy, a marital dummy, and city effects after the program. * indicates significance at the 1% level, ** indicates significance at the 5% level, and § indicates significance at the 10% level.

Table 4: First-Stage of Probability of Being Trained

	Women	Men
Selected	0.962*	0.966*
	(0.006)	(0.006)
Age	-0.001	0.001
	(0.001)	(0.002)
Head of Household	0.001	-0.020§
	(0.011)	(0.012)
Married	0.011§	-0.013
	(0.007)	(0.010)
City Effects	Yes	Yes
R²	1,908	1,622
N	0.927	0.933

Notes: The table reports the effect of being randomly selected into the program on the probability of having being trained. Standard errors are in parenthesis. Standard errors are clustered at the city level * indicates significance at the 1% level, ** indicates significance at the 5% level, and § indicates significance at the 10% level.

Table 5: IV Estimates of Effects of Training on Labor Market Outcomes

	Employed	Formal	Contract	Salary	Profits	Tenure	Days	Hours	Education
					A. Women				
Trained	0.062*	0.080*	0.082§	0.089**	0.326*	-2.725**	0.031§	0.035	0.300*
	(0.022)	(0.020)	(0.042)	(0.033)	(0.122)	(1.005)	(0.015)	(0.029)	(0.088)
R²	1,917	1,104	1,104	1,103	159	1,268	1,280	1,279	1,908
N	0.094	0.051	0.067	0.074	0.059	0.045	0.046	0.022	0.082
					B. Men				
Trained	0.011	0.062*	0.114*	0.039	-0.184§	-3.352*	-0.018	-0.030	0.279**
	(0.012)	(0.019)	(0.025)	(0.041)	(0.095)	(0.957)	(0.032)	(0.023)	(0.099)
R²	1,630	1,176	1,176	1,176	169	1.348	1,353	1,353	1,622
N	0.081	0.032	0.039	0.041	0.052	0.040	0.014	0.007	0.060

Notes: The table reports IV estimates of the effects of training on the probabilities of employment, being a formal worker, and having a written contract and on log salaries, log profits, tenure, log days worked per month, log hours worked per week, and years of education. Whether someone received training or not is instrumented with an indicator of whether the person was randomly selected into the program. Standard errors are reported in parenthesis. Standard errors are clustered at the city level. The regressions control for age, a head status dummy, a marital dummy, and city effects after the program. * indicates significance at the 1% level, ** indicates significance at the 5% level, and § indicates significance at the 10% level.

Table 6: IV Estimates of Effects of Training on Labor Market Outcomes,
Conditioning on Pre-Treatment Observables

	Employed	Formal	Contract	Salary	Profits	Tenure	Days	Hours	Education
					A. Women				
Trained	0.057**	0.066*	0.066	0.074**	0.486§	-3.102*	0.019	0.021	0.116§
	(0.022)	(0.018)	(0.038)	(0.030)	(0.238)	(0.725)	(0.019)	(0.026)	(0.053)
R^2	1,914	1,102	1,102	1,101	31	573	1,278	1,277	1,905
N	0.099	0.079	0.103	0.102	0.421	0.055	0.058	0.032	0.562
					B. Men				
Trained	0.009	0.054**	0.107§	0.032	-0.303	-2.975**	-0.025	-0.029	0.071
	(0.012)	(0.021)	(0.047)	(0.066)	(0.357)	(1.038)	(0.030)	(0.027)	(0.040)
R^2	1,619	1,167	531	523	54	698	844	842	1,611
N	0.087	0.044	0.177	0.044	0.224	0.055	0.034	0.038	0.658

Notes: The table reports IV estimates of the effects of training on the probabilities of employment, being a formal worker, and having a written contract and on log salaries, log profits, tenure, log days worked per month, log hours worked per week, and years of education. Whether someone received training or not is instrumented with an indicator of whether the person was randomly selected into the program. Standard errors are reported in parenthesis. Standard errors are clustered at the city level. The regressions control for age, a head status dummy, a marital dummy, and city effects after the program. In addition, the regressions for women control for age and education before the program and the profits and tenure regressions control for profits and tenure before the program. The regressions for men also control for age, education, and marital status before the program. In addition, the regressions for contract, salary, profits, tenure, days and hours control for these variables pre-treatment. * indicates significance at the 1% level, ** indicates significance at the 5% level, and § indicates significance at the 10% level.

Table 7: Propensity Score for Being Selected into Training

	Women			Men						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age	-0.437	0.995	-0.798§	-0.343	-0.503	-0.530	-0.639	-0.884**	-0.600	-0.546
	(0.288)	(0.978)	(0.455)	(0.300)	(0.538)	(0.531)	(0.719)	(0.446)	(0.397)	(0.398)
Age²	0.009	-0.019	0.019§	0.007	0.010	0.011	0.011	0.017§	0.013	0.012
	(0.006)	(0.021)	(0.010)	(0.007)	(0.012)	(0.011)	(0.016)	(0.010)	(0.009)	(0.009)
Education	0.197	1.090	0.309	0.274	0.016	0.023	0.016	-0.224	0.092	0.100
	(0.204)	(0.724)	(0.303)	(0.218)	(0.359)	(0.357)	(0.447)	(0.304)	(0.267)	(0.266)
Education²	-0.012§	-0.029	-0.017	-0.011	-0.008	-0.006	-0.011	0.001	-0.004	-0.004
	(0.007)	(0.026)	(0.012)	(0.008)	(0.012)	(0.012)	(0.018)	(0.011)	(0.010)	(0.010)
Married	1.960*	2.208*	3.001	-0.604	-1.938	-2.166	1.139	-0.919	-0.999	-0.864
	(0.742)	(2.512)	(1.212)	(1.120)	(1.677)	(1.666)	(2.726)	(1.438)	(1.294)	(1.294)
Age × Education	0.003	-0.022	0.001	-0.001	0.008	0.007	0.010	0.012	0.001	0.000
	(0.007)	(0.024)	(0.011)	(0.008)	(0.014)	(0.014)	(0.018)	(0.011)	(0.010)	(0.010)
Age × Married	-0.072**	-0.083*	-0.145	0.021	0.064	0.070	-0.004	0.035	0.034	0.030
	(0.031)	(0.102)	(0.052)	(0.047)	(0.071)	(0.070)	(0.111)	(0.060)	(0.053)	(0.053)
Education × Married	-0.039	-0.026	0.018	-0.003	0.021	0.029	-0.082	-0.010	0.010	0.007
	(0.030)	(0.097)	(0.049)	(0.047)	(0.074)	(0.073)	(0.100)	(0.060)	(0.054)	(0.054)
Log Profits	−	0.151	−	−	−	−	-0.371*	−	−	−
		(0.124)					(0.128)			
Tenure	−	−	0.009*	−	−	−	−	0.002	−	−
			(0.003)					(0.004)		
Contract	−	−	−	−	−	-0.231**	−	−	−	−
						(0.106)				
Log Salary	−	−	−	−	-0.387*	−	−	−	−	−
					(0.093)					
Log Days	−	−	−	−	−	−	−	−	-0.280*	−
									(0.110)	
Log Hours	−	−	−	−	−	−	−	−	−	-0.141§
										(0.80)
City Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Pseudo R²	0.009	0.098	0.032	0.011	0.001	0.035	0.026	0.095	0.019	0.017
N	2,313	210	948	2,018	820	829	259	1,002	1,188	1,184

Table 8: Kernel Estimates of Effects of Being Selected on Labor Market Outcomes

	Employed	Formal	Contract	Salary	Profits	Tenure	Days	Hours	Education
					A. Women				
Difference Selected and Non-selected	0.053[*] (0.019)	0.061[*] (0.025)	0.058[§] (0.032)	0.058[**] (0.028)	0.148 (0.298)	-3.193[*] (1.058)	0.021 (0.028)	0.022 (0.026)	0.093[§] (0.060)
N Selected	1,109	1,109	1,109	1,109	99	468	1,109	1,109	1,109
N Non-selected	1,200	1,200	1,200	1,200	109	477	1,200	1,200	1,200
					B. Men				
Difference Selected and Non-selected	0.001 (0.017)	0.051[**] (0.024)	0.126[*] (0.048)	-0.106 (0.273)	-0.303 (0.357)	-2.262[**] (0.993)	-0.025 (0.030)	-0.027 (0.028)	-0.011 (0.019)
N Selected	997	997	432	123	54	529	844	602	605
N Non-Selected	1,016	1,016	394	131	0.224	471	0.034	581	583

Notes: The table reports Kernel estimates of the effects of being selected on the probabilities of employment, being a formal worker, and having a written contract and on log salaries, log profits, tenure, log days worked per month, log hours worked per week, and years of education. The propensity scores for women include age, age squared, education, education squared, marital status, and interactions of age, education and marital status before the program and profits and tenure before the program for comparisons of these outcomes after the program. The propensity scores for men also include age, age squared, education, education squared, marital status, and interactions of age, education and marital status before the program. In addition, the propensity scores for men include contract, salary, profits, tenure, days and hours pre-treatment when comparing these outcomes after being selected into the program. Standard errors are reported in parenthesis. Analytical standard errors cannot be estimates, so we get bootstrapped standard errors. [*] indicates significance at the 1% level, [**] indicates significance at the 5% level, and [§] indicates significance at the 10% level.