An Evaluation of the CCT Program Familias en Acción on Grade Failure and School Drop-out Rates

Alkistis Zavakou

PhD Candidate at the Economics Department

European University Institute

Florence

ABSTRACT

This paper evaluates the effect of the Conditional Cash Transfer Program "Familias en Acción" on the school grade failure rates and the drop-out rates of children in Colombia. We combine propensity score matching with difference-in-differences. This approach yields much more robust results than a simple DD method, where systematic differences between the two groups would lead to misleading estimates. Moreover, compared to a simple propensity score matching it relaxes the Conditional Independence Assumption (CIA) and allows for selection on unobservables as long as the "parallel trend" assumption holds.

We find that on average the program had a small negative effect on the average drop-out rates of matched individuals. In particular, it decreased drop-out rates of girls and boys by between 2 and 4 and by 1 percentage point, respectively. The impact on average failure rates is larger. This is mainly driven by the effect on younger children (7-13 years old) and on children in urban areas whose failure rates decreased by 7 and 8.5 percentage points, respectively. Moreover, our results suggest a substantial fall in the failure rates of boys and girls in the region of 6 to 7 percentage points. However, only the former effect is statistically significant. Finally, there is no discernible impact on both failure and drop-out rates of children in rural areas. Given their considerably higher initial rates, this raises issues about changes in the experimental design in order to improve efficiency.

INTRODUCTION

Conditional cash transfers (CCTs) are programs which provide money to poor households who meet certain behavioural requirements related to investments in human capital, such as sending their children to school or bringing them to health centers on a regular basis. The aim of these programs is to alleviate poverty, increase human capital accumulation and promote gender equality. They do so by explicitly targeting the poor, focusing on children and delivering transfers to mothers. Since 1997 CCTs have been launched in more than 30 countries and there is clear evidence of success from the first generation of programs in Latin America (Colombia, Argentina, Ecuador, Mexico, Brazil, Honduras and Nicaragua) in improving the health status, increasing school enrolment rates and reducing poverty.

Referring to "Familias en Acción", it was initially designed as a safety net in response to a severe economic crisis in Colombia. In the late 1990s Colombia experienced the worst recession in 60 years: GDP fell significantly and poverty and inequality rates increased dramatically. As time progressed and given its positive observed impact on school enrolment rates and health outcomes, the program expanded beyond its initial purpose to address the crisis to the longer-term purpose of human capital accumulation. This will increase intergenerational mobility and will enable individuals from poor families to escape poverty and crime activities since better health and higher education leads to better jobs and higher wages (the average returns to education in Colombia range from 7% to 11%).

A few evaluations of "Familias en Acción" have taken place. Attanasio et al (2006 and 2008) find that the program increased school attendance and reduced child labour. Other studies show a positive effect on the nutrition and health status of the children and on consumption[1]. However, higher school enrolment rates don't necessarily lead to better educated people and hence to higher human capital accumulation since for example academic performance might have decreased and drop-out rates or grade repetition rates might have increased. After the intervention families might force their children to work in their free time instead of studying since the stipend is conditional on

1 Attanasio and Mesnard (2006), Attanasio et al (2005), Mesnard (2005) and Maldonado and Tejerina (2010)

school attendance and not on passing the grade. Consequently, this will have a negative impact on their academic performance. Actually, Attanasio et al (2008) find that although domestic work decreased considerably, the program didn't have any impact on the income-generating work of the children. This indicates that treated students continued to work in their free time. More importantly, given that the program is targeted to disadvantaged regions and aims to bring to school children with very low levels of education, its effects on the academic performance might be negative. On the other hand, the effects might be positive because these children are motivated to perform well and continue being at school in order to continue receiving the grant. Also, the fact that the grant doubles in secondary education provides children with incentives to pass primary education grades. Finally, there might be positive peer effects from school attendance for those beneficiaries that wouldn't have gone to school had the program not been launched. These positive effects could motivate them to study and pursue higher education. In their evaluation of the Mexican Progresa, Dubois et al (2004) demonstrate all the ways in which a CCT program may affect not only enrolment rates but also academic performance. They criticize all previous literature which focused on enrolment and assumed that the student will remain at school, pass the grade and consequently benefit from the return of one additional year of education. For example, they find evidence that students in secondary education are less likely to pass the last grades and graduate since the program design incentivizes them to prolong their school attendance in order to continue receiving the stipend. For all the above reasons, this paper departs from the existing literature and estimates the impact of Familias en Acción on two other important education indicators apart from school enrolment: grade failure and school drop-out rates.

According to the Economic Commission for Latin America and the Caribbean (ECLAC), in 2000 the overall drop-out rate for students before completing secondary education in Colombia ranged from 20% to 25%. Moreover, according to UNESCO the percentage of repeaters in primary and secondary education in 2002 was 7% and 5% respectively. A high repetition percentage leads to higher enrolment rates but this doesn't reflect any improvement in education. Apart from the quantity (obtained years of education) it is the quality of education (school completion and academic performance) that is equally – if not more – important. Maluccio and Flores (2004) show that Nicaragua's CCT (Red de Proteccion Social) increased retention rates by 6.5 percentage points for children in grades 1-4. Dubois, de Janvry and Sadoulet (2004) develop a dynamic education demand model in which progression to the next grade is endogenous to examine the effect of the Mexico's CCT Progresa on academic performance. They find that the program increased primary grade progression but reduced secondary grade progression. However, they attribute the latter effect to disincentives due to the fact that by default the program terminates at the third year of secondary school. Glewwe and Kassouf (2010) examine Brazil's Bolsa Escola and find that it decreased drop-out rates by 0.5 percentage points in grades 1-4 and by 0.4 percentage points in grades 5-8. They also provide evidence that it raised grade promotion rates by 0.9 percentage points in grades 1-4 and by 0.3 percentage points in grades 5-8. Finally, Glewwe and Olindo (2004) find that Honduras' CCT (Programa de Asignación Familiar) reduced drop-out rates by between 5 and 9 percentage points. Despite the large existing literature on Familias en Acción, its impact on other education indicators apart from school enrolment has not been evaluated yet. Only one paper (Garcia and Hill 2009) estimated the effect of Familias en Acción on school progression and academic performance (test scores) of the children who would have attended school anyway, even in the absence of the program. Students in the treatment group were matched to similar students in the control group and the results indicate on average no effect on test scores (positive effect for younger children but negative for older ones). However, the estimation strategy used in this paper has received strong criticism and the results are not likely to be robust (World Bank 2011). This dissertation aims to fill this gap in knowledge. In particular it aims to investigate whether the increased enrolment attributed to Familias en Acción was accompanied by lower school drop-out rates and lower grade failure rates of the beneficiaries.

The paper is organized as follows: Section 1 describes briefly the education system in Colombia and the program itself. Section 2 describes the experimental design and the data. Section 3 discusses the estimation strategy as well as the credibility of the key identification assumptions. Section 4 presents the results and Section 5 conducts several robustness checks to ensure the internal validity of the analysis. Finally, Section 6 concludes and raises some questions.

1.1 The Education System in Colombia

The National Ministry of Education offers two calendars for the academic year. The academic year can either start in February and finish in December or start in September and finish in June, depending on the climate of the region. Both calendars include the same number (198) of days of school attendance. Education is divided into three levels: primary (grades one to five), lower secondary (grades six to nine) and upper secondary (grades ten and eleven). Children usually start school between the age of six and seven and they have to attend school until the ninth grade (complete lower secondary). However, although school attendance rates for primary education in 2001 were high (93%), for secondary education they were significantly low (57%). Furthermore, there are large inequalities in school enrolment depending on income and on region. For example, in 2001 25% of children 16-17 years old, from families belonging in the top quintile of the income distribution were not enrolled in school whereas the equivalent percentage for children from the bottom quintile of the income distribution was 48% (Garcia and Hill 2009). This is because children from poor families start to work at a very early age in order to contribute to the family income. But the most striking differences exist between regions. During the period 1978-1987 the urban drop-out rate was 7% lower than the rural (McEwan 1998). Also, in the same period only 59% of rural children in the first grade were promoted to the second grade compared to the 74% for children in urban areas.

Table 1 below shows the determinants of failure and drop-out rates at the baseline survey (2002). These estimates where obtained by running two probit regressions. The failure rate is the percentage of enrolled students who, based on academic performance, failed the current grade attended while the drop-out rate is the percentage of enrolled students who left school before the academic year ended. We can see that females and children in higher grades are more likely to pass the grade and less likely to drop out from school. As expected, the education level at which the child wants to study has a large negative and statistically significant effect on both failure and drop-out rates. The impact of father's education is in line with the literature on education: higher education levels of the father are associated with lower failure and drop-out rates of the child. Quite surprisingly however, the education level of the mother has no statistically significant effect on the our two outcome variables. Consistent with what we mentioned above, children living in urban regions are more likely to pass the grade and less likely to drop out from school than those living in rural regions. The effect of the family's labour and non-labour income is negative and largest and the effect of the child's wage is positive and statistically significant as expected. In particular, the child's wage increases the probability of dropping out from school by 14 percentage points. Finally, both higher distance from school and having to pay for the meals at school increase the probability of dropping out by approximately 3 percentage points. These last three results are consistent with what the children claimed when they were surveyed. They mentioned as the major reason for their failure and drop-out rates the cost of education (Fedesarrollo 2005). This includes both direct costs (e.g. children must pay for enrollment every year, they must also pay for books, uniforms, meals, transport to school etc.) and the opportunity cost of education (foregone earnings). To overcome these problems the Colombian government launched Familias en Acción in 2002.

Table 1: Determinants of Failure and Drop-out Rates at Baseline Survey

VARIABLES	**Failure Rates**	**Drop-out Rates**
Current grade attended	-0.0313***	-0.109***
	(0.00469)	(0.0226)
Traditional School attended	-0.0256	-0.0244**
	(0.0485)	(0.0118)
Class size	0.00290*	0.0100
	(0.00166)	(0.000192)
Pay for food at school	0.0111	0.0380
	(0.0415)	(0.0925)
Distance from school	0.000519	0.0276***
	(0.00223)	(0.00604)
Education level at which the child wants to study	-0.133***	-0.109*
	(0.0312)	(0.0591)
Age entering primary school	0.00384	0.0009
	(0.00795)	(0.00139)
Age of father	-0.00276*	-0.00025
	(0.00163)	(0.00760)
Age of mother	0.0009	0.000156
	-0.0009	(0.000825)
Single parent	0.0213	-0.0032
	(0.0264)	(0.00466)
Father with primary education not completed	0.103**	0.00450
	(0.0495)	(0.0132)

VARIABLES	Failure Rates	Drop-out Rates
Father with completed primary education	-0.0081	-0.0215*
	(0.0194)	(0.0125)
Father with completed secondary education	-0.0885*	-0.0162*
	(0.0456)	(0.00953)
Mother with primary education not completed	0.0104	0.00131
	(0.0167)	(0.0132)
Mother with completed primary education	-0.0200	0.00623
	(0.0197)	(0.0117)
Mother with completed secondary education	-0.0394	-0.0108
	(0.0458)	(0.0213)
Urban region	-0.0309**	-0.0213**
	(0.0123)	(0.0079)
Female	-0.116***	-0.0934
	(0.0408)	(0.0779)
Number of students per teacher	0.00101	-0.000443
	(0.00112)	(0.000357)
Household with 1 family	-0.0253	-0.0152**
	(0.0223)	(0.00713)
Household with 2 or more families	-0.04	-0.00978
	(0.0234)	(0.00353)
Father works in the agricultural sector	0.0197	-0.00616
	(0.0121)	(0.00387)
Family's labour income	-0.125***	-0.65905***
	(0.0331)	(0.0315)
Family's non-labour income	-0.129	-0.72
	(0.104)	(0.168)
Child wage	0.00678	0.14105***
	(0.00970)	(0.012605)
Child has to buy books	0.0678*	0.000629
	(0.00705)	(0.00225)
Constant	0.152	0.139***
	(0.159)	(0.307)
Observations	3,382	3,382
R-squared	0.089	0.011

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

1.2 Familias en Acción

The program comprises three components: nutrition, education and health. The nutrition component consists of an unconditional subsidy given to poor families with children under the age of 7. The health component consists of payments to poor families conditional on bringing their children to the health centers on a regular basis for vaccination and check-ups and also conditional on the mothers attending courses on nutrition, hygiene and contraception. The education component is the largest. This consists of payments to poor families with children 7-17 years old conditional on the children attending at least 80% of school classes (mothers must provide certificates of school attendance from schools every two months). However, it is worth mentioning that the three components interact. For example, education outcomes such as academic performance will be affected not only directly by the grant conditional on enrolment but also indirectly by improved nutrition and health status. When the program was launched in 2001 the monthly grant stood at 12,000 Colombian pesos (approximately US$7) for children enrolled in primary education and 24,000 Colombian pesos (approximately US$14) for those enrolled in secondary education (in order to capture the higher wage that older children could earn). This stipend is almost equal to the sum of the estimated direct and opportunity costs of going to school. As time progressed it increased to follow inflation rates and finally in 2007 it stood at 15,000 and 30,000 Colombian pesos respectively.

Household eligibility depended on the SISBEN rate. SISBEN (Selection System for Identifying and Selecting Beneficiaries) is a welfare indicator assigning families into 6 levels. Eligible for the stipend were all SISBEN 1 families (the poorest 20% families) with children between 7-17 years old. Moreover, there were four additional eligibility criteria for municipalities:

- The municipality had to be registered for the program and provide a list of documents e.g. a list of SISBEN 1 eligibles.
- The municipality had to have at least one bank.
- The municipality had to have access to basic education and health infrastructure.
- The municipality had to have less than 100,000 inhabitants and not belong in the regions that received aid after the 1995 earthquake.

The program started operating in 2002 depending on the municipality. Initially, from the 1,024 municipalities in Colombia, 691 qualified for the program. Within them 407,076 families were eligible but only 362,403 of them were enrolled and became beneficiaries. At this point we can see that there are two major issues arising that we should take into account in our analysis: non-random assignment of treatment and low compliance (86,5%). Under the presence of non-random assignment to treatment, treated and control regions might differ in many aspects (the program targeted disadvantaged municipalities in which the quality, demand and supply of education are likely to be very low). Moreover, some eligible families didn't register to the program and this of course is not random and can be attributed to several factors. Incomplete participation might be due to lack of information about the program or about eligibility. For example, since the program was advertised by schools, the radio and the television, very poor families who didn't have radio or television or whose children didn't go to school might not be aware of its existence. It might also be due to deliberate unwillingness to participate. Under the presence of low compliance, we need to be very careful in our comparison since compliers and eligibles might differ significantly. For example, the eligibles who registered to the program are likely to be more responsible parents and care more about their children' s education than the non-compliers. This will affect children's education outcomes and if we don't take it into account, we will and up with biased estimates (selection bias). We will examine these two issues in more detail in Section 3.

2 DATA

In this section we describe the research design and data collection and present some descriptive statistics related to the sample.

2.1 Data Collection

Data at individual, family and municipality level were collected by a consortium consisting of the the IFS, the SEI and Econometria at three different periods:

- Baseline survey conducted between June and October 2002
- First survey conducted between July and November 2003 and
- Second survey conducted in April 2006

As mentioned above, randomization was not feasible since the program aimed to mitigate the consequences of the deep crisis that affected Colombia in 2000 and therefore was introduced in the municipalities facing the biggest problems. Thus, it was decided to construct a stratified sample of treated areas (they were assigned to 25 strata based on an index taking into account region, infrastructure, number of eligible households and health and school characteristics). Then the control municipalities were chosen, from the same stratum, to be as similar as possible to the treated municipalities in terms of population, area and an index of quality of life. Consequently, they end up with a final sample consisting of 122 municipalities, of which 57 are treated and 65 are controls. Within each municipality they randomly chose eligible families and schools who were interviewed in the three different periods mentioned above. By definition, the baseline survey should have taken place before the program commenced. However, due to political pressure[2] the program started in 26 from the 57 treated municipalities at the beginning of 2002 (before the baseline survey took place). We will refer to the former ones as TCP from “tratamiento con pago” (treatment with payment) and to the latter ones as TSP from “tratamiento sin pago” (treatment without payment). This means that for these 26 municipalities the data from the baseline survey were collected when the program was already running and even though in the rest municipalities the program hadn't commenced during the baseline survey, it was common knowledge and the

2 The Colombian government wanted to start the program as quickly as possible since the elections were close.

registration procedure had begun. This can lead to serious bias in our estimates since the data from the baseline survey are either directly affected by the treatment in the case of the TCPs or indirectly affected by the expectations of treatment in the case of the TSPs (anticipation effects). For these reasons we exclude from our baseline sample failure rates and drop-out rates for academic years February 2002 - December 2002, September 2001 - June 2002 and September 2002 - June 2003 and use as a baseline failure rates and drop-out rates only for the academic year February 2001 - December 2001.We compare participants (eligible individuals who registered with the program) in treated municipalities to eligibles in control municipalities. The sub-sample for our analysis consists of 16,437 children between 7 and 17 years old in the baseline survey out of which 11,062 were treated and 5,375 were not treated. As the program expanded this number increased to 19,254 children between 7 and 17 years old in the second survey out of which 13,522 were treated (among which 2,124 were the new treated / controls in the baseline but treated in the second survey) and 6,732 were not treated.

Using school responses regarding the results obtained in each grade attended, we define our outcome variables as follows: a) failure is a dummy variable equal to 1 if the child, based on academic performance, failed the current grade attended and equal to 0 if it passed it and b) drop-out is a dummy variable equal to 1 if the child left school before the current academic year ended and 0 otherwise. It is important that failure rates are defined with respect to the population of the enrolled in school children after having excluded those who dropped out. We use as a baseline the academic period February 2001 - December 2001 and compare the outcomes to those observed in the academic year February 2005 - December 2005. Thus, we have two repeated cross-sections. It is worth mentioning that although some consider Familias en Acción as a school program, schools had no role in the funding of the program and didn't receive any money. Therefore, they had no incentives to misreport data on enrolment, attendance and academic performance. Table 1 compares failure rates and drop-out rates between treated and control groups before and after the intervention. There is a clear problem of academic performance in Colombia: in 2001 almost 13% of students didn't promote to the next grade. However, we can see that even though treatment areas were more disadvantaged, they had lower failure and drop-out rates than controls. Furthermore, the table demonstrates a substantial reduction in both failure and drop-out rates in 2005. Overall, among the 13,522 children consisting our treatment group in 2005 1,261 failed the grade and 250 dropped out from school. However since we don't know the counterfactual (what would have happened in treated municipalities in the absence of the program), we can't infer anything yet about the effects of the program. Tables 3 and 4 present our two outcome variables by gender and location. We can see that girls have lower failure and drop-out rates than boys. This is in contrast to evidence from Mexico, where the higher drop-out rates for girls accounted for them receiving larger subsidies (IFS Baseline Report of the evaluation of FA 2004). Finally, both drop-out and failure rates are lower in urban compared to rural areas.

Table 2: Failure and Drop-out Rates Before and After the Program

Failure Rates	**Before (2001)**	**After (2005)**
Treatment	12.52%	9.33%
Control	13.26%	10.58%
Total	12.81%	9.69%

Drop-out Rates	**Before (2001)**	**After (2005)**
Treatment	2.21%	1.85%
Control	2.98%	2.93%
Total	2.51%	2.15%

Table 3: Distribution of Failure Rates across Socio-demographic Categories

	Before (2001)	**After (2005)**
Females	10.84%	7.76%
Males	14.65%	10.52%
Urban areas	11.85%	9.35%
Rural areas	13.77%	10.35%

Table 4: Distribution of Drop-out Rates across Socio-demographic Categories

	Before (2001)	**After (2005)**
Females	1.96%	1.85%
Males	3.04%	2.28%
Urban areas	2.47%	2.23%
Rural areas	2.56%	2.36%

Figure 1: Distribution of Failure Rates in Treatment Municipalities in 2001

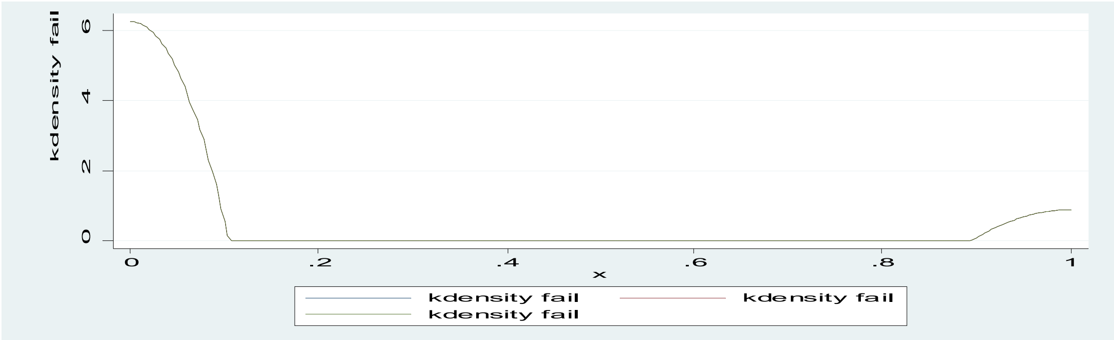


Figure 2: Distribution of Failure Rates in Control Municipalities in 2005

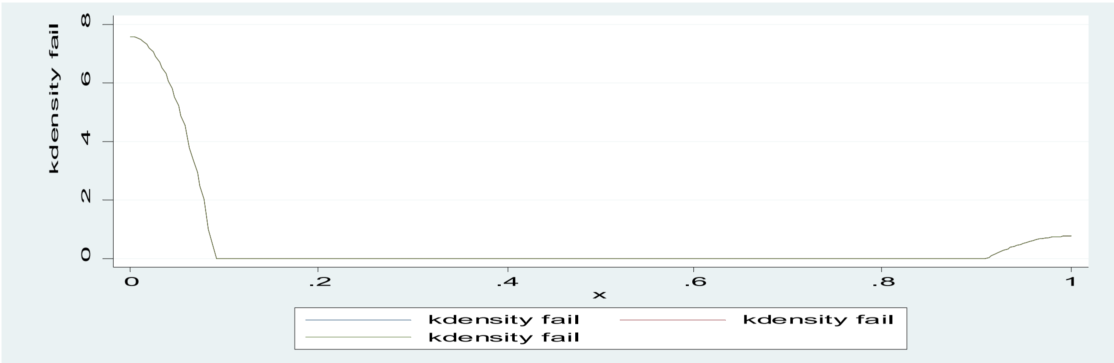
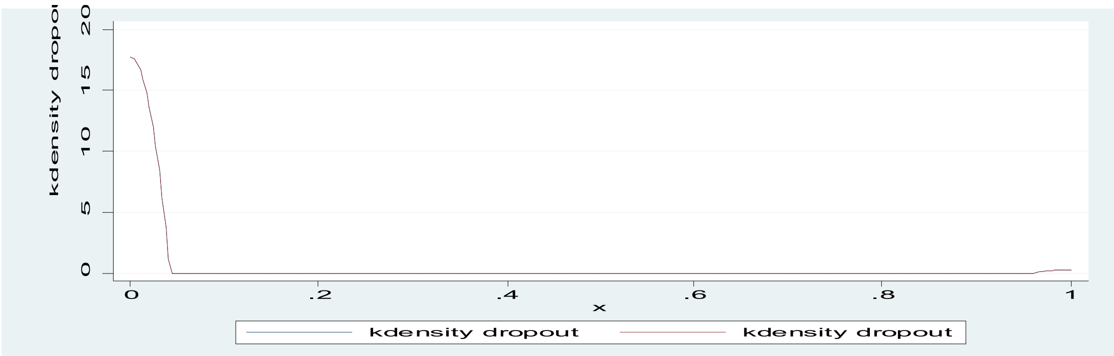


Figure 3: Distribution of Drop-out Rates in Treatment Municipalities in 2001

Figure 4: Distribution of Drop-out Rates in Control Municipalities in 2005



2.2 Comparison between Treatment and Control Groups at the Baseline Survey

The methodology used in this paper is based on the comparison of the outcomes of interest (failure rates and drop-out rates) between beneficiaries in treated municipalities and eligible households in non-treated municipalities. The necessary assumption for our results to be valid is that the 2 groups are quite similar and thus comparable. Although the control municipalities were chosen from the same stratum as treated ones in order to be as similar as possible to them, as can be seen from Table 1 in Appendix A there are some pre-program differences between the two. Treated municipalities have on average fewer schools despite the fact that they have higher population and a significantly higher number of families with children 0-17. Moreover, as expected they are more deprived compared to controls in terms of wages, labour and non-labour income and they are poorer according to SISBEN (more families with SISBEN1). Nevertheless, they face higher food prices. With propensity score matching we control for all these differences.

3 ESTIMATION STRATEGY

This section describes the estimation strategy. We want to estimate the average effect of the program on the treated. To do that we must compare the average failure and drop-out rates of the treated individuals to their average education outcome in the absence of treatment (had the program not occurred). However, the latter one is never observed and we need to construct it using outcomes from the control group. There are three main problems we have to overcome in order to construct the missing counterfactual: non-random treatment assignment, imperfect compliance and anticipation effects.

Since the experiment was non-randomized, there might be differences between treated and control municipalities. For example, less developed municipalities with less than 100,000 inhabitants like those targeted by “Familias en Acción” are likely to have less infrastructure and school inputs and lower quality of supplied education. In that case our estimates would be downward biased. Actually, as we can see from Table 1 in Appendix A, there are some differences between the two groups even after a careful method was used in order to select comparable controls (belonging to the same strata). And in general, no matter how similar treatment and control groups are on the basis of observed characteristics, it is always of concern that there might be unobserved differences between the two groups affecting our outcome variables (Attanasio and Mesnard 2005). Thus, using a simple difference-in differences method might lead to bias even if the assumption of parallel trends between the two regions holds. Moreover, even if treatment and control group assignment were random, we wouldn't obtain the causal effect of the program on the treated because of the existence of non-compliers. As mentioned above, the decision to participate to the program is endogenous and only 86.5% of the eligible individuals registered. This imperfect compliance might lead to selection bias since participants are likely to differ from non-participants (e.g. participants might be more responsible parents and this will affect the education outcomes of their children). Thus, we can't compare participants in treatment groups to eligibles in control groups. To overcome these problems we use a difference-in-differences combined with propensity score matching approach. This is the difference in outcome variables before and after the program between treatment and control groups after having matched the former ones to similar latter ones. Since treated are not a random sample of the population, the comparison group should also not be a random sample but rather have similar characteristics to those treated. Matching eliminates any selection bias due to non-random assignment and endogenous participation decision and ensures that the groups we compare are similar.

Finally, our results will also be contaminated if eligible individuals increase their demand for education in anticipation of the program. As mentioned above, to deal with anticipation effects we use as a baseline data for failure and drop-out rates one year before “Familias en Acción” was announced.

3.1 Estimation of ATT

In the case of a binary treatment $Di \in \{0, 1\}$ – as in our program – we define the potential outcomes as $Y1i$ and $Y0i$, where $Y1i$ is the outcome variable of individual i if he is exposed to treatment and $Y0i$ is the outcome variable of individual i if he is not exposed to treatment.

The observed outcome is: $Yi = Di \cdot Y1i + (1 - Di) \cdot Y0i$

Because of imperfect compliance there is a difference between the Average Intention to Treat (AIT), which is the effect on the eligibles irrespective of final participation and the Average Treatment effect on the Treated (ATT), which is the effect on the participants (those eligibles who registered with the program). The AIT is equal to:

$$E(Y1i) - E(Y0i)$$

The AIT will provide a lower bound for the ATT if the ATT is positive and an upper bound for the ATT if the latter one is negative. From a policy perspective our parameter of interest is the Average Treatment Effect on the Treated (ATT), defined as:

$$E(Y1i|Di = 1) - E(Y0i|Di = 1) = E(Y1i - Y0i|Di = 1)$$

where treatment $Di = 1$ for the registered participants and 0 otherwise.

A problem arises because the potential outcome of treated had they not been treated $E(Y0i|D = 1)$ is never observed and thus we have to construct it. This is what Paul Holland (1986) referred to as the "fundamental problem of causal inference". In the case of a randomized experiment, random assignment to treatment guarantees that $E(Y0i|Di = 1) = E(Y0i|D = 0) = E(Yi|D = 0)$. Consequently, we can use as the counterfactual the observed expected outcome in the control group $E(Yi|D = 0)$. However, we can't use this in the case of a non-experimental study as the one examined because it will lead to selection bias:

$$E(Yi|Di = 1) - E(Yi|Di = 0) = E(Y1i|Di = 1) - E(Y0i|Di = 0) =$$

$$= E(Y1i|Di = 1) - E(Y0i|Di = 1) + E(Y0i|Di = 1) - E(Y0i|Di = 0) =$$

$$= ATT + E(Y0i|Di = 1) - E(Y0i|Di = 0)$$

The difference $E(Y0i|Di = 1) - E(Y0i|Di = 0)$ is known as "selection bias". Those treated for example come from poor families with low education levels and thus are likely to have higher failure rates and drop-out rates than the non-treated anyway (even in the absence of the program). Propensity score matching can eliminate the bias if the assumptions made are credible. The availability of many pre-treatment characteristics from the baseline survey that are very likely to affect both treatment status and potential outcomes makes propensity score matching a suitable methodology for the evaluation of "Familias en Acción".

3.2 Propensity Score Matching

Matching methods are used to correct for non-random assignment to treatment and/or endogenous participation in non-randomized experiments by pairing program participants with individuals from the control group who have similar observed characteristics. Matching actually changes the distribution of the observed characteristics in the control group so that it resembles the one in the treatment group and estimates the ATT by subtracting the average outcomes of matched controls from the average outcomes of matched treated individuals. However, if we have many observed characteristics, it is difficult to much on all of them. In such settings it is preferred to match instead on one variable (the propensity score) in order to reduce the dimension of the set of controlling covariates (Rosenbaum and Rubin 1983).

The propensity score is the probability of receiving treatment conditional on observed covariates *X*, defined as:

$$P(X) = Pr(D = 1|X)$$

Heckman et al (1998) examine whether it is better in terms of efficiency to match on *X* directly rather than on *P(X)*. Their results indicate that for the ATT the 2 methods are equivalent. In order to use the propensity score three necessary assumptions are required:

<u>Assumption 1</u>: SUTVA (Stable Unit Treatment Value Assumption): Potential outcomes of one individual are not affected by the participation status of others. This can be written:

$(Y1i, Y0i)$ do NOT depend on $j \neq i$

For the ATT the assumption needed is that *Y0i* does NOT depend on $j \neq i$. It is sufficient to ensure that the potential outcomes for the control group are not affected by treatment (Attanasio and Angelucci 2006).

<u>Assumption 2</u>: CIA (Conditional Independence Assumption): Potential outcomes are independent of treatment status conditional on observable characteristics *X*. This can be

written:

$$(Y1i,\ Y0i) \perp Di|X$$

Another way of thinking of this is that given *X* the allocation to treatment is random. This implies that we observe all variables that affect both treatment status and potential outcomes. Since we have defined the propensity score as $P(X) = Pr(D = 1|X)$, if the CIA holds it implies that $(Y1i,\ Y0i) \perp Di|P(X)$. This is known as the Rosenbaum and Rubin Theorem and it states that if treatment assignment is random given the covariates X, then it will be random as well given the propensity score.

Assumption 3: Common Support: There does not exist any *x* such that the propensity score equals 1 or 0.

$$0 < Prob(D = 1|X) < 1$$

This ensures that individuals with the same characteristics have the same positive probability of treatment and rules out cases of perfect predictability of treatment given *X*. Heckman et al (1998) show that for the ATT the assumption needed is $Prob(D = 1|X) < 1$. It is sufficient to ensure the existence of potential matches in the control group only. If some individuals in the treatment group can't be matched to individuals in the control group on the basis of their propensity score, then we can't construct the counterfactual for this particular subgroup and consequently we can't estimate the ATT.

Under these three assumptions the PSM estimator for ATT can be written as:

$$ATT^{PSM} = E_{P(X)|D=1}\,\{E[Y1i|D = 1,\ P(X)] - E[Y0i|D = 0,\ P(X)]\}$$

3.2.1 The validity of the three assumptions

The validity of the propensity score matching depends crucially on the three assumptions: SUTVA, CIA and Common Support. In order for the first assumption to hold, treatment status and potential outcomes of one individual must not be affected by the treatment status of other individuals. The SUTVA will be violated if there are spill-over effects. For example it is often of a concern that eligible participants might give part of the stipend to non-eligible individuals in the same municipalities. Also, in the case where more than one families live in a household and the one is registered for the program whereas the other isn't, treatment might lead to a reallocation of responsibilities within the household. Non-treated children might work more in the household and study less. However, these are not issues in our setting since we compare compliers in treated areas to eligible individuals in control areas and not to non-participants in the same municipality. So, the only way in which the SUTVA can be violated is if participants share the stipend with eligibles in the control municipalities, which in practice is very unlikely.

The second necessary assumption is the CIA. Exactly as the SUTVA, the CIA is not directly testable. It might be violated if individuals select into treatment on the basis of unobserved characteristics. Although we try to control for as many pre-treatment characteristics that are likely to affect treatment status as possible, selection into treatment might be explained as well by factors such as parental support and child's ability that are unobserved. However, even if the CIA does not hold, the propensity score matching combined with the difference-in-differences approach can still yield an unbiased estimate of the ATT, provided that the unobserved factors affecting both potential outcomes and exposure to treatment are time-invariant. We discuss this in more detail in Section 3.3.

The last key identification assumption is common support. As opposed to the previous two assumptions this is directly testable and it depends only on the covariates we include. As we will see below this is satisfied with our preferred model, we have full common support.

3.2.2 Estimating the propensity score

Propensity score matching involves three steps:

1. Model the probability of participation and estimate its value for each individual
2. Choose a matching algorithm and match individuals with similar propensity scores
3. Estimate the impact of the program on matched individuals

In the multiple treatment case (e.g. when the individual has to choose between different subsidy programs) the multinomial probit model is preferred to the multinomial logit since the former one is built on less strong assumptions (Imbens 2000 and Lechner 2001). However, in our case where treatment is binary, logit and probit yield quite similar results and the choice between them is not crucial. Thus, we use a probit model. Then, we have to choose the set of the covariates that we will include as well as the matching algorithm.

As mentioned above, the validity of matching depends on whether the CIA assumption holds. Heckman, Ichimura and Todd (1997) show that omitting explanatory variables can lead to a violation of the CIA and consequently to serious bias. This however doesn't mean that the more variables we include the better our estimates will be. There is a trade-off between bias and efficiency and as Bryson, Dorsett and Purdon (2002) prove, the inclusion of insignificant variables will increase the variance of the estimates. Thus, we want to include all covariates that affect both the probability of treatment and our two outcome variables (failure and drop-out rates).

We use kernel matching. As opposed to other matching algorithms that are frequently used in the literature (nearest neighbor and caliper) which use only one comparison individual for each treated individual, kernel uses a weighted average of individuals in the control group and consequently leads to more accurate estimates that have lower variance. It is true that all matching algorithms have advantages and disadvantages and there doesn't exist an ideal one. For example kernel might lead to bad matches if there don't exist enough comparable units in the control group. Nevertheless, there is broad consensus in the literature that when there is a substantial number of similar individuals in the control group as in our case, kernel and local-linear matching are preferred to any other algorithm.

We first match children at the baseline using three different specifications to estimate the propensity score. The results of the three probits and the 3 psgraphs are reported in Appendix B. We begin with a parsimonious specification (PSM1) that includes both individual and municipality characteristics: the geographic region, number of families with children 7-17 and families with SISBEN 1 in the municipality, age, gender and prior academic achievement of the child, parental income, age and education, the number of families living in a household, a dummy variable for single-headed households and dummy variables for whether the household has telephone and receives gas and water by pipe. The results indicate that this model leads to sizable common support (among 11,062 treated individuals 7,926 are matched); however it does not control for many important variables determining both treatment and potential outcomes. This can be also seen by the pseudo R2, which demonstrates how well the regressors explain the probability of being treated and in our case it is very low. Consequently, the quality of matching will be low. From the psgraph we can see that indeed there do not exist any control individuals for some treated ones in the right tail of the distribution whose density is quite high. Next we check whether the propensity score balances the distribution of the regressors in the treated and control groups and whether it reduces the standarised bias, as defined by Rosenbaum and Rubin (1985). The results of the pstests we conducted are reported in Tables 1 to 3 in Appendix C. Due to the large number of variables included, we report the distribution of the mean standarised bias before and after matching only for our preferred model (PSM3 as we will see). The results for PSM1 indicate that although it reduces the mean standarised bias, the difference in means between the two groups for several variables after matching is statistically significant.

In the second specification (PSM2) we drop the number of families with SISBEN 1 and families with children 7-17 in the municipality which as can be seen in Table 1 in Appendix B didn't have any effect on the propensity score and they were significantly more unbalanced after the matching than before. We add some individual characteristics related to schooling (whether the child receives lunch at school, school absence during the previous academic year and the educational level at which he would like to study). We also include two dummies for whether the household suffered a death or a serious illness and some further municipality characteristics such as teacher per pupil ratio, number of schools and monthly male and female wage. We think that controlling for these municipality-level variables might be important since wages and availability of schools and teachers are likely to affect both the treatment and the outcome variables. However, this leads to no improvement compared to our previous model. The number of treated individuals decreases (almost 40% of individuals remain unmatched) and we still don't have any common support for those on the right tail of the distribution. Moreover, the differences in means between the two groups after matching remain statistically significant for many covariates and the standarised bias doesn't decrease further.

Finally, in the third model we drop the municipality average female and male wage since as can be seen in Table 2 (Appendix B) and in Table 2 (Appendix C) they didn't have any effect on the propensity score and they were significantly more unbalanced after the matching than before. We include instead some other municipality characteristics such as the population and public infrastructure. We also include some further individual characteristics indicating the socio-economic status of the family (employment status of parents, ownership of the house, the material of which walls are constructed and whether the house has a toilet connected to the sewerage). Moreover, as the children reported as one of the most important reasons for not going to school the distance and the high transportation cost, we include dummies for the distance from school and whether the child owns a bike or a motorbike (we don't have any data on transportation cost). The advantage of this specification is that as it controls for many individual and municipality characteristics, it makes the CIA more credible. However, we are afraid that this might be achieved at the cost of lower common support. On the contrary, 9,360 among 11,062 treated individuals are now matched and the average common support increases to 0.82. Moreover, the average quality of matching increases substantially. The t-test we conducted indicates that for those on the support the difference in means is not statistically significant for most variables. Thus this propensity score matching balances well the distribution of covariates in the two groups and as can be seen in Table 3b, the mean standarised bias decreases by approximately 80%. Different matching algorithms including nearest neighbor and caliper and different bandwidths were tested but they didn't lead to any improvement in the results. Consequently, PSM3 is our preferred model and we will use it in the rest of our analysis.

Since we have two repeated cross sections and not panel data we match individuals in the second period as well on the basis of their propensity score. As this is done exactly in the same way as for the baseline described above (controlling for the same preferred set of variables and using again kernel), we won't describe the procedure again. PSM3 is again our preferred model leading to a large common support, balancing the distribution of covariates and reducing the mean standarised bias.

3.3 Difference-in-differences Propensity Score Matching

After matching we could have used a simple difference method comparing the outcome variables in the two groups after the program. Given the CIA, the propensity score matching ATT is the average difference in outcomes between the two groups in the region of common support:

$$ATT = E[Y1 - Y0|Di = 1] = E_{p(x)|D=1}\{E[Y1 - Y0|Di = 1, P(X_{it})]\} =$$

$$= E_{p(x)|D=1}\{E[Y|Di = 1, P(X_{it})] - E[Y\ |Di = 0, P(X_{it})]\}$$

This can be estimated either non-parametrically or assuming a linear specification of the following form:

$$Y_{it} = \beta_0 + \beta_1 \cdot Di + \beta_2 \cdot P(X_{it}) + \varepsilon_{it}$$

The ATT is the OLS estimate for β_1.

Allowing for heterogeneous effects:

$$Y_{it} = \beta_0 + \beta_1 \cdot Di + \beta_2 \cdot P(X_{it}) + \beta 3 \cdot [Di \cdot P(X_{it})] + \varepsilon_{it}$$

$$ATT = E_{p(x)|D=1}\{E[Yi|Di = 1, P(X_{it})] - E[Yi|Di = 0, P(X_{it})]\} = \beta_1 + \beta_2 \cdot 1/N\, \Sigma_{\,i:Di=1} P(X_{it})$$

However, matching will eliminate the bias due to systematic differences between the two groups and yield the ATT only if the CIA holds. In the language of Heckman and Robb (1985) matching assumes that selection is on observables. As we saw above however, the CIA is a very strong assumption and it might not hold in our context. In that case the matching estimates will be severely biased. This problem can be tackled by using difference-in-differences propensity score matching. As opposed to propensity score matching, the difference-in-difference allows for selection on unobservables as long as it is time invariant. Then, when taking differences over time for the same group this fixed component will cancel out. If matching is combined with difference-in-differences, then the CIA can be replaced by:

$$(Y^{t1}1i - Y^{t0}0i) \perp Di|P(X)$$

or alternatively:

$$[Y_{1i} - Y_{0i}|Di = 1, t = 1, P(X_{it})] = E[Y_{1i} - Y_{0i}|Di = 0, t = 1, P(X_{it})] \; for \; P(X_{it}) \in S$$

where S denotes the region of common support. This requires the evolution of potential outcomes to be independent of treatment status conditional on the propensity score.

Somebody would argue that instead of using matching initially we could have directly used a simple difference-in-differences:

$$Y_{it} = \alpha + \beta \cdot Di + \gamma \cdot post_t + \delta \cdot (Di \cdot post_t) + \varepsilon_{it}$$

where $post_t$ is a dummy variable equal to 1 for the period after the implementation of the program, Di is a dummy variable equal to 1 for those who received treatment and the interaction term $Di \cdot post_t$ is equal to 1 for those treated in the period after the implementation of the program.

The difference-in-differences estimate is the OLS estimate of δ:

$$\delta = \{E[Y_1|D_i = 1] - E[Y_0|D_i = 1]\} - \{E[Y_1|D_i = 0] - E[Y_0|D_i = 0]\} =$$

$$= E(Y_1 - Y_0|D_i = 1) - E(Y_1 - Y_0|D_i = 0)$$

where Y_1 now is the outcome variable in period 1 (after the intervention), Y_0 is the outcome variable in period 0 (before the intervention), $E(Y_1 - Y_0|Di = 1)$ is the average difference in the outcome variable before and after the program in the treatment group and $E(Y_1 - Y_0|D_i = 0)$ is the average difference in the outcome variable before and after the program in the control group.

Admittedly, difference-in-differences is a very popular method to estimate treatment effects when the outcome variables are observed in both groups before and after treatment. However, it might lead to serious bias if the two groups differ significantly prior to the policy. Thus, to improve on the simple difference-in-differences or the simple matching we will combine the two methods following Heckman et al (1997) and Heckman et al (1998). That is, we compare outcome variables before and after the intervention in treatment and control groups but only for the matched individuals (those on the region of common support) according to their weights.

Heckman et al (1997) use as well two repeated cross sections and extend the simple difference-in-differences described above by defining outcomes conditional on X. Then the

$$ATT = E(Y_1 - Y_0|X_{it}, D_i = 1) - E(Y_1 - Y_0|X_{it}, D_i = 0)$$

Assuming a linear specification:

$$Y_{it} = \beta_0 + \beta_1 \cdot D_i + \beta_2 \cdot post_t + \beta_3 \cdot (D_i \cdot post_t) + \beta_4 \cdot X_{it} + \varepsilon_{it}$$

where Y_{it} denotes failure (drop-out) rates of individual i in period t

The ATT is: $ATT = E(Y_1 - Y_0|X_{it}, D_i = 1) - E(Y_1 - Y_0|X_{it}, D_i = 0) = \beta_3$

Abadie (2005) instead of controlling for X, he controls for the propensity score and the ATT becomes:

$$ATT = E(Y_1 - Y_0|P(X), D_i = 1) - E(Y_1 - Y_0|P(X), D_i = 0) \;\; for\ P(X_{it}) \in S \qquad (1)$$

In equation (1), $E(Y_1 - Y_0|P(X), D_i = 0)$ can be written:

$$E(Y_1|P(X), D_i = 0) - E(Y_0|P(X), D_i = 0) =$$

$$= \Sigma_{\iota \in Co} W(P(X))\cdot(Y_1|D_i = 0) - \Sigma_{\iota \in C1} W(P(X))\cdot(Y_0|D_i = 0)$$

where Y_1 is the outcome variable in period 1 (after the intervention), Y_0 is the outcome variable in period 0 (before the intervention), C_1 is the set of control individuals in period 1, C_0 is the set of control individuals in period 0 (since we have cross-sectional and not panel data, individuals in the two periods differ and have different propensity scores) and W is the weight attached to each individual in the control group, which is a function of their propensity score. The counterfactual for each beneficiary in each period is constructed as a weighted average of the outcomes of non-beneficiaries. The weight depends on the matching algorithm that has been chosen. We used kernel matching and thus the average places higher weight on control individuals with propensity score close to the one of the treated individual and lower weight on more distant observations.

To estimate (1) we first use a linear specification of the form:

$$Y_{it} = \beta_0 + \beta_1 \cdot D_i + \beta_2 \cdot post_t + \beta_3 \cdot (D_i \cdot post_t) + \beta_4 \cdot P(X_{it}) + \varepsilon_{it} \qquad (2)$$

where $ATT = E(Y_1 - Y_0|P(X), D_i = 1) - E(Y_1 - Y_0|P(X), D_i = 0) = \beta_3$

Then we want to allow for heterogeneous effects (treatment might have different effects on individuals depending on their propensity score). Thus, we include the interaction between the propensity score, the treatment variable and the dummy variable for the post-program period and estimate:

$$Y_{it} = \beta_0 + \beta_1 \cdot D_i + \beta_2 \cdot post_t + \beta_3 \cdot (D_i \cdot post_t) + \beta_4 \cdot P(X_{it}) + \beta_5 \cdot [post \cdot D_i \cdot P(X_{it})] + \varepsilon_{it} \quad (3)$$

The $ATT = E(Y_1 - Y_0|P(X), D_i = 1) - E(Y_1 - Y_0|P(X), D_i = 0) = \beta_3 + \beta_5 \cdot 1/N\, \Sigma_{i:Di=1}\, P(X_{it})$ and the variance of the ATT is: $Var\ (ATT) = Var\ (\beta_3) + 1/N\ \Sigma_{i:Di=1}\ P(X_{it})^2 \cdot Var\ (\beta_5) + 2 \cdot Cov\ (\beta 3, \beta_5)$.

After estimating β_3 and β_5 and bootstrapping to correct for the fact that the propensity score is estimated and it is not the true value in the population, we compute the ATT and its standard error as the square root of the variance defined above.

However, the assumption of linearity is too restrictive. There doesn't exist any theoretical justification for a linear relationship between failure/drop-out rates and treatment and thus in order to avoid bias due to misspecification we estimate equation (1) non-parametrically using a difference-in-differences propensity score matching approach. As opposed to (2) and (3) which impose a linearity assumption and can be considered as semi-parametric methods, this fully non-parametric approach is more flexible. However, all three estimations yield quite similar results as we will see in Section 4.

In all cases the propensity score matching will alleviate the bias due to systematic differences between the two groups. Heckman et al (1997) show that this combined method performs almost as well as randomization. Moreover, Smith and Todd (2004) find that it performs much better than simple matching or simple difference-in-differences when treated and control individuals belong to different geographic regions as in our case. However, in order for the estimate to be unbiased the key identification assumption is the "parallel trend" assumption as mentioned above. This means that in the absence of the program the average change in the outcome variables would have been the same between the two groups (same time trends):

$$[Y_{1i} - Y_{0i}|D_i = 1, t = 1, P(X_{it})] = E[Y_{1i} - Y_{0i}|D_i = 0, t = 1, P(X_{it})] \text{ for } P(X_{it}) \in S$$

Intuitively, the outcome variables in the control group must have evolved over time in the same way that they would have evolved in the treatment group had they not been treated.

4 RESULTS

In this Section we present the estimates of the Average Treatment effect on the Treated of Familias en Acción on failure and school drop-out rates. It is worth mentioning again that in all specifications we compare participants in the treatment groups to eligibles in the control groups. The results from the difference-in-differences propensity score matching are reported in Appendix D. Table 1 shows the effects of the program on failure rates from estimating equation (2). In column 1 we run the regression only for the matched (those who lie within the common support using our preferred propensity score matching model 3). Thus, the estimate of the coefficient for the interaction term can be interpreted as the Average Treatment effect on the Treated. We can see that on average the program had a large and statistically significant effect on the failure rates of the matched treated children: they decreased by almost 6 percentage points. In the next column we run the same regression for the whole sample (not only for those on support) without controlling for the propensity score. This would be the estimate from a simple difference-in-differences, not being combined with matching. The impact is the same qualitatively but smaller in absolute value. This estimate however is biased since as can be seen from the pstest in Table 3 in Appendix B, the distributions of some covariates of treated and control individuals who are off support differ significantly. This underlines the importance of comparing treated individuals to similar controls, otherwise the results can be misleading. In our case the bias is small since the treatment municipalities are a stratified random sample of municipalities where the program is operating and controls were chosen from the same stratum to be as similar as possible to treatment ones.

To take into account the heterogeneity of program effects by gender we continue to restrict our analysis to matched individuals only but estimate (2) for females and males separately. The results are reported in columns 3 and 4 respectively. Both estimates are negative and similar in magnitude, however only the effect for boys is statistically significant. The program reduced failure rates of boys by 6.3 percentage points. In Table 1b we perform the same analysis by location (urban or rural) and by age. Column 1 shows the estimates for matched individuals in urban areas while column 2 shows the equivalent estimates in rural areas. The results are striking. The large, negative average effect of Familias en Acción on failure rates is mainly driven by the effect in urban areas. In particularly, the program led to a decrease of failure rates in urban areas of 8.5 percentage points and to a decrease in rural areas by 3.3 percentage points. However, only the former effect is statistically significant. This can be attributed to the fact that school enrollment and academic performance were substantially lower prior to the intervention in rural areas. As Attanasio et al (2005) demonstrate, not even 50% of 12-17-year-olds in rural areas would have attended school in the absence of the program. When these children are induced into education they are expected to fail without any educational assistance. Finally, in columns 3 and 4 we run the regression separately for 7-13 and 14-17 year olds. These two groups are chosen on the basis of the dramatic reduction in school enrolment in Colombia at age 14 (Attanasio et al 2008). The estimates suggest that the program had no impact on failure rates of older children whereas it reduced failure rates of young children by almost 7 percentage points. In sum, when matching approximately 82% of treated with control individuals we find evidence that the students benefited considerably from treatment assignment in terms of better academic performance.

The results for drop-out rates from estimating equation (2) are reported in Tables 2a and 2b. The Average Treatment effect on the Treated (ATT) for those matched is very small and negative suggesting that on average the program decreased the drop-out rates of the recipients by almost 1 percentage point. If we don't control for the propensity score and don't restrict our sample to those on support only but examine the whole sample, we can see from column 2 that in contrast to what we observed for the failure rates, the results now do not differ considerably. The simple difference-in-differences estimate is almost 0.01 again although now it is statistically significant. Columns 3 and 4 contain the results for females and males respectively. We can see that the impact on drop-out rates is larger for girls although neither of the two effects is statistically significant. Familias en Acción reduced the drop-out rates of girls by 2 percentage points whereas it reduced the drop-out rates of boys by only 0.5 percentage points. This can be attributed to the fact that girls do mainly domestic work which is more flexible and also a reallocation of responsibilities within the household is likely to have taken place in order for them to attend school. Barrera-Osorio et al (2007) find spillover effects on the siblings of the treated children. In particular they find that in families who received the subsidy 93% of registered children attended school whereas only 75% of unregistered children in the same municipality did. Boys work mainly outside the house which is less flexible. Table 2b shows the distribution of the program effects by location and age. As in the case of failure rates, the results for drop-out rates indicate that the effect is larger for urban areas although neither of the two effects is statistically significant. More precisely, the program reduced drop-out rates of matched individuals in urban areas by almost 2 percentage points whereas it reduced them in rural areas by only 1 percentage point. Finally, from columns 3 and 4 we can see that the program led to a decrease in drop-out rates of 14-17 year old children of 6 percentage points whereas it had no impact on the drop-out rates of younger (7-13) children. A possible explanation for that is that as expected the drop-out rates of younger children prior to the intervention were very low (the probability that a child 7 years old would drop out from school was 1.5% whereas the equivalent for a child 15 years old was 4.2%) and consequently there was little room for improvement. In sum, although the subsidy increased considerably academic performance it had only a small effect on drop-out rates.

To allow for heterogeneous effects we next estimate by OLS equation (3). Table 3 contains the estimates from this regression as well as their bootstrapped standard errors. We can see that the results are similar qualitatively just larger in absolute value than those with constant effects. These results indicate that it is important who is treated: we find that individuals with high propensity scores benefited more from the program than those with low propensity scores in terms of reduced failure and drop-out rates. We also run the same regression but instead of conditioning on the propensity score, we condition on all the covariates X we used to estimate it. As expected the results are very similar and we don't report them.

Finally, we implement the difference-in-differences propensity score matching non-parametrically and again estimate the standard errors using bootstrap to correct for the fact that the propensity score has been estimated and is not the true value in the population. The results are reported in Table 4 and serve as a comparison to those obtained from the parametric estimation. We can see again that the program had a large negative impact on average failure rates of matched individuals and a small negative impact on their drop-out rates. These results are totally consistent with those reported above.

In conclusion, the above results are very important because they imply that despite their lower education level those who are induced into education by the subsidy do not find the courses extremely difficult or beyond their capacity, perform equally well or even better than students already enrolled and drop out from school less. Moreover, they indicate that the stipend was enough to cover the opportunity cost of education.

5 ROBUSTNESS ANALYSIS

Propensity score matching eliminates the selection bias under the common support assumption and the assumption that there do not exist unobservable characteristics that affect both treatment status and potential outcomes. The propensity score graphs in Appendix B allow us to test the common support assumption, whether there exist non treated for the treated at any value of the propensity score. We can see that for our chosen model (model 3) the region of common support is quite large, approximately 82% of treated individuals have a suitable match among the set of controls. So the first necessary assumption is met. However, individuals are likely to select into treatment on the basis of unobserved characteristics even though we used a large number of detailed pretreatment characteristics at the individual and municipality level to predict treatment. Since we can't directly test the CIA we use a difference-in-differences method that will yield unbiased estimates even if the CIA doesn't hold under the assumption of similar time trends between treatment and control groups. Unfortunately, we can't test this since we don't have data for the outcome variables (failure and dropout rates) in the two groups for more than one periods prior to the program. Nevertheless, we have data for more than one pre-program periods for some other variables (e.g. school attendance) affecting both the outcome variables and the treatment status. From Table 1 in Appendix E we can see that the evolution of school attendance in the pre-program period is similar in treatment and control groups (the estimates for the coefficients of the two interactions are not statistically significant). Although this is not sufficient, it is at least not inconsistent with the parallel trend assumption.

Another concern is that the school supply (quantity and quality of education) might differ significantly over time between treatment and control groups and this will in turn affect our outcome variables. Thus we conduct a t-test for the differences in average changes in several school inputs over time between the two groups. The results are reported in Table 2 in Appendix E and with the exception of the school having a library, we find no statistically significant differences over time between treatment and control regions. Finally, another important issue when implementing difference-in-differences is the known as “Ashenfelter dip”. Many papers evaluating job training programs in the US compared earnings of treated and non-treated individuals before and after the program. However, Ashenfelter's (1978) criticism is that participants' earnings fall substantially just before entering the program, which is precisely the reason why they were offered the program. In our example this means that eligible students are most likely those having high failure and drop-out rates just before the commencement of Familias en Acción and not all of these are likely to be permanent. For example, a child that would usually attend school and perform well might dropped out the previous year because his father lost his job because of the economic crisis. However this child will work temporarily and then return to school and perform well again when his father finds a job. In that case, a difference-in-differences estimate that doesn't take into account this pre-program dip will be upward biased. However, this isn't a concern in our case since the program targeted schools and not individuals and the failure and drop-out rates were measured at school and not at individual level.

As mentioned above, in some municipalities (TCPs) the program had already started during the baseline survey and even in those that it hadn't started (TSPs), it was publicly known and registration had begun. Anticipation effects will lead to biased estimates since our outcomes variables in treatment groups prior to the program are already affected and thus not representative. We deal with this problem by excluding failure and drop-out rates in 2002 and use as our baseline data from one year before the announcement of the program. Moreover, the existence of municipalities where the program started earlier not only doesn't constitute a problem for our evaluation but also it enables us to do a robustness check by comparing TCPs and TSPs in the pre-baseline and in the baseline survey (academic year 2002) when only the former ones were treated. The results from this check will serve as a comparison to our estimates. The choice of the municipalities in which the program started earlier was completely random, determined only by the order in which the names of the municipalities were written in the official papers (Attanasio et al 2008). Also, these two groups are likely to be significantly more similar than the treatment and control groups we examined. The propensity score now is the probability that the individual lives in a TCP municipality and has enrolled to the program. The results from the propensity score matching are reported in Tables 3 and 4 in Appendix E. We use our preferred model (model 3) and among the 7,755 early treated 7,553 (almost all) are on support. The pstest suggests that with the exception of the region, for all the other covariates the differences between the two groups after matching are not statistically significant and this confirms our hypothesis that TCP and TSP municipalities are very similar.

After having matched the early treated (TCP) to late-treated individuals (TSP), we implement the difference-in-differences and our specification becomes:

$$Y_{it} = \beta_0 + \beta_1 \cdot TCP + \beta_2 \cdot post_t + \beta_3 \cdot (TCP \cdot post_t) + \beta_4 \cdot P(X_{it}) + \varepsilon_{it}$$

where Y_{it} denotes failure (drop-out) rates of individual i in period t, $post_t$ is a dummy variable equal to 1 for the baseline survey (2002) when the program had commenced in the TCPs but not in the TSPs, *TCP* is a dummy variable equal to 1 if the region was an early-treated and 0 otherwise and the interaction term $TCP \cdot post_t$ is equal to 1 for the early treated at the baseline survey.

The difference-in-differences estimate is the OLS estimate for β3. The results are reported in Tables 5a and 5b and serve as a comparison for our obtained estimates reported in Appendix D. Having restricted again our sample to those on support only, we can see that the estimates for failure rates are the same qualitatively but smaller in absolute value whereas the estimates for drop-out rates are the same qualitatively but slightly larger in absolute value. Overall, the program reduced grade failure rates by almost 3 percentage points and school drop-out rates by 1.2 percentage points (column 1). When comparing treatment effects between females and males (columns 2 and 3 respectively), we find that the early treated girls had significantly lower grade drop-out rates whereas the effect for boys is almost zero. Finally, the program led to a decrease of almost 6 percentage points in failure rates of boys and to a decrease of 2 percentage points for girls. However, only the former effect is statistically significant. The above results are totally consistent with those we find when comparing treatment to control municipalities.

Finally, the internal validity will be undermined if the estimates for the ATT capture the effects of other contemporaneous programs. The only major contemporaneous welfare programs were Hogares Comunitarios, which has been in operation since 1986 and the Breakfast Program which was launched in 2003. However, the former is targeted only to children between 0 and 6 years old and mothers having children in both age categories had to choose between Hogares Comunitarios and

Familias en Acción, they couldn't enroll in both. The latter is also targeted to younger children but it operates only in municipalities not targeted by Familias en Acción. Therefore, our estimates can't be affected by these two programs.

6 CONCLUSION

Despite the large existing literature reporting significant positive effects of Familias en Acción on enrolment, gaps in knowledge exist as to whether these effects were combined with positive effects in grade progression and negative effects in school drop-out rates. These gaps motivated us to conduct this analysis. To deal with non random assignment and low compliance we use a difference-in-differences propensity score matching comparing participants in treated municipalities to eligible individuals in control municipalities. The main advantage (and drawback) of propensity score matching relies on the credibility of the CIA. If selection bias from unobserved characteristics is likely to be negligible, then propensity score matching can yield as good estimates as a randomized experiment. But if the two groups differ on the basis of unobserved characteristics that affect failure and drop-out rates, then a simple propensity score matching can lead to serious bias. Therefore, we combine difference-in-differences with matching by imposing a common support condition on the difference-in-differences. This guarantees that for each treated individual in the region of common support there exists a similar/comparable individual in the control group and therefore the estimated effects of the program are not a consequence of pre-program differences between the two groups. We find that the estimates from a simple difference-in-differences are significantly smaller in absolute value and this bias justifies our choice for the combination of matching with difference-in-differences.

The results indicate that treated children are less likely to fail and to drop out from school than similar non-treated children. Overall, the program reduced average drop-out rates by approximately 1 percentage point. The effects are larger for urban areas and for girls, whose drop-out rates decreased by 2 and by between 2 and 4 percentage points, respectively compared to the decrease of 1 percentage point in the drop-out rates of both children in rural areas and boys. However, only the effect for girls is statistically significant. As expected, the impact is larger for older children (14-17) whose drop-out rates decreased by 6 percentage points compared to the fall of only 0.3 percentage points in the drop-out rates for younger children (7-13). Of course, this can be explained by the fact that drop-out rates for younger children were very low prior to the policy. The effect on failure rates is larger, in the region of 6 to 7 percentage points but continues to vary substantially across different groups. Contrary to the estimated effects on drop-out rates, the effects on failure rates for boys are larger compared to those on girls. The estimates are also largest for children in urban areas and for younger children (7 to 13 years old) whose failure rates decreased by 8.5 and almost 7 percentage points, respectively. The fact that the main findings of this paper are stable across different methods and specifications is reassuring and indicates the robustness of our results.

There are at least two effects that can account for the significant decrease in failure and drop-out rates. The first is an income effect arising from the stipend which increases the family's income and enables children to attend school instead of working. The second is a substitution effect since by default the program design reduces the relative price of education and incentivizes children to substitute education for work. In addition, Familias en Acción is expected to improve academic performance through the positive effects on nutrition and health status which increase human capital. However, as mentioned in Section 1 there are also other factors operating in the opposite direction that are likely to prevent the effects on drop-out rates from being larger in magnitude.

Familias en Acción has been expanding significantly over time: it started by targeting rural municipalities and today it has achieved national coverage with approximately 3.8 million recipient children. Its cost is relatively low, constituting 0.27% of Colombian GDP. A cost-benefit analysis implemented by the IFS et al (2006) suggests that the program will increase wages and will have benefits that are 1.59 times larger than its total cost. Our analysis indicates that the program was moderately successful in its main target: foster human capital accumulation. An extrapolation to the current recipient population suggests that on average 219,640 more children would pass the grade and 44,840 less children would drop out from school due to Familias en Acción. This increases the estimated benefits and consequently the cost-effectiveness of the program. However, the lack of evidence indicating any significant effect on failure and drop-out rates in rural areas as well as the small magnitude of the overall negative effect on drop-out rates raise issues concerning changes in the program design to increase efficiency. A possible solution is to make CCTs conditional on academic performance and school completion rather than on simply enrolment (e.g. in Progresa children lose eligibility if they fail the grade) and at the same time intervene in the supply-side of education e.g. provide these children with educational assistance. Conditionality based on graduation rather than simply on attendance has been proven to be particularly effective, leading to even higher levels of enrolment and academic performance (Barrera-Osorio et al 2008).

REFERENCES

A. ABADIEU (2005): “Semi-parametric Difference-in-Differences Estimators”, *Review of Economic Studies 72, 1–19*

M. ANGELUCCI AND O. ATTANASIO (): “Estimating ATT effects with non-experimental data and low compliance.”, *unpublished manuscript, Department of Economics, University College London*

M. ANGELUCCI AND O. ATTANASIO (2009): “Oportunidades: program effect on consumption, low participation and methodological issues”, *Economic Development and Cultural Change, 57:3, 479-506*

J. ANGRIST, E. BEETINGER, E. BLOOM, E. KING, M. KREMER (2002*)*: “Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment”, *the American Economic Association, vol. 9(5), pages 1535-1558*

J. ANGRIST AND J. PISCHKE (2009): “Mostly harmless econometrics”, *Princeton, NJ*

O. ATTANASIO AND A. MESNARD (2006): “The impact of a Conditional Cash Transfer Programme on Consumption in Colombia”, *IFS (2006) vol.27, Issue 4, pages 421-442*

O. ATTANASIO, E. FITZSIMOND, A. GOMEZ, D. LOPEZ, C. MEGHIR AND A. MESNARD (2006): “Child education and work choices in the presence of a conditional cash transfer program in rural Colombia”, *Economic Development and Cultural Change, vol.58, No. 2, pp 181-210, University of Chicago Press, Journal Articles*

O. ATTANASIO, C. MEGHIR AND M.V. HERNANDEZ (2004): “Baseline report on the evaluation of Familias en Acción”, *Institute for Fiscal Studies Report*

F. BARRERA-OSORIO, L. BERTRAND, L. LINDEN AND F. PEREZ-CALLE (2008): “Conditional Cash Transfers in Education Design Features, Peer and Siblings effects, Evidence from a Randomized Experiment in Colombia", *NBER Working Papers 13890, National Bureau of Economic Research, Inc*

J. BEHRMAN, S. PARKER AND P. E. TODD (2008): "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico", *Economic Development and Cultural Change, vol. 57, Issue 3, pp 439-477,University of Chicago Press, Journal Articles*

R. BLUNDELL AND M. COSTA DIAS (2000): "Evaluation Methods for Non-Experimental Data", *IFS (2000) vol. 21, no. 4, pp. 427–468*

M. CALIENDO AND S. KOPEINIG (2008): "Some practical guidance for the implementation of propensity score matching", *IZA Discussion Paper No. 1588*

D. CARD AND A. KRUEGER (1994): "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania", *The American Economic Review, vol. 84, n. 4, pp. 772-793*

J. DAS, D. QUY TOAN AND B. OZLER (2005): "Reassessing Conditional Cash Transfer Programs", *World Bank Observer 20(1): 57-80*

L. DEARDEN, C. EMMERSON, C. FRAYNE AND C. MEGHIR (2009): "Conditional Cash Tranfers and School Dropout Rates", *The Review of Economic Studies August 2, 2011 0:rdr015v1-rdr015*

L. DEARDEN, C. EMMERSON, C. FRAYNE AND C. MEGHIR (2005): "Education Subsidies and School Dropout Rates", *IFS WPO5/11*

R.H. DEHEJLA AND S. WAHBA (1998): "Propensity score-matching methods for non-experimental causal studies", *NBER Working Paper No. 6829*

DUBOIS, DE JANVRY AND SADOULET (2004): "Effects on School Enrollment and Performance of a Conditional Transfers Program in Mexico", *IDEI Working Papers 114, Institut d' Économie Industrielle (IDEI), Toulouse*

E. DUFLO (2002): “Empirical methods”, *Handout of MIT courses (2002), 14.*

J. HECKMAN, H. ICHIMURA, J. SMITH AND P. TODD (1998): “Characterizing Selection Bias Using Experimental data”, *Econometrica, Econometric Society, vol. 66(5), pages 1017-1098*

J. HECKMAN, H. ICHIMURA AND P. TODD (1997): “Matching as an Econometric Evaluation Estimator”, *Review of Economic Studies 65, pp. 261-294*

K. HIRANO, G. IMBENS AND G. RIDDER (2003): “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score” *Econometrica, 71(4): 1161-1189*

P. J. McEWAN (1998): “The effectiveness of multigrade schools in Colombia”, *International Journal of Educational Development, Volume 18, Number 6, 1998 , pp. 435-452(18)*

B. D. MEYER (1995): “Natural and Quasi-Experiments in Economics”, *Journal of Business & Economic Statistics, 13, pp. 151–161*

R. MORENO-SERRA (2009): “Health Programme Evaluation by Propensity Score Matching: Accounting for Treatment Intensity and Health Externalities with an Application to Brazil”, *HEDG Working Paper 09/05*

L. RAWLINGS AND G. RUBIO (2005): “Evaluating the impact of Conditional Cash Transfer Programs”, *World Bank Research Observer 20(1): 29-55*

P. R. ROSENBAUM AND D. B. RUBIN (1985): “Constructing a control group using multivariate matched sampling methods that incorporate the propensity score”, *the American Statistician, Volume 39, No. 1*

P. R. ROSENBAUM AND D. B. RUBIN (1983): “The central role of the propensity score in observational studies for causal effects”, *Biometrika 70 (1), 41-55*

T. P. SCHULTZ (2004): “School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program”, *Journal of Development Economics 74(1): 199-250*

F. V. SOARES, R. P. RIBAS, R. G. OSORIO (2010): “Evaluating the Impact of Brazil's Bolsa Família: Cash Transfer Programs in Comparative Perspective”, *Latin American Research Review - Volume 45, Number 2, pp. 173-190*

P. E. TODD (2006): “Evaluating social programs with endogenous program placement and selection of treated”, *Handbook of Agricultural Economics, Volume 4, North Holland*

APPENDIX A: STATISTICS

Table 1: Average Characteristics of Treatment and Control Municipalities at Baseline Survey

	Treatment	Control	Total
Monthly labour income	227387.7	272524.2	285427.5
	(451286.0)	(362367.4)	(385794.2)
Monthly non-labour income	1108.1	3044.5	2589.0
	(14655.5)	(24413.3)	(22516.1)
Proportion of households with 1 family	0.782	0.769	0.772
	(0.413)	(0.422)	(0.420)
Proportion of households with 2 families	0.179	0.177	0.177
	(0.383)	(0.382)	(0.382)
Proportion of households with 3 or more families	0.0392	0.0542	0.0507
	(0.194)	(0.226)	(0.219)
Proportion of mothers with completed primary education	0.891	0.907	0.903
	(0.119)	(0.139)	(0.135)
Proportion of mothers with completed secondary education	0.407	0.399	0.401
	(0.229)	(0.274)	(0.264)
Price of fish	5325.5	5205.9	5234.0
	(1275.7)	(2022.1)	(1874.0)
Price of chicken	4719.7	4381.4	4460.9
	(758.4)	(995.6)	(955.9)
Price of beef	6517.7	6193.1	6264.4
	(772.5)	(618.9)	(669.3)
Price of rice	1392.2	1239.5	1275.4
	(101.4)	(274.0)	(253.0)

	Treatment	Control	Total
Urban hourly wage	1214.3	1236.2	1231.0
	(159.2)	(405.8)	(363.3)
Rural hourly wage	1219.2	1222.6	1221.8
	(243.4)	(331.7)	(313.2)
Population in 2000	27712.8	30285.9	29680.7
	(12893.2)	(20613.8)	(19111.3)
Families with SISBEN1	1063.2	628.5	730.8
	(926.3)	(1490.0)	(1390.5)
Families with children 0-17	768.9	424.6	505.5
	(664.1)	(919.0)	(878.0)
Class m^2 per student	2.269	2.989	2.820
	(0.460)	(2.741)	(2.427)
Number of students per teacher	25.00	22.20	22.86
	(4.734)	(3.866)	(4.256)
Number of hospitals	0.879	0.866	0.869
	(0.326)	(0.341)	(0.338)
Number of public hospitals	0.927	0.826	0.849
	(0.260)	(0.379)	(0.358)
Number of rural public schools	38.71	50.10	47.42
	(21.99)	(25.86)	(25.46)
Number of urban public schools	6.968	9.323	8.770
	(4.864)	(8.276)	(7.677)
Number of banks	0.768	1.867	1.608
	(2.124)	(2.432)	(2.408)

	Treatment	Control	Total
Amount of tax on business	280672927.5	185873423.4	225022419.3
	(715093746.6)	(438598947.0)	(571192779.4)
Proportion of toilets connected to sewerage	0.491	0.557	0.542
	(0.186)	(0.144)	(0.158)
Proportion of households with piped water	0.587	0.556	0.563
	(0.249)	(0.179)	(0.198

Mean coefficients; standard deviations in parentheses

APPENDIX B: RESULTS FROM PROBIT REGRESSIONS AND PSGRAPHS

Table 1: Probit Regression (Model 1)

Number of obs	= 15438
LR chi2(29)	= 1632.96
Prob > chi2	= 0.0000
Pseudo R2	= 0.0789

Log likelihood = -9528.3891

Treat	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
region	.0499085	.0120099	4.16	0.000	.0263694	.0734475
age	-.0032109	.0054598	-0.59	0.556	-.0139118	.0074901
age_head	.0000974	.0016402	0.06	0.953	-.0031173	.0033121
age_spouse	-.0026358	.0018181	-1.45	0.147	-.0061992	.0009275
single	.1229206	.0304647	4.03	0.000	.0632108	.1826303
ed_level_1	.0074409	.0178691	0.42	0.677	-.0275818	.0424637
edu_h2	.040971	.0284094	1.44	0.149	-.0147104	.0966523
edu_h3	.053553	.0380432	1.41	0.159	-.0210103	.1281163
edu_h4	-.0076051	.0454632	-0.17	0.867	-.0967113	.0815012
edu_h5	-.1502661	.0658917	-2.28	0.023	-.2794114	-.0211208
edu_s2	.0590787	.0299789	1.97	0.049	.0003212	.1178363
edu_s3	-.0323172	.0380328	-0.85	0.395	-.1068601	.0422257
edu_s4	-.0972124	.0462305	-2.10	0.035	-.1878225	-.0066023
edu_s5	-.1527141	.0652806	-2.34	0.019	-.2806617	-.0247665
pershog	-.0045554	.0055677	-0.82	0.413	-.0154679	.0063571
house	.1256865	.0658453	1.91	0.056	-.0033679	.254741
gasbypipe	.1141745	.0417187	2.74	0.006	.0324074	.1959416
waterbypipe	-.0164625	.0231998	-0.71	0.478	-.0619333	.0290083
phone_3	.1257952	.0379253	3.32	0.001	.0514629	.2001275
houseown_2	.1183625	.0429438	2.76	0.006	.0341943	.2025308
houseown_1	.1005248	.0255412	3.94	0.000	.0504648	.1505847
no_hos_alc	.2885034	.0275406	10.48	0.000	.2345249	.3424819
no_colurb_~c	-.0011314	.0013789	-0.82	0.412	-.0038341	.0015712
no_colrur_~c	.0132386	.0004372	30.28	0.000	.0123817	.0140955
fam017d	-.0018396	.0001614	-11.40	0.000	-.002156	-.0015233
famsis1d	.0011987	.0001113	10.77	0.000	.0009806	.0014167
incm_lab_h	-.0002967	.000043	-6.91	0.000	-.0003809	-.0002125
nucleos_1	-.1517185	.060574	-2.50	0.012	-.2704415	-.0329956
nucleos_2	-.208811	.0598688	-3.49	0.000	-.3261517	-.0914702
_cons	-.1683658	.1431202	-1.18	0.239	-.4488763	.1121446

Table 2: Probit Regression (Model 2)

Number of obs	=	15438
LR chi2(32)	=	1508.46
Prob > chi2	=	0.0000
Pseudo R2	=	0.0729

Log likelihood = -9591.1376

Treat	Coef.	Std. Err.	Z	P>|z|	[95% Conf.	Interval]
region	.0534453	.0122003	4.38	0.000	.0295331	.0773575
age	.0065272	.0060626	1.08	0.282	-.0053553	.0184096
single	.0958288	.028097	3.41	0.001	.0407598	.1508979
ed_level_1	-.0111317	.0194826	-0.57	0.568	-.049317	.0270536
edu_h2	.031395	.0281113	1.12	0.264	-.0237021	.0864921
edu_h3	.0524384	.0374556	1.40	0.162	-.0209731	.12585
edu_h4	.0014665	.0447031	0.03	0.974	-.0861499	.0890829
edu_h5	-.1404628	.0650681	-2.16	0.031	-.267994	-.0129316
edu_s2	.0578292	.0295341	1.96	0.050	-.0000567	.1157151
edu_s3	-.0171346	.0373598	-0.46	0.646	-.0903585	.0560893
edu_s4	-.0973101	.0452917	-2.15	0.032	-.1860801	-.0085401
edu_s5	-.1202918	.0643128	-1.87	0.061	-.2463426	.005759
gasbypipe	.153678	.0413033	3.72	0.000	.0727251	.2346309
waterbypipe	-.0016836	.023057	-0.07	0.942	-.0468746	.0435073
phone_3	.1030734	.0377575	2.73	0.006	.02907	.1770768
houseown_2	.1084986	.0424613	2.56	0.011	.025275	.1917212
houseown_1	.0947103	.0251766	3.76	0.000	.045365	.1440555
no_colurb_~c	.00364	.0013112	2.78	0.006	.0010701	.0062098
no_colrur_~c	.0129407	.0004557	28.40	0.000	.0120477	.0138338
nucleos_1	-.1432455	.0528218	-2.71	0.007	-.2467743	-.0397167
nucleos_2	-.2092261	.0572214	-3.66	0.000	-.321378	-.0970742
dum_death	.1101566	.0527697	2.09	0.037	.00673	.2135832
dum_ill01	.0730454	.0491635	1.49	0.137	-.0233133	.1694041
b_Lunch	-.13872	.0372405	-3.72	0.000	-.2117099	-.0657301
b_BuyBks	.0370829	.0313053	1.18	0.236	-.0242744	.0984402
abs_JanDec01	-.1754605	.0415589	-4.22	0.000	-.2569145	-.0940064
EdLevWant	-.1465586	.0183011	-8.01	0.000	-.182428	-.1106892
alumxpro	-.0083507	.0021123	-3.95	0.000	-.0124907	-.0042107
n_pagos	.1350755	.0262343	5.15	0.000	.0836573	.1864937
wage_male_m	-.0000601	.000036	-1.67	0.094	-.0001306	.0000104
wage_femal~m	-.0000258	.0000282	-0.91	0.361	-.0000812	.0000295
age_ent	-.0289426	.0088856	-3.26	0.001	-.046358	-.0115272
_cons	.4512273	.146584	3.08	0.002	.163928	.7385267

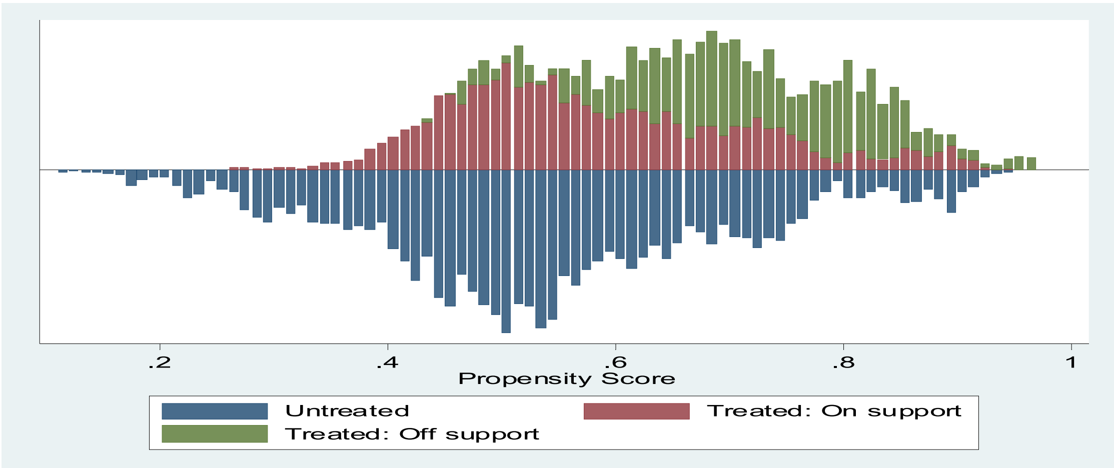
Table 3a: Probit Regression (Model 3)

		Number of obs	= 15438
		LR chi2(44)	= 489.55
		Prob > chi2	= 0.0000
Log likelihood	= -577.11373	Pseudo R2	= 0.4978

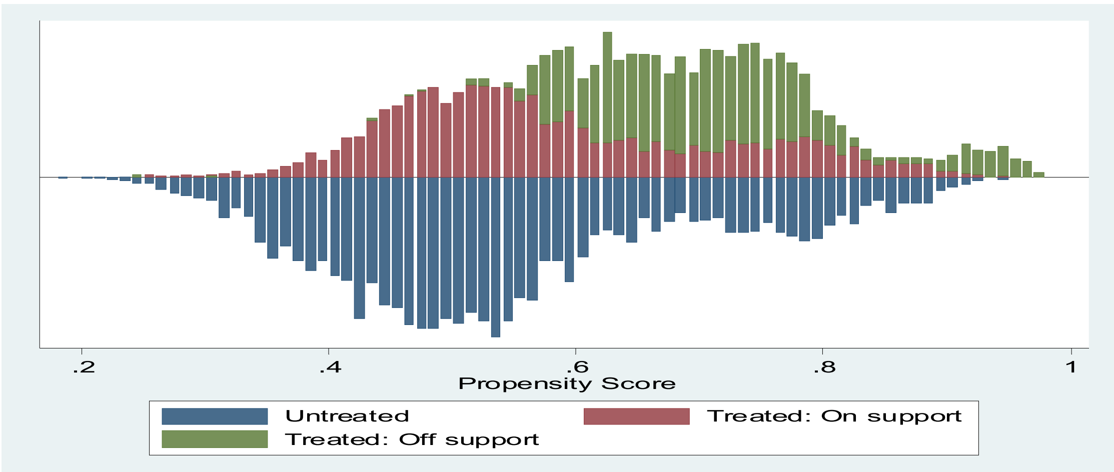
Treat	Coef.	Std. Err	z	P>|z|	[95% Conf.	Interval]
region	.0285028	.0614992	0.46	0.643	-.0920334	.149039
age	-.0226028	.0244502	-0.92	0.355	-.0705242	.0253187
age_head	-.0026436	.0070752	-0.37	0.709	-.0165108	.0112236
age_spouse	-.0030409	.0076502	-0.40	0.691	-.0180351	.0119533
single	.0445948	.1266855	0.35	0.725	-.2037041	.2928938
ed_level_1	.0498963	.0827932	0.60	0.547	-.1123753	.2121679
edu_h2	.0764273	.1328591	0.58	0.565	-.1839717	.3368263
edu_h3	-.2008753	.1578879	-1.27	0.203	-.5103299	.1085793
edu_h4	.1403083	.1750972	0.80	0.423	-.2028759	.4834925
edu_h5	-.267374	.242077	-1.10	0.269	-.7418362	.2070883
edu_s2	.3367958	.134966	2.50	0.013	.0722673	.6013243
edu_s3	.1046208	.1572468	0.67	0.506	-.2035772	.4128188
edu_s4	.0490298	.1739833	0.28	0.778	-.2919712	.3900308
edu_s5	.3638487	.2497544	1.46	0.145	-.125661	.8533584
pershog	.0263069	.0272075	0.97	0.334	-.0270188	.0796327
house	.0163445	.225942	0.07	0.942	-.4264937	.4591828
gasbypipe	-.1938684	.1569731	-1.24	0.217	-.5015301	.1137932
waterbypipe	-.1456636	.1039639	-1.40	0.161	-.3494291	.058102
phone_3	-.0626681	.1591346	-0.39	0.694	-.3745661	.2492299
houseown_2	-.2194103	.1636666	-1.34	0.180	-.5401909	.1013704
houseown_1	.3091845	.1048412	2.95	0.003	.1036995	.5146694
no_hos_alc	.9220238	.1208828	7.63	0.000	.6850979	1.15895
no_colurb_~c	.0220753	.0066268	3.33	0.001	.0090869	.0350636
no_colrur_~c	.0091058	.0021221	4.29	0.000	.0049466	.013265
fam017d	-.0022514	.000654	-3.44	0.001	-.0035332	-.0009697
famsis1d	.0014923	.0004516	3.30	0.001	.0006073	.0023774
incm_lab_h	-.0000976	.0001993	-0.49	0.624	-.0004882	.000293
price_rice	-.0025877	.0003315	-7.81	0.000	-.0032374	-.0019381
price_chic~n	.0002411	.0000917	2.63	0.009	.0000613	.0004209
nucleos_2	-.0993326	.1388392	-0.72	0.474	-.3714525	.1727873
nucleos_3	.0432772	.2427994	0.18	0.859	-.432601	.5191553
dum_death	.0868354	.2078691	0.42	0.676	-.3205806	.4942513
dum_ill01	.4606662	.1991276	2.31	0.021	.0703834	.8509491
b_Lunch	-.3611013	.0953077	-3.79	0.000	-.5479008	-.1743017

Treat	Coef.	Std. Err	z	P>|z|	[95% Conf. Interval]
b_BuyBks	.0533333	.0589234	0.91	0.365-.0621544	.1688209
abs_JanDec01	-.271835	.2015416	-1.35	0.177-.6668493	.1231793
EdLevWant	-.3092918	.0830567	-3.72	0.000-.4720799	-.1465036
alumxpro	-.0427737	.0102991	-4.15	0.000-.0629595	-.0225878
edu3	-.2984332	.3490878	-0.85	0.393-.9826326	.3857663
walls_mate~1	.070287	.3103518	0.23	0.821-.5379913	.6785653
walls_mate~2	.302572	.3190867	0.95	0.343-.3228265	.9279704
walls_mate~3	-.3226968	.3249515	-0.99	0.321-.95959	.3141965
walls_mate~4	.4931764	.3955041	1.25	0.212-.2819973	1.26835
motorbike	1.674063	.3062358	5.47	0.0001.073852	2.274274
_cons	3.833849	.8953039	4.28	0.0002.079085	5.588612

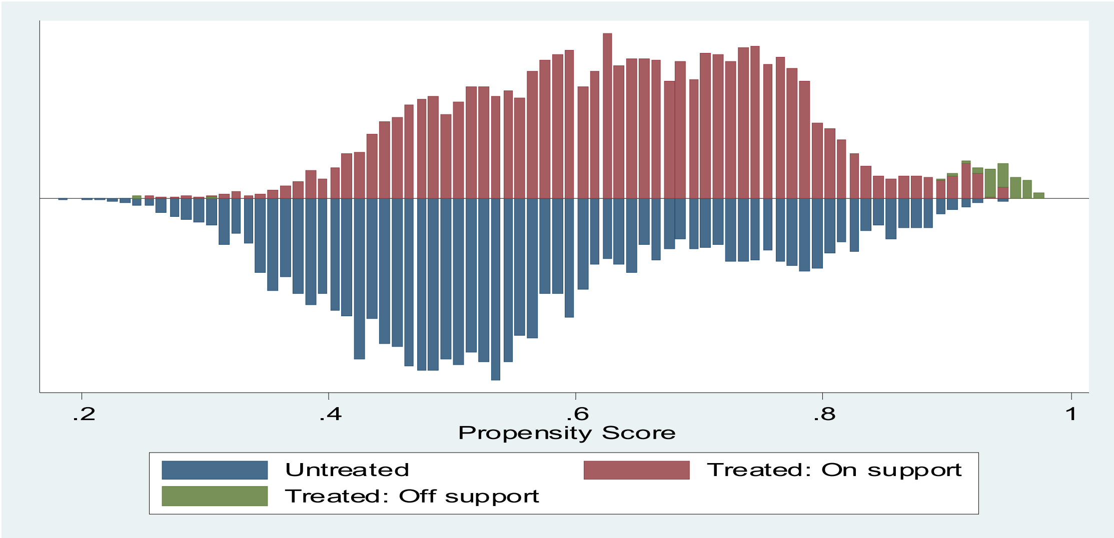
PSGRAPH (MODEL 1)



PSGRAPH (MODEL 2)



PSGRAPH (MODEL 3)



APPENDIX C: PSTESTS

Table 1: pstest Model 1

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	p>|t|
treat	Unmatched	1	0	.		.	.
	Matched	1	0	.	.	.	.
region	Unmatched	1.9993	1.8309	16.9		10.15	0.000
	Matched	1.9031	1.9355	-3.2	80.7	-1.65	0.100
age	Unmatched	11.312	11.372	-2.1		-1.29	0.198
	Matched	11.345	11.351	-0.2	90.4	-0.10	0.919
age_head	Unmatched	44.143	44.148	-0.0		-0.03	0.977
	Matched	44.402	44.507	-1.0	-1894.0	-0.47	0.638
age_spouse	Unmatched	40.1	40.164	-0.6		-0.38	0.705
	Matched	40.312	40.428	-1.1	-80.0	-0.55	0.580
single	Unmatched	.19193	.16461	7.1		4.31	0.000
	Matched	.18346	.17625	1.9	73.6	0.94	0.348
ed_level_1	Unmatched	2.4484	2.4922	-4.9		-3.01	0.003
	Matched	2.4749	2.461	1.6	68.4	0.78	0.436
edu_h2	Unmatched	.46894	.43071	7.7		4.66	0.000
	Matched	.44262	.44903	-1.3	83.2	-0.64	0.519
edu_h3	Unmatched	.15019	.15258	-0.7		-0.40	0.686
	Matched	.14681	.15041	-1.0	-51.1	-0.51	0.613
edu_h4	Unmatched	.08678	.10282	-5.5		-3.35	0.001
	Matched	.09994	.08832	4.0	27.5	1.99	0.047
edu_h5	Unmatched	.03192	.04663	-7.6		-4.69	0.000
	Matched	.04326	.03405	4.7	37.4	2.39	0.017
edu_s2	Unmatched	.4858	.42857	11.5		6.98	0.000
	Matched	.45123	.45784	-1.3	88.5	-0.66	0.507
edu_s3	Unmatched	.16749	.18356	-4.2		-2.57	0.010
	Matched	.17364	.17865	-1.3	68.8	-0.66	0.511
edu_s4	Unmatched	.08892	.11468	-8.5		-5.24	0.000
	Matched	.10855	.09533	4.4	48.7	2.18	0.029
edu_s5	Unmatched	.03565	.04894	-6.6		-4.07	0.000
	Matched	.04446	.03785	3.3	50.2	1.66	0.096
pershog	Unmatched	6.5381	6.5378	0.0		0.01	0.994
	Matched	6.5772	6.593	-0.7	-5294.0	-0.34	0.736
house	Unmatched	.97513	.9677	4.5		2.74	0.006
	Matched	.96956	.97336	-2.3	48.7	-1.14	0.254
gasbypipe	Unmatched	.08038	.08239	-0.7		-0.45	0.656
	Matched	.09273	.0743	6.7	-818.6	3.33	0.001
waterbypipe	Unmatched	.58529	.63487	-10.2		-6.16	0.000
	Matched	.63569	.61366	4.5	55.6	2.27	0.023
phone_3	Unmatched	.91343	.89636	5.8		3.56	0.000
	Matched	.90326	.90046	1.0	83.6	0.47	0.638
houseown_2	Unmatched	.08945	.08519	1.5		0.91	0.360
	Matched	.09674	.08933	2.6	-73.7	1.27	0.202
houseown_1	Unmatched	.67186	.64261	6.2		3.75	0.000
	Matched	.66153	.67595	-3.0	50.7	-1.53	0.126
no_hos_alc	Unmatched	.77114	.67145	22.4		13.74	0.000
	Matched	.73323	.75205	-4.2	81.1	-2.15	0.031
no_colurb_~c	Unmatched	8.6681	7.9558	7.9		4.92	0.000
	Matched	8.853	8.0499	8.9	-12.8	4.20	0.000
no_colrur_~c	Unmatched	42.637	27.34	57.7		34.75	0.000
	Matched	33.083	31.779	4.9	91.5	2.57	0.010
fam017d	Unmatched	390.96	366.3	3.6		2.10	0.035
	Matched	375.81	319.14	8.3	-129.9	4.22	0.000
famsis1d	Unmatched	561.59	515.99	4.6		2.63	0.009
	Matched	534.35	452.07	8.2	-80.4	4.17	0.000
incm_lab_h	Unmatched	1090.1	1132.4	-16.1		-9.72	0.000
	Matched	1123.1	1109.6	5.1	68.1	2.58	0.010
nucleos_1	Unmatched	.79206	.79222	-0.0		-0.02	0.980
	Matched	.7837	.79551	-2.9	-7074.5	-1.45	0.147
nucleos_2	Unmatched	.15969	.16955	-2.7		-1.62	0.106
	Matched	.16944	.16263	1.8	30.9	0.91	0.361
nucleos_3	Unmatched	.04825	.03823	4.9		2.96	0.003
	Matched	.04687	.04186	2.5	50.0	1.22	0.224

Table 2: pstest Model 2

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
treat	Unmatched	1	0	.		.	.
	Matched	1	0	.	.	.	.
region	Unmatched	1.9991	1.8309	16.8		10.14	0.000
	Matched	1.8675	1.9249	-5.8	65.9	-2.83	0.005
age	Unmatched	11.312	11.372	-2.1		-1.29	0.199
	Matched	11.347	11.345	0.1	96.9	0.03	0.975
single	Unmatched	.19191	.16461	7.1		4.30	0.000
	Matched	.1774	.18171	-1.1	84.2	-0.55	0.580
ed_level_1	Unmatched	2.4484	2.4922	-4.9		-3.02	0.003
	Matched	2.4672	2.4619	0.6	87.8	0.30	0.766
edu_h2	Unmatched	.46889	.43071	7.7		4.66	0.000
	Matched	.43909	.45509	-3.2	58.1	-1.59	0.112
edu_h3	Unmatched	.15018	.15258	-0.7		-0.41	0.684
	Matched	.15012	.15258	-0.7	-2.4	-0.34	0.735
edu_h4	Unmatched	.08678	.10282	-5.5		-3.35	0.001
	Matched	.09516	.08573	3.2	41.2	1.62	0.104
edu_h5	Unmatched	.03191	.04663	-7.6		-4.69	0.000
	Matched	.04696	.03733	5.0	34.5	2.37	0.018
edu_s2	Unmatched	.48575	.42857	11.5		6.97	0.000
	Matched	.44217	.4639	-4.4	62.0	-2.16	0.031
edu_s3	Unmatched	.16747	.18356	-4.2		-2.58	0.010
	Matched	.17494	.17884	-1.0	75.8	-0.50	0.614
edu_s4	Unmatched	.08891	.11468	-8.5		-5.24	0.000
	Matched	.1087	.09106	5.8	31.6	2.91	0.004
edu_s5	Unmatched	.03565	.04894	-6.6		-4.07	0.000
	Matched	.04594	.03856	3.7	44.4	1.81	0.070
gasbypipe	Unmatched	.08037	.08239	-0.7		-0.45	0.654
	Matched	.09454	.08183	4.7	-531.2	2.21	0.027
waterbypipe	Unmatched	.58533	.63487	-10.2		-6.16	0.000
	Matched	.63515	.62039	3.0	70.2	1.51	0.132
phone_3	Unmatched	.91344	.89636	5.8		3.57	0.000
	Matched	.90505	.90566	-0.2	96.4	-0.10	0.917
houseown_2	Unmatched	.08944	.08519	1.5		0.91	0.361
	Matched	.09639	.08737	3.2	-112.0	1.54	0.123
houseown_1	Unmatched	.6719	.64261	6.2		3.75	0.000
	Matched	.66161	.67186	-2.2	65.0	-1.07	0.283
no_colurb_~c	Unmatched	8.6683	7.9558	7.9		4.92	0.000
	Matched	8.8829	8.0736	9.0	-13.6	4.18	0.000
no_colrur_~c	Unmatched	42.64	27.34	57.7		34.75	0.000
	Matched	32.455	31.608	3.2	94.5	1.68	0.093
nucleos_1	Unmatched	.79197	.79222	-0.1		-0.04	0.970
	Matched	.78548	.7943	-2.2	-3438.2	-1.07	0.285
nucleos_2	Unmatched	.15978	.16955	-2.6		-1.60	0.109
	Matched	.16715	.16304	1.1	58.0	0.55	0.585
nucleos_3	Unmatched	.04824	.03823	4.9		2.95	0.003
	Matched	.04737	.04266	2.3	52.9	1.12	0.261
dum_death	Unmatched	.04643	.0379	4.2		2.55	0.011
	Matched	.04163	.04286	-0.6	85.6	-0.30	0.763
dum_ill01	Unmatched	.05198	.04498	3.3		1.96	0.050
	Matched	.04902	.0482	0.4	88.3	0.19	0.851
b_Lunch	Unmatched	.62312	.63123	-2.8		-1.72	0.086
	Matched	.62305	.6301	-2.5	13.0	-1.21	0.226
b_BuyBks	Unmatched	1.4156	1.4028	3.8		2.29	0.022
	Matched	1.4117	1.4136	-0.6	84.5	-0.29	0.773
abs_JanDec01	Unmatched	1.0369	1.0507	-5.4		-3.34	0.001
	Matched	1.0468	1.0367	3.9	26.9	2.01	0.044
EdLevWant	Unmatched	2.3691	2.4905	-19.9		-12.05	0.000
	Matched	2.442	2.4212	3.4	82.8	1.69	0.092
alumxpro	Unmatched	22.531	23.177	-12.5		-7.62	0.000
	Matched	22.761	22.498	5.1	59.4	2.47	0.014
n_pagos	Unmatched	1.2281	1.1988	7.3		4.01	0.000
	Matched	1.1862	1.1988	-3.1	56.7	-1.73	0.083
wage_male_m	Unmatched	1250.3	1253.3	-1.0		-0.59	0.556
	Matched	1255.7	1244.4	3.7	-270.7	1.83	0.068
wage_femal~m	Unmatched	956.8	1011.8	-12.6		-7.85	0.000
	Matched	1001.2	992.3	2.0	83.8	1.00	0.318
age_ent	Unmatched	6.6962	6.6902	0.4		0.26	0.792
	Matched	6.717	6.7247	-0.6	-28.7	-0.27	0.787

Table 3a: pstest Model 3

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	p>|t|
treat	Unmatched	1	0	.		.	.
	Matched	1	0	.	.	.	.
region	Unmatched	1.9043	1.6697	23.8		3.94	0.000
	Matched	1.8746	2.045	-17.3	27.3	-1.87	0.062
age	Unmatched	10.189	10.406	-8.8		-1.49	0.136
	Matched	9.9936	10.212	-8.9	-0.8	-1.06	0.288
age_head	Unmatched	42.116	42.828	-6.8		-1.17	0.243
	Matched	42.942	44.711	-17.0	-148.3	-2.10	0.036
age_spouse	Unmatched	38.754	39.55	-8.0		-1.37	0.170
	Matched	39.106	40.91	-18.2	-126.5	-2.33	0.020
single	Unmatched	.21868	.19037	7.0		1.18	0.238
	Matched	.19936	.12219	19.1	-172.6	2.63	0.009
ed_level_1	Unmatched	2.2128	2.2936	-10.7		-1.84	0.067
	Matched	2.1383	2.1994	-8.1	24.4	-1.06	0.290
edu_h2	Unmatched	.45863	.3555	21.1		3.55	0.000
	Matched	.41801	.39228	5.3	75.1	0.65	0.514
edu_h3	Unmatched	.16312	.20872	-11.7		-2.02	0.044
	Matched	.17042	.18971	-5.0	57.7	-0.63	0.532
edu_h4	Unmatched	.12766	.16514	-10.6		-1.83	0.067
	Matched	.13826	.06431	20.9	-97.3	3.07	0.002
edu_h5	Unmatched	.05674	.06651	-4.1		-0.70	0.486
	Matched	.05466	.09968	-18.7	-360.5	-2.11	0.035
edu_s2	Unmatched	.41726	.3211	20.0		3.36	0.001
	Matched	.36977	.38264	-2.7	86.6	-0.33	0.741
edu_s3	Unmatched	.22104	.23624	-3.6		-0.62	0.538
	Matched	.23794	.22186	3.8	-5.8	0.48	0.634
edu_s4	Unmatched	.13475	.18349	-13.3		-2.31	0.021
	Matched	.15113	.11254	10.6	20.8	1.42	0.155
edu_s5	Unmatched	.06265	.05046	5.3		0.88	0.379
	Matched	.06752	.07717	-4.2	20.9	-0.46	0.643
pershog	Unmatched	6.0816	6.0023	3.8		0.66	0.512
	Matched	6.1672	6.3537	-9.0	-135.3	-1.14	0.255
house	Unmatched	.95981	.94495	7.0		1.21	0.226
	Matched	.96141	.97749	-7.5	-8.2	-1.16	0.245
gasbypipe	Unmatched	.09102	.12156	-9.9		-1.72	0.086
	Matched	.09646	.12219	-8.3	15.8	-1.03	0.305
waterbypipe	Unmatched	.63948	.70642	-14.3		-2.40	0.016
	Matched	.66881	.62379	9.6	32.8	1.17	0.241
phone_3	Unmatched	.89598	.92661	-10.8		-1.78	0.075
	Matched	.88746	.8746	4.5	58.0	0.49	0.621
houseown_2	Unmatched	.08274	.11927	-12.1		-2.11	0.035
	Matched	.07717	.07717	0.0	100.0	-0.00	1.000
houseown_1	Unmatched	.66312	.54358	24.6		4.21	0.000
	Matched	.65916	.74277	-17.2	30.1	-2.28	0.023
no_hos_alc	Unmatched	.86052	.58257	65.2		11.71	0.000
	Matched	.84887	.86495	-3.8	94.2	-0.57	0.568
no_colurb_~c	Unmatched	10.543	7.6353	35.3		5.85	0.000
	Matched	10.196	8.8553	16.3	53.9	2.05	0.041
no_colrur_~c	Unmatched	37.098	19.337	71.8		11.80	0.000
	Matched	30.405	37.875	-30.2	57.9	-3.83	0.000
fam017d	Unmatched	391.33	385.63	0.8		0.13	0.897
	Matched	367.49	480.66	-16.4	-1886.2	-2.16	0.031
famsis1d	Unmatched	562.19	549.25	1.3		0.20	0.844
	Matched	533.94	681.23	-14.5	-1037.7	-1.91	0.057
incm_lab_h	Unmatched	1190.6	1217.8	-9.4		-1.59	0.113
	Matched	1240	1146.4	32.4	-243.8	4.80	0.000
price_rice	Unmatched	1273.7	1401.3	-60.8		-9.80	0.000
	Matched	1360.2	1383.6	-11.1	81.7	-2.42	0.016
price_chic~n	Unmatched	4457.5	4574.1	-16.0		-2.59	0.010
	Matched	4623.4	4754.6	-18.0	-12.6	-2.54	0.011
nucleos_1	Unmatched	.78723	.78899	-0.4		-0.07	0.942
	Matched	.77492	.81029	-8.6	-1913.3	-1.09	0.277
nucleos_2	Unmatched	.15957	.15596	1.0		0.17	0.867
	Matched	.15434	.13826	4.4	-345.2	0.57	0.571
nucleos_3	Unmatched	.05319	.05505	-0.8		-0.14	0.889
	Matched	.07074	.05145	8.5	-940.4	1.00	0.316
dum_death	Unmatched	.06619	.04128	11.1		1.81	0.070
	Matched	.04823	.05145	-1.4	87.1	-0.18	0.854
dum_ill01	Unmatched	.05437	.04128	6.1		1.02	0.308
	Matched	.06109	.05788	1.5	75.4	0.17	0.866

Table 3b: The Distribution of the Standarised Bias before and after Matching

BEFORE MATCHING

	Percentiles	Smallest		
1%	.2249482	.2249482		
5%	.9694404	.9694404		
10%	1.054334	1.054064	Obs	34
25%	2.417302	1.054334	Sum of Wgt.	34
50%	4.217777		Mean	9.764024
		Largest	Std. Dev.	14.72748
75%	8.468779	26.84944		
90%	26.84944	35.47184	Variance	216.8987
95%	58.63744	58.63744	Skewness	2.630717
99%	61.95455	61.95455	Kurtosis	9.096919

AFTER MATCHING

	Percentiles	Smallest		
1%	.2000306	.2000306		
5%	.289721	.289721		
10%	.3606214	.2930272	Obs	34
25%	.8157857	.3606214	Sum of Wgt.	34
50%	1.19327		Mean	1.607125
		Largest	Std. Dev.	1.284578
75%	2.352936	3.61925		
90%	3.61925	3.776485	Variance	1.650142
95%	3.961799	3.961799	Skewness	1.4837
99%	5.908238	5.908238	Kurtosis	5.035521

APPENDIX D: RESULTS FROM DIFFERENCE-IN-DIFFERENCES PROPENSITY SCORE MATCHING

Table 1a: OLS Estimates for the Effect of the Program on Grade Failure Rates (linear difference-in-differences after controlling for propensity score matching)

VARIABLES	**(1)** fail	**(2)** fail	**(3)** fail	**(4)** fail
Treatment area	-0.0140**	-0.0126**	-0.0148	-0.0142*
	(0.00593)	(0.00580)	(0.00920)	(0.00775)
Post	0.0560***	0.0450***	0.0507***	0.0591***
	(0.00885)	(0.00769)	(0.0161)	(0.0115)
Treatment area·post	-0.0582***	-0.0367***	-0.0591	-0.0634***
	(0.0184)	(0.0127)	(0.0339)	(0.0220)
Pscore	-0.00330		0.0117	-0.0184
	(0.0184)		(0.0285)	(0.0228)
Constant	0.105***	0.105***	0.0852***	0.122***
	(0.0113)	(0.00551)	(0.0180)	(0.0138)
Observations	13,903	33,161	4,982	8,921
Number of llavefin	9,415	23,057	4,042	6,574

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 1b: OLS Estimates for the Effect of the Program on Grade Failure Rates (linear difference-in-differences after controlling for propensity score matching)

VARIABLES	(1) fail	(2) fail	(3) fail	(4) fail
Treatment area	-0.00621	-0.0253***	-0.0207***	0.0149
	(0.00784)	(0.00909)	(0.00680)	(0.0107)
Post	0.0613***	0.0506***	0.0646***	0.0234
	(0.0116)	(0.0134)	(0.00946)	(0.0275)
Treatment area·post	-0.0850***	-0.0332	-0.0680***	-0.00166
	(0.0237)	(0.0276)	(0.0193)	(0.0717)
Pscore	0.0230	-0.0321	0.00862	-0.0601*
	(0.0260)	(0.0260)	(0.0215)	(0.0342)
Constant	0.0826***	0.133***	0.111***	0.0920***
	(0.0154)	(0.0168)	(0.0130)	(0.0234)
Observations	7,319	6,584	11,556	2,347
Number of llavefin	4,939	4,476	7,294	2,121

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2a: OLS Estimates for the Effect of the Program on School Drop-out Rates (linear difference-in-differences propensity score matching model)

	(1)	(2)	(3)	(4)
VARIABLES	**drop-out**	**drop-out**	**drop-out**	**drop-out**
Treatment area	-0.00943***	-0.00692***	-0.00720*	-0.0115***
	(0.00294)	(0.00219)	(0.00418)	(0.00391)
Post	-0.00297	-0.00323	0.00960	-0.00454
	(0.00401)	(0.00234)	(0.00818)	(0.00494)
Treatment area·post	-0.00986	-0.00827**	-0.0208	-0.00577
	(0.00759)	(0.00349)	(0.0184)	(0.00804)
Pscore	0.00366		-0.00932	0.00693
	(0.0100)		(0.0146)	(0.0122)
Constant	0.0303***	0.0334***	0.0269***	0.0312***
	(0.00620)	(0.00219)	(0.00983)	(0.00735)
Observations	13,903	33,161	4,982	8,921
Number of llavefin	9,415	23,057	4,042	6,574

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 2b: OLS Estimates for the Effect of the Program on School Drop-out Rates (difference-in-differences propensity score matching model)

	(1)	(2)	(3)	(4)
VARIABLES	**drop-out**	**drop-out**	**drop-out**	**drop-out**
Treatment area	-0.00916**	-0.0106**	-0.0111***	-0.000274
	(0.00383)	(0.00455)	(0.00318)	(0.00658)
Post	-0.00134	-0.00582	-0.00723*	0.0277**
	(0.00585)	(0.00564)	(0.00419)	(0.0122)
Treatment area·post	-0.01865	-0.0103	-0.00328	-0.0629**
	(0.0129)	(0.00859)	(0.00776)	(0.0313)
Pscore	0.00471	0.00180	0.00257	0.0102
	(0.0150)	(0.0135)	(0.0104)	(0.0282)
Constant	0.0268***	0.0355***	0.0308***	0.0207
	(0.00901)	(0.00869)	(0.00646)	(0.0170)
Observations	7,319	6,584	11,556	2,347
Number of llavefin	4,939	4,476	7,294	2,121

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table 3: OLS Estimates for the Effect of the Program with Heterogeneous Effects

Outcome Variable	ATT	ATT (females)	ATT (males)
Failure rates	-0.06759*	-0.06737	-0.07264***
	(0,056733)	(0.045039)	(0.030881)
Drop-out Rates	-0.01258	-0.03981*	-0.01017
	(0.010364)	(0.022939)	(0.01195)

Bootstrapped standard errors based on 100 replications in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 4: Estimates for the Effect of the Program using Non-parametric Difference-in-differences Propensity Score Matching

Outcome Variable	ATT	ATT (females)	ATT (males)
Failure rates	-0.055696***	-0.062687*	-0.069307***
	(0.019556)	(0.032078)	(0.026159)
Drop-out Rates	-0.003807	-0.010627	-0.003773
	(0.009457)	(0.014920)	(0.012894)

Bootstrapped standard errors based on 100 replications in parentheses

*** p<0.01, ** p<0.05, * p<0.1

APPENDIX E: ROBUSTNESS CHECKS

Table 1: Pre-program Time Trends in School Attendance in Treatment and Control Groups

VARIABLES	**School Attendance in 1999**
Treatment area	0.0648**
	(0.00588)
Year 2001	-0.0361**
	(0.00445)
Treatment area* Year 2000	-0.033
	-0.052
Treatment area* Year 2001	-0.801
	-0.492
Observations	41,985
Number of llavefin	22,131

Year 2000 is dropped due to collinearity; robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 2: Changes in School Supply over Time

Variable	Changes in treatment group over time	Changes in control group over time	Differences in changes over time between the two groups
Number of students/teacher	0,82408 (0,184196)	0.560495 (0.278354)	0.263585 (0.167339)
School provides books	-0.007811 (0.006027)	-0.001784 (0.007986)	-0.006027 (0.004505)
Number of teachers/school	-0.043978 (0.057838)	-0.049059 (0.091356)	0.005081 (0.058811)
School provides meals	-0.032744 (0.007259)	-0.071836 (0.010385)	-0.039092 (0.059583)
Traditional over New schools	-0.011091 (0.005641)	-0,017937 (0.007699)	0.006841 (0.004632)
School has a library	0.017112 (0.004918)	0.005737 (0.006886)	0.011375** (0.004493)

Differences in means; standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Probit Regression (Model 3) for TCP and TSP Municipalities

	Number of obs	= 13309
	LR chi2(33)	= 2601.50
	Prob > chi2	= 0.0000
Log likelihood = -7688.3299	Pseudo R2	= 0.1447

TCP	**Coef.**	**Std. Err.**	**Z**	**P>|z|**	**[95% Conf. Interval]**	
region	.133754	.0152536	8.77	0.000	.1038576	.1636505
age	-.0107958	.0060394	-1.79	0.074	-.0226328	.0010412
age_head	.0007071	.0018055	0.39	0.695	-.0028316	.0042458
age_spouse	-.0013337	.0019919	-0.67	0.503	-.0052378	.0025704
single	.1517586	.033551	4.52	0.000	.0859998	.2175173
ed_level_1	.0281292	.019726	1.43	0.154	-.010533	.0667914
edu_h2	.059771	.0312901	1.91	0.056	-.0015565	.1210985
edu_h3	.0285659	.0421139	0.68	0.498	-.0539758	.1111076
edu_h4	.00559	.0494868	0.11	0.910	-.0914022	.1025823
edu_h5	-.1277954	.0707901	-1.81	0.071	-.2665414	.0109506
edu_s2	.1318342	.033045	3.99	0.000	.0670672	.1966013
edu_s3	.0118027	.0422459	0.28	0.780	-.0709978	.0946033
edu_s4	-.0409772	.0505216	-0.81	0.417	-.1399977	.0580433
edu_s5	-.1531667	.0703415	-2.18	0.029	-.2910335	-.0152999
pershog	-.0019486	.0062087	-0.31	0.754	-.0141174	.0102202
house	.0482835	.0758289	0.64	0.524	-.1003384	.1969055
gasbypipe	.0808587	.0442169	1.83	0.067	-.0058049	.1675223
waterbypipe	-.0441586	.0259864	-1.70	0.089	-.0950909	.0067737
phone_3	.0551538	.0427162	1.29	0.197	-.0285683	.138876
houseown_2	.1029836	.0482294	2.14	0.033	.0084557	.1975114
houseown_1	.0242912	.0286373	0.85	0.396	-.0318368	.0804192
no_hos_alc	.5822374	.0312766	18.62	0.000	.5209364	.6435384
no_colurb_~c	-.0076624	.0014787	-5.18	0.000	-.0105606	-.0047643
no_colrur_~c	.0104879	.0004987	21.03	0.000	.0095105	.0114653
fam017d	-.0016647	.0001662	-10.02	0.000	-.0019904	-.001339
famsis1d	.0010755	.0001155	9.32	0.000	.0008492	.0013018
incm_lab_h	-.0002188	.0000527	-4.15	0.000	-.000322	-.0001156
price_rice	-.0020941	.0000852	-24.57	0.000	-.0022611	-.0019271
price_chic~n	-.0002846	.0000202	-14.12	0.000	-.0003241	-.0002451
nucleos_1	-.1529903	.0664516	-2.30	0.021	-.283233	-.0227476
nucleos_2	-.22674	.0651984	-3.48	0.001	-.3545266	-.0989534
dum_death	.0301953	.057295	0.53	0.598	-.0821008	.1424914
dum_ill01	.0354209	.0561237	0.63	0.528	-.0745796	.1454214
_cons	3.757044	.2147683	17.49	0.000	3.336105	4.177982

Table 4: pstest for TCP and TSP Municipalities

Variable	Sample	Mean Treated	Mean Control	%bias	%reduct |bias|	t-test t	t-test p>|t|
tcp	Unmatched	0	1	.		.	.
	Matched	0	1	.	.	.	.
region	Unmatched	1.8564	1.7872	6.9		3.89	0.000
	Matched	1.8482	1.9077	-5.9	14.0	-2.59	0.010
age	Unmatched	11.353	11.422	-2.4		-1.36	0.174
	Matched	11.365	11.4	-1.2	50.1	-0.53	0.599
age_head	Unmatched	44.278	44.161	1.1		0.60	0.550
	Matched	44.132	44.154	-0.2	81.0	-0.09	0.931
age_spouse	Unmatched	40.272	40.171	1.0		0.55	0.583
	Matched	40.127	40.247	-1.1	-18.2	-0.50	0.616
single	Unmatched	.20167	.16577	9.3		5.22	0.000
	Matched	.18236	.18469	-0.6	93.5	-0.26	0.791
ed_level_1	Unmatched	2.4634	2.4932	-3.3		-1.90	0.058
	Matched	2.4687	2.4713	-0.3	91.3	-0.13	0.899
edu_h2	Unmatched	.45218	.4111	8.3		4.70	0.000
	Matched	.4224	.4374	-3.0	63.5	-1.33	0.183
edu_h3	Unmatched	.1441	.14783	-1.1		-0.60	0.550
	Matched	.1464	.14537	0.3	72.2	0.13	0.897
edu_h4	Unmatched	.09198	.10657	-4.9		-2.78	0.005
	Matched	.1014	.09312	2.8	43.3	1.23	0.219
edu_h5	Unmatched	.03416	.0494	-7.6		-4.39	0.000
	Matched	.04656	.04346	1.6	79.6	0.66	0.510
edu_s2	Unmatched	.47685	.41314	12.8		7.27	0.000
	Matched	.43766	.44413	-1.3	89.9	-0.57	0.567
edu_s3	Unmatched	.16435	.17872	-3.8		-2.17	0.030
	Matched	.17434	.1715	0.8	80.2	0.33	0.741
edu_s4	Unmatched	.0916	.11674	-8.2		-4.71	0.000
	Matched	.10683	.10321	1.2	85.6	0.52	0.604
edu_s5	Unmatched	.03859	.05199	-6.4		-3.70	0.000
	Matched	.05044	.04863	0.9	86.5	0.37	0.714
pershog	Unmatched	6.5205	6.5863	-2.8		-1.57	0.118
	Matched	6.6223	6.6125	0.4	85.1	0.18	0.856
house	Unmatched	.97457	.97132	2.0		1.14	0.254
	Matched	.97232	.97569	-2.1	-3.6	-0.93	0.353
gasbypipe	Unmatched	.09223	.09158	0.2		0.13	0.899
	Matched	.09208	.08381	2.9	-1173.6	1.28	0.199
waterbypipe	Unmatched	.58831	.6296	-8.5		-4.79	0.000
	Matched	.63373	.62209	2.4	71.8	1.06	0.290
phone_3	Unmatched	.9146	.90435	3.6		2.03	0.042
	Matched	.90404	.90507	-0.4	89.9	-0.15	0.877
houseown_2	Unmatched	.09033	.07956	3.9		2.18	0.029
	Matched	.09286	.08898	1.4	64.0	0.59	0.553
houseown_1	Unmatched	.66903	.6568	2.6		1.47	0.142
	Matched	.66218	.67331	-2.4	9.0	-1.04	0.299
no_hos_alc	Unmatched	.81022	.65569	35.5		20.46	0.000
	Matched	.76643	.76073	1.3	96.3	0.59	0.556
no_colurb_~c	Unmatched	9.5363	8.1802	14.4		8.35	0.000
	Matched	8.9984	8.6249	4.0	72.5	1.73	0.083
no_colrur_~c	Unmatched	43.144	27.167	58.6		32.90	0.000
	Matched	33.324	33.096	0.8	98.6	0.38	0.703
fam017d	Unmatched	411.23	391.38	2.8		1.52	0.129
	Matched	397.22	391.25	0.8	69.9	0.38	0.704
famsis1d	Unmatched	588.51	553.92	3.3		1.79	0.074
	Matched	553.92	546.15	0.7	77.5	0.35	0.728
incm_lab_h	Unmatched	1117.4	1146.6	-12.0		-6.83	0.000
	Matched	1164.8	1155.6	3.8	68.5	1.60	0.109
price_rice	Unmatched	1281.4	1395.4	-62.0		-34.49	0.000
	Matched	1356.2	1358.9	-1.5	97.7	-0.79	0.427
price_chic~n	Unmatched	4538.9	4746.7	-26.8		-15.30	0.000
	Matched	4772.5	4780	-1.0	96.4	-0.45	0.652
nucleos_1	Unmatched	.78606	.78002	1.5		0.83	0.406
	Matched	.77884	.7822	-0.8	44.3	-0.36	0.721
nucleos_2	Unmatched	.16334	.18057	-4.6		-2.60	0.009
	Matched	.17175	.17589	-1.1	76.0	-0.48	0.631
nucleos_3	Unmatched	.05061	.03941	5.4		3.03	0.002
	Matched	.04941	.0419	3.6	33.0	1.58	0.114
dum_death	Unmatched	.04542	.04052	2.4		1.36	0.173
	Matched	.04863	.04578	1.4	42.0	0.59	0.555
dum_ill01	Unmatched	.04808	.04366	2.1		1.19	0.234
	Matched	.04863	.04682	0.9	59.0	0.37	0.709

Table 5a: OLS Estimates for the Effect of the Program on Grade Failure Rates (difference-in-differences model) when Comparing TCP and TSP Municipalities

	(1)	**(2)**	**(3)**
VARIABLES	**fail**	**fail**	**fail**
TCP	0.0215**	0.0246*	0.0176
	(0.0107)	(0.0145)	(0.0157)
Post	-0.123***	-0.103***	-0.141***
	(0.00675)	(0.00905)	(0.00995)
TCP·post	-0.0293***	-0.0222	-0.0557***
	(0.0101)	(0.0147)	(0.0139)
Pscore	-0.00471	-0.00536	-0.00526
	(0.0224)	(0.0294)	(0.0334)
Constant	0.125***	0.105***	0.144***
	(0.0135)	(0.0180)	(0.0199)
Observations	7,689	3,660	4,029
Number of llavefin	4,077	1,953	2,124

Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5b: OLS Estimates for the Effect of the Program on School Drop-out Rates (difference-in-differences model) when Comparing TCP and TSP Municipalities

	(1)	(2)	(3)
VARIABLES	**drop-out**	**drop-out**	**drop-out**
TCP	0.0129***	0.00254	0.0221***
	(0.00496)	(0.00625)	(0.00754)
Post	-0.0185***	-0.0166***	-0.0203***
	(0.00268)	(0.00363)	(0.00394)
TCP·post	-0.0121***	-0.0191***	-0.00413
	(0.00463)	(0.00584)	(0.00707)
Pscore	0.0167	0.0289**	0.00535
	(0.0112)	(0.0145)	(0.0167)
Constant	0.0105*	0.00311	0.0175*
	(0.00637)	(0.00788)	(0.00985)
Observations	7,689	3,660	4,029
Number of llavefin	4,077	1,953	2,124

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1