On the importance of families and public policies for child development outcomes

Rosero Moncayo, J.A.

Publication date
2012

Chapter 3

On the effectiveness of child care centers in promoting child development in Ecuador

3.1 Introduction

Human capital investments have been widely referred to as a key factor to achieve sustained growth and as one of the main determinants of the chances of an individual to break the intergenerational transmission mechanism of poverty (Mankiw et al., 1992; Solon, 1999). Moreover, a growing body of evidence suggests that investments made at early ages, as early as an individual is born, are critical to enhance the cognitive development of a person (Currie, 2001; Engle et al., 2007; Grantham-McGregor et al., 2007). The lack of investments during this sensitive period is considered to have long lasting consequences which are reflected later in life in poor levels of education and low levels of earnings. Using long term panels, Currie and Thomas (1999) and Case and Paxson (2008) have shown that cognitive tests taken at ages 5 and 7 are good predictors of educational attainment and wages at ages 30 and 33. In a related study, Feinstein (2003) finds that test scores taken as early as 22 months old predicts educational qualifications of children at age 26.

To prevent the adverse consequences of underinvestment at early ages, governments in developed and developing countries have implemented special programs targeted to children exposed to risk factors such as poverty, malnutrition and unstimulating environments. Their core motivation is to equalize allocations of early endowments and provide disadvantaged children a better start. The effects of these interventions in developed countries have been extensively analyzed (Currie, 2001; Almond and Currie, 2011). The experiences of small-scale child care interventions such as the Perry School and the Abecedarian programs in the United States have shown positive short-term effects in improving the performance of exposed children on several tests of cognitive development. In the long term, these programs also showed a positive effect on the likelihood of completion of tertiary education of treated individuals, as well as increasing the probability of being employed and receiving higher incomes in adulthood. In contrast to these findings, the results of the effects of large-scale interventions in developed countries are more mixed and suggest the absence or even a detrimental effect of child care interventions on cognitive tests (Barnett, 2011; Baker et al., 2008).

Compared to the literature for developed countries the evidence on the effects of early childhood programs in developing countries is still thin (Schady, 2006; Engle et al., 2007). Regarding large scale interventions, Behrman et al. (2004) evaluate the impact of a child care program in Bolivia using propensity score matching. The authors find positive effects on motor and language skills for children older than 37 months. The effects are more pronounced for children exposed for at least one year. Attanasio and Vera-Hernandez (2004) use the distance between the child's home and the care center as an instrumental variable to estimate the impact of a child care program in Colombia. They find positive impacts on child height, school enrollment at age 13 and mother's employment rates. Berlinski and Galiani (2007) and Berlinski et al. (2008) evaluate the impact of a program to construct preschool facilities in Argentina. They exploit variation in treatment intensity over time across regions and cohorts in a difference-in-differences framework and find a positive effect of the program on preschool enrollment and test scores of children between 3 and 5 years old. The analysis also finds that the program results in an increase of maternal labor participation.

In the second chapter of this thesis, a regression discontinuity design is used to evaluate the effects of a large scale program in Ecuador, the Child Development Fund (FODI), that offered two independent interventions: child care centers and home visits. In contrast with the positive impact of child care centers found in previous studies in developing countries, it finds no evidence of a positive impact on cognitive development of children exposed to this intervention. This result is robust across different types of tests of cognitive development. Regarding the home visit intervention, the results in chapter 2 find a positive effect across the different tests. The strength of the RD design is that it allows to estimate an effect that is internally valid. However, the effect is identified in a sample of children that are in the neighborhood of the discontinuity point. In case of heterogeneous effects the estimated effects might differ from the average effect that we would have obtained if we would have observed the effects for the whole distribution of children.

Considering the opposing results of the effect of child care centers and the empirical approach used in the previous chapter, this chapter assesses the effect of another provider of centers in Ecuador, the Child Rescue Program (ORI). This program shares the same purpose and target population as FODI. In order to deal with selection bias, this study uses propensity score matching methods to find a control group that is comparable to the children exposed to the program. Unlike RD estimators, matching methods identify an average effect that is valid for the whole distribution of treated children. Nevertheless, matching assumes that selection is driven only by observable characteristics. With the intention of reducing the influence of bias due to unobservables, matching is done using a sample of control children that have been selected a priori to be eligible for this type of intervention. For comparison, the analysis also presents impact results using the regression version of the efficient weighted estimator proposed by Hirano et al. (2003).

The results show that children exposed to the child care centers have no better scores on different cognitive tests than the children in the control group. Furthermore, the program has negative effects on health outcomes of the treated children and increases the probability of labor participation of their mothers as well as household income. These findings are in line with the ones presented in chapter 2 and are consistent with the findings presented in Baker et al. (2008) for the introduction of universal child care in the province of Quebec-Canada and Barnett (2011) for the largest child care provider in the United States.

The rest of the chapter is organized as follows. Section 3.2 describes the context of childhood development at early ages in Ecuador and introduces the Child Rescue Program (ORI). Section 3.3 presents the empirical approach used to identify the effect of the program and highlights its assumptions. Section 3.4 provides information of the data and the different outcomes which are employed in the analysis. Section 3.5 discusses empirically the identifying assumptions and presents the main results of this chapter. Section 3.6 elaborates on the potential mechanisms that may explain the effects and Section 3.7 summarizes and concludes.

3.2 Context and intervention

3.2.1 Context

As most Latin American countries, Ecuador is characterized by high levels of poverty and inequality. In 2006, around 38% of the Ecuadorian population were considered poor as their per capita consumption was below the national poverty line. Of these, circa 650 thousand were children between 0 and 5 years old. The situation of children's cognitive development in Ecuador is considered to be precarious and highly associated with family's socioeconomic background. Using a sample of nearly 3,000 children under 6 years old living in poor rural families in Ecuador, Paxson and Schady (2007) find that children who are in the poorest quartile of the socioeconomic distribution have an equivalent of 18 months of delay in terms of receptive vocabulary relative to children in the richest quartile at the time they start school. After a second data collection of the same children, Schady (2011) shows that the socioeconomic gradient found at pre-school age is maintained at the time children are in primary school. He also finds significant differences in other cognitive tests such as memory and visual integration.

The coverage of early child development care in Ecuador is low. It is estimated that only 24% of children at age 3 and almost 51% of children at age 4 attend a child development program.[1] Programs are offered by both private and public providers. Public providers account for 62% of all children covered and are mainly concentrated on children in poor families of rural and marginal-urban areas. Until 2009, the Ecuadorian government managed three child development programs. The Child Rescue Program (ORI), or "Operacion Rescate Infantil" in Spanish, was the smallest of these programs and covered a total of 50 thousand children.[2] ORI started in 1989 and was merged with the other two public providers in 2009.

3.2.2 Intervention

The Child Rescue Program (ORI) was designed, mainly, to promote the integral development of children under 6 years of low-income households in rural and marginal-urban areas of Ecuador. This was achieved through a child care based intervention which included an education/child care component as well as a nutritional component. Child care was provided 52 weeks per year, five days a week, eight hours per day. Food was given at least twice per day: breakfast and lunch.

An intervention was supplied at the community or neighborhood level to all children that fulfill the age requirement (children under 6 years). It was carried out in "centers" adapted for this purpose and provided by the guidance of trained "community mothers" (in Spanish "madres comunitarias"). Within the community, the intervention was designed to be highly participatory as it comprised the creation of a parent association among the parents of eligible children. The parents association provided the physical space for the operation of the center and selected community mothers among their members. ORI not only funded equipment, materials and food but also provided training and technical assistance for community mothers and the parents association. An additional objective of the program was to stimulate the labor participation of women whose children participated in the intervention.

By design, a center served a minimum of 20 children and a maximum of 60 children.

[1] Living Conditions Measurement Survey (ECV), 2006.

[2] The other two programs are INNFA (for "Instituto del Niño y la Familia") and FODI (for "Fondo de Desarrollo Infantil").

Every community mother worked with a group of 8 to 10 children. The decision to open a new center was based on the specific demand of a community or neighborhood. After receiving the demand, program's officials had the role to evaluate the request and decide upon the center's opening. The sole criterion for evaluation was the poverty condition of the requesting community/neighborhood and the decision was conditional on the availability of program resources. Within a community/neighborhood, there were no clear rules to allocate treatment across children apart from age. The annual cost of the intervention was US$ 530 per child.

3.3 Empirical approach

If the intervention in child development centers had been randomly assigned among children, its effect could be measured by comparing outcomes of children who received the intervention to outcomes of children who did not receive the intervention. However, as mentioned in section 3.2, the program targeted its intervention to communities with a high concentration of poor and vulnerable families. Additionally, even within a community considered eligible, the decision of a family to participate is endogenous.

To obtain a consistent estimate of the effect of the program a control group is needed to infer the counterfactual. In the absence of objective rules for the allocation of treatment across families or a baseline survey that could have been used in a quasi-experimental design, this chapter employs matching methods to identify the counterfactual outcome (Rosenbaum and Rubin, 1983; Heckman et al., 1997; Dehejia and Wahba, 1999). Intuitively, these methods create a counterfactual by matching children who participate in the program with untreated children that have similar observable attributes or a similar probability of treatment. The effect of the program is identified by the difference between the average outcome among treated individuals and the average among individuals that did not participate on the program. By aligning the distribution of observed characteristics between the treated and control children, matching attempts to imitate a randomized experiment (Heckman et al., 1997).

Formally, this identification strategy assumes that, conditional on a set of observable attributes X_i, the potential outcomes with or without intervention (Y_{1i} and Y_{0i}) are independent of treatment T_i (conditional independence assumption or unconfoundedness). Therefore, matching assumes that selection bias is driven purely by observable characteristics. The latter rules out selection into the program based on unobserved characteristics. If unobserved characteristics determine program participation and are also correlated with potential outcomes, matching estimates will be biased. If the conditional independence assumption is met, Rosenbaum and Rubin (1983) demonstrated that the potential outcomes are also independent of treatment conditional on a function of the probability of treatment based on X_i (propensity score). This is denoted by $Y_0, Y_1 \perp T_i \,|\, Pr(T_i = 1 \,|\, X_i)$. The main advantage of matching on the propensity score is that it reduces the dimensionality problem that results from trying to match individuals on each of the observable attributes X_i.

A second assumption needed for identification is the one of common support which states that for each child in the treated group there is a positive probability of finding an untreated child with a comparable propensity score to be matched. This implies that the estimates of the effect of the program are limited to propensity score values where treated and control observations are found (region of common support). The lack of common support for the entire sample of treated children is another potential cause of bias since the region of common support after matching may differ from the original sample to be analyzed.

If we denote the number of treated individuals N_1 and the number of untreated individuals as N_0, the average treatment effect on the treated is formally defined by:

$$ATT = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_{1i} - \sum_{j=1}^{N_0} W(i,j)\, Y_{0j} \right) \tag{3.1}$$

where no program outcomes (Y_0) for treated children i are estimated using a weighted average of outcomes of children j that did not participate from the program but have a similar propensity score. Based on equation 3.1, different matching methods can be generated on the basis of the choice of $W(i,j)$. The empirical analysis in this chapter uses the n nearest-neighbor method and the kernel method. The n nearest-neighbor method matches a treated observation with a set of n comparison individuals whose propensity score is the closest ($W(i,j) = \frac{1}{n}$). I consider two versions of this method: the first nearest neighbor ($n = 1$) and the five nearest neighbors ($n = 5$). Alternatively, the kernel method uses a weighted average of all individuals in the control group to estimate a counterfactual for each treated unit. In this case, the weight $W(i,j)$ is given by equation 3.2 where $K(.)$ is a kernel function and p_j is the propensity score of an untreated observation. Observations that are closer to the propensity score of the treated unit p_i receive a larger weight.

$$W(i,j) = \frac{K(p_j - p_i)}{\sum_{k=1}^{N_0} K(p_j - p_i)} \tag{3.2}$$

Inferring confidence intervals of propensity matching estimators based on equation 3.1 is a complicated task as the distributions of outcomes are not parametrically specified. Usually the problem is tackled by calculating standard errors using bootstrap. However, Abadie and Imbens (2008) show that bootstrap fails to provide asymptotically valid standard errors for matching estimators that use a fixed number of matches and, therefore, is not accurate as an inference method for the nearest neighbor estimators. On the other hand, the bootstrap method is considered to be appropriate to derive standard errors of kernel matching estimators. In the analysis that follows, confidence intervals for the propensity matching estimators are computed using bootstrap. By including the kernel estimator I test for inconsistencies of the bootstrap's confidence interval estimation of the nearest neighbor method.

Another drawback of matching estimators that use a fixed number of matches is related with its asymptotic properties. Abadie and Imbens (2006) analyzes the large sample properties of nearest neighbor estimators and find that even in a case where these estimators are consistent they entail a loss of efficiency compared to the efficiency bound derived by Hahn (1998). There are, however, fully efficient estimators in the econometric literature. Hirano et al. (2003) propose an estimator based on weighting observations by the inverse of the estimated propensity score. The authors show that their estimator is not only more efficient than matching estimators but lead to a fully efficient estimator since its large sample properties achieve the efficiency bound.[3] In this case, the estimator of the average treatment effect on the treatment is computed as follows:

$$ATT = \frac{\sum_{i=1}^{n} p_i \left(\frac{T_i}{p_i} - \frac{(1-T_i)Y_i}{(1-p_i)}\right)}{\sum_{i=1}^{n} p_i} \tag{3.3}$$

The weighting estimator of equation 3.3 can be rewritten by estimating the following regression function by weighted least squares:

$$Y_i = \alpha + \beta T_i + \theta X_i + \varepsilon_i \tag{3.4}$$

with weights equal to 1 for the treated observations and $p_i/(1-p_i)$ for the untreated observations. β is the average treatment effect on the treated and X_i is a set of covariates included to increase precision of the estimate. In the empirical analysis I present estimates of the program's effect using propensity matching estimators (nearest neighbor and kernel) and the regression version of the efficient weighted estimator proposed by Hirano et al. (2003). The comparison of results between different type of methods is used to evaluate the robustness of the impact estimates.

3.4 Data and outcomes

As noted above, matching assumes that selection bias is driven only by observable characteristics. This is a strong assumption since it rules out the possibility of selection

[3] Other fully efficient estimators in the literature are the ones proposed by Hahn (1998) using nonparametric series estimation and by Heckman et al. (1998) based on local linear kernel methods.

into the program based on unobserved characteristics. An extensive number of studies have evaluated the performance of non experimental methods in estimating an unbiased treatment effect by using experimental data as a benchmark (LaLonde, 1986; Heckman et al., 1997; Dehejia and Wahba, 1999; Smith and Todd, 2005; Dehejia, 2005). All of them conclude that the quality of the data available for the analysis is important for a reliable estimation.

Heckman et al. (1997) argue that matching methods are more likely to reduce the bias due to unobservables if (i) treated and untreated individuals come from similar settings and are likely to be affected by the same unobservable variables in their decision to participate in the program, (ii) the data available for the analysis is rich in terms of variables that are relevant to model the program participation choice, (iii) treated and control observations come from similar sources of data. As will be explained below, the sample design in this chapter follows the guidelines given by Heckman et al. in an attempt to reduce any bias driven by unobservables as it (i) uses as a control group a pool of children that were considered eligible for an early development intervention and also their families had expressed their willingness to participate in the intervention, (ii) uses a rich data set that includes an extensive number of variables that are relevant to estimate program participation and (iii) uses information for the treated and control individuals that was collected at the same time and using the same questionnaire and cognitive tests.

3.4.1 Sample design

The data used in this analysis comes from the first round of the ENEVIN survey. This self-collected survey was designed to investigate the effects of early development programs in Ecuador on children under 72 months old. The sample design of this survey includes three subsamples. The first subsample surveyed children that were treated by the ORI program. It consists of 469 families with at least one child younger than 72 months who participated in the program in June 2008 in a total of 26 centers previously stratified by geographic area (rural or marginal urban) and care center's size (measured in number of children). Within centers, a sample of children was randomly selected from a list of children served by the program.

The other two subsamples surveyed treated and untreated children of different child development programs in Ecuador that were part of other impact evaluation designs.[4] To construct a reliable comparison group, I use the children that were part of the untreated groups for a child care intervention as the pool of individuals to be matched with the children treated by ORI. By design, these children are considered to be eligible

[4]The two other programs that were evaluated using this survey are FODI and INNFA. The results of the impact evaluation of FODI are presented in chapter 2 of this thesis.

Table 3.1. Number of children by program participation and age group

Age	Untreated	Treated	Total
all	1246	660	1906
age >36 months	709	454	1163
age <60 months	957	541	1498

for a child care intervention and are also part of a list of children whose families have expressed their willingness to be part of this type of intervention.[5]By using this pool of children, I make sure that there are enough children of the same age range for the matching procedure and that they might share similar observable and unobservable characteristics. Although untestable, I assume that using this pool of untreated children for the matching procedure will eliminate endogeneity bias. The pool of potential control children come from 1,179 families with at least one child younger than 72 months old. Table 3.1 shows the number of observations in the analysis sample by treatment eligibility and age group.[6]

The final sample of treated and untreated children live in several provinces of Ecuador.[7] Teams of enumerators visited the homes of all children included in the sample between September 2008 and January 2009. At the moment of the data collection, treated children had been exposed to the program during 24 months, on average.

The main characteristics of the sample of analysis, before matching, by program participation are described in Table 3.2. The characteristics include an extensive group of variables at the child, household and community levels which are reasonably assumed not to be influenced by the treatment. Although the untreated children come from a pool of eligible children, table 3.2 shows that while most of the characteristics are balanced between the groups, some differences are statistically significant. Children in the treated group are on average three months older, they are more likely to be part of a family that receives a cash transfer program and they are less likely to either have their father present at home or live in municipalities with inpatient health facilities. The existence of differences between groups makes clear that a simple comparison of outcomes between groups would result in a biased estimate of the effect of the program.

[5]The data also includes untreated children that were elegible for a child development program but were willing to be part of a home visit intervention. These children were not considered in the pool of untreated individuals for matching since the decision of participating in home visits is endogenous and arguably different from the decision to participate in child care centers.

[6]In table 3.1 a division by age is presented since the impact analysis on cognitive outcomes uses tests that are administered to children at different ages.

[7]The sample includes observations of nine provinces of Ecuador (Bolivar, Chimborazo, Esmeraldas, Guayas, Imbabura, Los Rios, Manabi, Pichincha and Tungurahua) and two different regions (Highlands and the Coast).

Table 3.2. Differences by treatment status

Variable	Sample	Treated (1)	Control (2)	t (3)	p (4)
Boy (dummy)	Unmatched	0.521	0.513	0.35	0.726
	Matched	0.525	0.542	-0.45	0.654
Age (months)	Unmatched	43.908	40.146	1.73	0.087*
	Matched	43.798	43.910	-0.04	0.968
Household size	Unmatched	5.7515	5.9141	-0.49	0.627
	Matched	5.7632	5.6514	0.60	0.547
HH members <6	Unmatched	1.965	1.907	0.55	0.585
	Matched	1.950	1.956	-0.06	0.950
HH members 6-17	Unmatched	1.379	1.429	-0.23	0.821
	Matched	1.386	1.322	0.45	0.651
HH members: adults	Unmatched	2.132	2.195	-1.15	0.253
	Matched	2.143	2.085	0.87	0.388
HH members > 65	Unmatched	0.068	0.080	-0.79	0.434
	Matched	0.070	0.064	0.38	0.705
Urban	Unmatched	1.559	1.693	-1.02	0.312
	Matched	1.555	1.497	0.39	0.694
Cash transfer (dummy)	Unmatched	0.605	0.491	2.01	0.047**
	Matched	0.600	0.617	-0.25	0.803
Mother's age (years)	Unmatched	29.571	29.688	-0.18	0.858
	Matched	29.609	29.647	-0.06	0.954
Schooling mother (years)	Unmatched	6.082	6.244	-0.31	0.760
	Matched	6.098	6.240	-0.38	0.703
Schooling head (years)	Unmatched	6.105	6.252	-0.4	0.693
	Matched	6.131	6.365	-0.62	0.539
Language score mother	Unmatched	62.911	66.523	-0.74	0.459
	Matched	63.232	63.232	0	1.000
Father present	Unmatched	0.794	0.853	-1.88	0.064*
	Matched	0.796	0.812	-0.44	0.658
Mother present	Unmatched	0.971	0.971	0.01	0.991
	Matched	0.970	0.964	0.41	0.682
Poverty index[a]	Unmatched	0.569	0.629	-1.17	0.243
	Matched	0.577	0.602	-0.49	0.626
Illiteracy index[a]	Unmatched	16.278	20.753	-0.85	0.400
	Matched	16.402	15.499	0.41	0.682
School supply[a]	Unmatched	192.790	220.040	-0.28	0.780
	Matched	197.950	205.640	-0.08	0.934
High school supply[a]	Unmatched	350.140	383.820	-0.27	0.788
	Matched	358.410	356.670	0.01	0.989
Higher education supply [a]	Unmatched	108.850	128.560	-0.38	0.701
	Matched	111.720	114.320	-0.05	0.958
Density[a]	Unmatched	193.830	157.880	0.49	0.623
	Matched	197.540	244.780	-0.65	0.516
Outpatient health facilities[a]	Unmatched	2.615	2.544	0.2	0.845
	Matched	2.538	2.358	0.77	0.442
Inpatient health facilities[a]	Unmatched	0.399	0.603	-2.31	0.023**
	Matched	0.409	0.411	-0.02	0.985

Note: Columns (1) and (2) presents mean values. t and$p$ values are presented in columns (3) and (4) and are based on a test for equality of means. */**/*** denote significance at a 10/5/1% confidence level. The test includes all children in the sample of analysis. The matched sample is obtained using the 5 nearest neighbor propensity score method.

[a]Variables at the municipality level

3.4.2 Outcomes

The survey included an extensive questionnaire with detailed information of socioeconomic and demographic characteristics of all members of the families in the sample. The key feature of the data collection is that it contains a comprehensive battery of standard tests that measure different dimensions of child cognitive development. Some of these tests have been validated for children older than 36 months while others have been validated for children younger than 60 months old. For children older than 36 months the survey collected the following tests: (i) the Spanish version of the Peabody Picture Vocabulary test (TVIP), which measures the level of receptive vocabulary (language); (ii) the Woodcock-Muñoz test which measures the long-term memory skills; and (iii) the Pegboard test, which measures fine motor skills of the children. For those younger than 60 months cognitive development is assessed using the Nelson Ortiz test. This test provides measurements of child development in four specific dimensions: gross motor skills, fine motor skills, language skills and social behavior.

Scores on the tests can be standardized by age using the tables provided by the developers of the tests. For the analysis we normalize the results of the tests for children older than 36 months to have a mean equal to zero and a standard deviation of one. For the analysis we have converted the Nelson Ortiz scores to binary variables that equal 1 if a child has a level of development above the average score for his/her age, and 0 otherwise.[8]

The survey also collected data on measures related to children's health. Specifically, the survey has information on children's weight and height, as well as the outcome of a test that measures levels of hemoglobin in the blood. By using these variables I am able to create three health indicators: height for age, weight for age and the presence of iron deficiency anemia. Height and weight measurements were standardized by age and sex according to guidelines of the World Health Organization (WHO) in 2005. Hemoglobin levels were corrected for ground altitude differences given the GPS coordinates of each household collected in the sample. This adjustment was done following the CDC guidelines (Centers for Disease Control, 1989).

With respect to mother's outcomes, the empirical analysis include a number of measures of mother's mental health as well as variables that are related to the labor market. Mother's mental health is measured by the Center for Epidemiology Studies Depression Scale (CESD-D) which is a measurement of the degree of depression and psychological stress of the mother (Radloff, 1977) and the Home Observation for Measurement of the Environment scale (HOME) which measures maternal punitiveness

[8]The definition of the Nelson Ortiz score as a binary variable is in accordance to the criterion adopted by Ecuadorian authorities to evaluate the development of children in the early educational programs provided by the State.

and the degree of responsiveness to children (Bradley, 1993). CESD-D and HOME scores are standardized to have a zero mean and a standard deviation of one. Maternal labor market participation is captured either by a binary variable of participation or by the number of working hours during a week. The return to participation in the labor market is measured by monthly income.

3.5 Results

3.5.1 Identifying assumptions

All the matching estimators described in this chapter depend on the estimation of the propensity score. I estimate the probability of participation in the program using a probit specification. The model includes a set of covariates that are likely to determine program participation and that are not influenced by the program. Following Heckman et al. (1997) who show that selection bias can be reduced substantially if geographic variables are included in the estimation of the propensity score, I included variables at the municipality level such as poverty rate, population density, relative supply of schools, presences of high school and higher education institutions, and the relative availability of outpatient and inpatient health facilities. The final selection of variables for the model follows the hit-or-miss criterion suggested by Heckman et al. (1998) which aims to maximize the percentage of observations that are correctly specified as treated or untreated. By this criterion, almost 75% of the observations is correctly specified (72% in the treatment group and 77% in the control group).

Table 3.3 presents the results of the participation models. Since the analysis uses outcomes that apply to children at different age ranges, I estimate separate models by the age group of the children: older than 36 months, younger than 60 months and the entire sample. The results highlight the importance of socioeconomic and geographical characteristics to improve the quality of the match and to capture the potential effect of unobserved variables that might bias the estimation of the program's impacts.

The conditional independence assumption of a matching design implies that selection bias is explained only by observables and that, after controlling for these differences, potential outcomes are independent on the treatment status. While this assumption cannot be tested, we can test whether matching on the propensity score achieves to balance the observable characteristics. Table 3.2 shows mean values for background characteristics by program participation for the entire sample of children after applying the five nearest neighbors propensity score method. Using a t-test for equality of means, Table 3.2 indicates that after matching, there are no systematic differences by program participation in any of the characteristics and, therefore, the use

Table 3.3. Probit Participation Model

Variable	Age >36 months	Age < 60 months	All
	(1)	(2)	(3)
Boy (dummy)	0.033	0.016	0.008
	(0.089)	(0.079)	(0.068)
Age (months)	-0.025***	0.022***	0.008***
	(0.004)	(0.003)	(0.002)
HH members <6	0.414***	0.403***	0.446***
	(0.103)	(0.097)	(0.080)
HH members 6-17	0.164*	0.354***	0.285***
	(0.090)	(0.087)	(0.072)
HH members: adults	0.221**	0.335***	0.304***
	(0.098)	(0.095)	(0.080)
HH members > 65	0.186	0.352**	0.228*
	(0.160)	(0.153)	(0.133)
Urban (dummy)	-0.092	-0.103	-0.068
	(0.108)	(0.096)	(0.083)
Cash transfer (dummy)	0.283***	0.155*	0.197**
	(0.104)	(0.092)	(0.079)
Mother's age (years)	0.010	-0.000	0.003
	(0.008)	(0.007)	(0.006)
Schooling mother (yrs)	-0.037**	-0.039**	-0.036***
	(0.017)	(0.016)	(0.013)
Schooling head (yrs)	-0.005	0.031**	0.001
	(0.011)	(0.014)	(0.009)
Language score mother	-0.008***	-0.010***	-0.008***
	(0.002)	(0.002)	(0.002)
Father present	-0.210*	-0.385***	-0.310***
	(0.124)	(0.113)	(0.097)
Mother present	0.105	-0.324	-0.163
	(0.274)	(0.272)	(0.231)
Poverty index[a]	-3.438***	-2.754***	-2.588***
	(0.424)	(0.354)	(0.307)
Illiteracy index[a]	-0.040***	-0.035***	-0.038***
	(0.005)	(0.005)	(0.004)
School supply[a]	0.037***	0.039***	0.037***
	(0.004)	(0.004)	(0.003)
High school supply[a]	0.002***	0.003***	0.002***
	(0.001)	(0.000)	(0.000)
Higher education supply [a]	-0.076***	-0.081***	-0.076***
	(0.008)	(0.007)	(0.006)
Density[a]	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Outpatient health facilities[a]	0.228***	0.124**	0.201***
	(0.061)	(0.050)	(0.044)
Inpatient health facilities[a]	-1.294***	-1.018***	-0.995***
	(0.141)	(0.122)	(0.106)
Constant	5.191***	3.668***	2.962***
	(0.798)	(0.670)	(0.579)
N	1163	1498	1906
Pseudo R square	0.304	0.305	0.265

Note: */**/*** denote significance at a 10/5/1% confidence level. Robust standard errors that are clustered at the community-level in parentheses. Number of clusters equals 52.

[a]Variables at the municipality level

of propensity score as a balancing tool is successful in eliminating the bias associated with observable characteristics.

Since achieving balance for the entire sample of children does not guarantee balance for the different subsamples of children according to their age range, tables A1 and A2 in the appendix presents the results of the balancing test for the subsample of children older than 36 months and the one of children younger than 60 months. The tables in the appendix also present results of this test using different matching methods (nearest neighbor, five nearest neighbors and kernel). As this table shows, the achievement of balanced characteristics for both groups is independent of the age group and the specific matching method.

To check the common support assumption, Figures 3.1 and 3.2 plot the propensity score distributions for the treated and untreated groups separately for the two subsamples of children. Both histograms show that there is considerable overlap in support as I can find untreated units that can be matched with the treated units in any range of propensity score values.[9] However, the figures also warn about the quality of the overlap in the upper segment of the distributions since there are treated observations with a high propensity score that are comparable with only few observations in the control group.

3.5.2 Results on cognitive, motor and social outcomes

Table 3.4 reports the impact estimates on cognitive, motor and social outcomes. The first three columns show results for children older than 36 months. The last four columns report results for children younger than 60 months . As mentioned in section 3.3 I present estimates using different propensity matching methods. As usual, the estimates for the matching estimators are restricted to the area of common support and the standard errors are calculated by bootstrap with 100 repetitions.

The table also presents two type of results based on the regression version of the efficient weighted estimator. While the first uses the full sample of observations (all the children in a specific age group), the second type uses a trimmed sample which eliminates untreated observations that have an estimated propensity score close to one. Frölich (2004) showed that the efficient weighted estimator has erratic small sample properties since the weight $p_i/(1-p_i)$ can become very large for values of p_i that are close to one. Hence, the idea behind trimming the sample is to remove observations that tend to dominate the estimator since they represent a large share of the weights. We follow the trimming criterion suggested by Imbens (2004) dropping

[9]Observations with weak common support are eliminated for the empirical analysis. Depending on the subsample under analysis, between 2% and 3% of the treated units are dropped after imposing common support.

Figure 3.1. Common Support Assumption. Sample of children older than 36 months

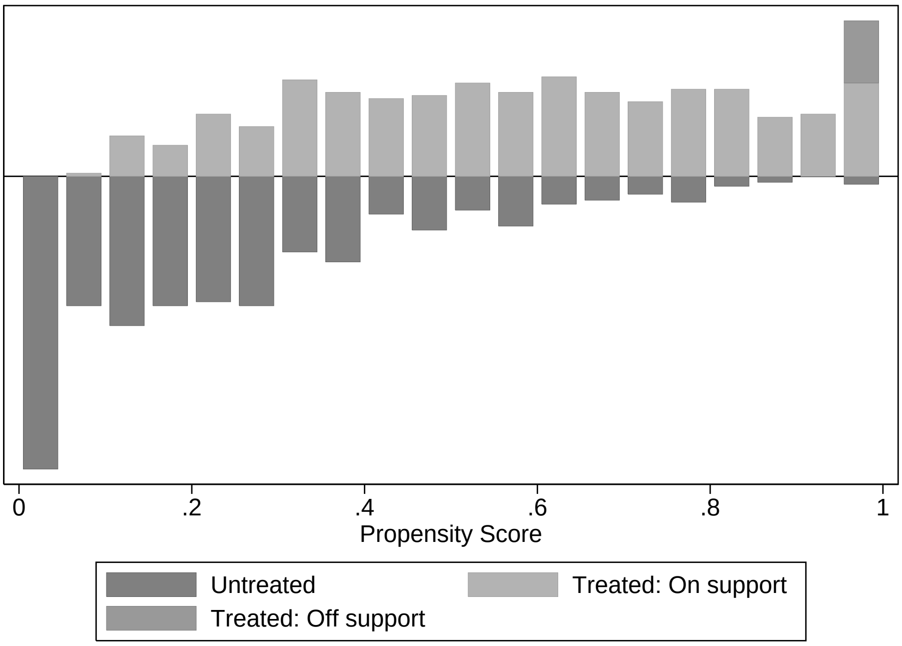
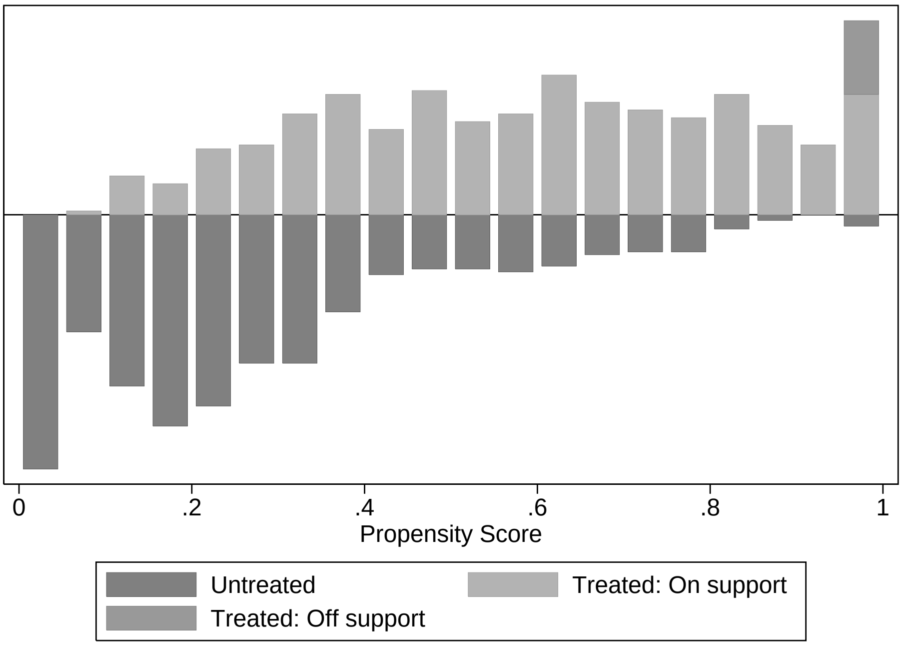


Figure 3.2. Common Support Assumption. Sample of children younger than 60 months



observations that have a weight share greater than 5%.

As presented in table 3.4 the results are not sensitive to the selection of a specific matching method or to the use of the efficient weighting estimator. Of the seven outcomes considered, only the estimate of the impact on the language test for children above 36 months is significantly negative, implying that child care centers harm the language skills of children exposed to the program. The estimates of the impacts on the other six tests do not reject the null hypothesis of being equal to zero. Taken together, I interpret the results as evidence that there is no effect of child care centers on cognitive, motor and social outcomes.

3.6 Potential mechanisms

The results from the previous section show a neutral impact of child care centers of cognitive outcomes. Although in terms of public policy this finding is relevant, it is equally important to inquire possible mechanisms that explain this zero effect. As explained in section 3.2, a child care program consists on a complex set of components and incentives which affect not only the outcomes of treated children but are also likely to affect the behavior of other members of the family. On the one hand, the program offers an early stimulation/educational component which is complemented with a nutritional component. Both components have the intention to improve the cognitive development of the children. On the other hand, by releasing time of mothers out of child care activities, the intervention is likely to have an effect on mother's labor participation, income and even on mother's parenting style. All these factors in turn might have a direct effect on children's cognitive outcomes. This is consistent with a cognitive production function of a child which is determined by at least four inputs: quantity of time devoted to the children, quality of this time, household income and health outcomes. The total effect on child development is, in theory, ambiguous, and depends both on the magnitude of the partial effects of the program, on the inputs and the way they interact (offsetting or reinforcing each other).[10] In the remainder of this section I present evidence on the effects of the program on outcomes that might serve as mechanisms to explain the results found on cognitive outcomes.

3.6.1 Results on children's health outcomes

Table 3.5 presents estimates of the effect of the program on children's health outcomes: anemia, underweight and below height. As in table 3.4 results are presented using different propensity matching methods and the regression version of the efficient weighted

[10]Section 2.7 presents a simple model of the production function of cognitive development that shows how different inputs interrelate.

Table 3.4. Results on cognitive, motor and social outcomes

	Older than 36 months			Younger than 60 months			
	Language	Fine motor	Memory	Gross motor	Fine motor	Language	Social
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Nearest neighbor	-0.391**	0.002	-0.096	0.046	-0.036	-0.034	-0.057
	(0.160)	(0.203)	(0.194)	(0.090)	(0.051)	(0.059)	(0.071)
5 nearest neighbors	-0.315**	0.055	-0.074	0.061	-0.054	-0.030	-0.021
	(0.124)	(0.228)	(0.151)	(0.060)	(0.038)	(0.057)	(0.049)
Kernel	-0.276**	0.079	-0.056	0.058	-0.055	-0.037	-0.034
	(0.130)	(0.197)	(0.184)	(0.053)	(0.035)	(0.052)	(0.048)
Weighted OLS	-0.161**	0.317**	0.004	0.033	-0.042	-0.023	-0.005
	(0.067)	(0.141)	(0.079)	(0.057)	(0.027)	(0.038)	(0.034)
N	1163	1163	1163	1498	1498	1498	1498
Weighted OLS (trimmed)	-0.171**	0.007	-0.010	0.017	-0.055**	-0.040	-0.021
	(0.067)	(0.078)	(0.073)	(0.051)	(0.023)	(0.038)	(0.034)
N	1132	1132	1132	1480	1480	1480	1480

Note: */**/*** denote significance at a 10/5/1% confidence level. Standard errors for matching estimators are in parentheses and are computed using bootstrap with 100 repetitions. Robust standard errors for the efficient weighted estimators that are clustered at the community-level in parentheses. Number of clusters equals 52.

Table 3.5. Results on health outcomes

	Anemia (1)	Under weight (2)	Below height (3)
Nearest neighbor	0.034	0.029	0.174**
	(0.075)	(0.033)	(0.083)
5 nearest neighbors	0.070	0.021	0.181**
	(0.056)	(0.031)	(0.085)
Kernel	0.074	0.018	0.171***
	(0.052)	(0.028)	(0.058)
Weighted OLS	0.047	0.026	0.181***
	(0.038)	(0.018)	(0.040)
N	1610	1867	1869
Weighted OLS (trimmed)	0.042	0.026	0.156***
	(0.039)	(0.019)	(0.047)
N	1590	1846	1848

Note: */**/*** denote significance at a 10/5/1% confidence level. Standard errors for matching estimators are in parentheses and are computed using bootstrap with 100 repetitions. Robust standard errors for the efficient weighted estimators that are clustered at the community-level in parentheses. Number of clusters equals 52.

estimator. The table indicates that while there is no statistically significant effect of the program in changing the prevalence of anemia or the likelihood of being underweight, child care centers increase the probability of a child to be below height by 17 to 18 percentage points (relative to a baseline of 0.28). These results are robust to the method used in the estimation.

Taking into account that the program has a nutritional component, the results on health outcomes are particularly striking. Two mechanisms can explain this result. The first is that food in schools is a poor substitute of food that children otherwise would have received at home. The second is that, given the food that children receive in the center, families of treated children change their behavior in a way that is detrimental to the children by not feeding them properly at home. Regarding the first mechanism, the information that is available in the survey is insufficient to test it. In addition, the program does not have a monitoring system of the nutritional component to assess the quality of the food that children receive in the center.

Regarding the second mechanism, the survey contains questions such as the perception of parents about the nutritional quality of the program and whether the center informs the parents about the food that they receive. Although 85% of families with a child in a center consider that the quality of the food provided is good, only 55% report that the center has ever given them information related to the food that their children received. This shows a high degree of ignorance on the side of the families about the true state of nutrition and/or if the child is being fed properly in the center. The survey also collects data on whether families have reduced the number of meals provided to the child and on whether they have decreased the amount of food in each meal. Using this information I find that, families with a treated child are 13 percentage points more likely to reduce the number of daily meals and 12 percentage points more likely to reduce the amount of food in each meal, than a family with a child in the control group. Although a decrease on the number of meals that a child receives at home is consistent with a program that gives free food to children in a center, a decrease in the amount of food per meal might indicate a detrimental change in family's behavior and explain the negative effect on children's height.

3.6.2 Results on mother's outcomes

Table 3.6 presents the effect of the program on outcomes related to mother's labor participation and income (columns 1 to 5) as well as on outcomes related to mother's psychological well being (columns 6 to 7). Focusing first on labor market outcomes, the results show a significantly positive effect of the program on the probability that a mother of a treated child works. The magnitude and significance of the result is independent of the matching method and points to an effect that is between 20 and 22 percentage points higher than for mothers of children not exposed to the program. This is also reflected in working hours, when children are treated their mothers work between 9 and 10 hours more per week.

Higher labor participation of the mother has the potential to increase family income if it does not induce the decrease of the supply of labor of other household members. I find that the effect of the program on household income is positive and significant. Household income increases in a range of 23 to 30 USD per month relative to a baseline of US$260. The size of this effect is comparable to the observed increase of mother's income which indicates that the program does not induce a negative change on the labor supply of other members of the household.

Finally, I look at the effect of the program on mother's psychological outcomes. The program, through its impact on maternal labor participation has the potential to affect these outcomes which in turn might have an impact on the quality of the time spent with their children as well as on their cognitive development. Table 3.6 shows that being exposed to the program has no significant effect on mother's level of depression according to the CES-D scale and also has no significant effect on mother's level of responsiveness measured by the Home scale.

Table 3.6. Results on labor market and parenting outcomes

	Labor mother	Hours mother	Income mother	Income head	Income HH	Stress/depression	Non-responsiveness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Nearest neighbor	0.201***	10.253***	30.346***	-6.123	23.206	-0.129	0.004
	(0.054)	(3.079)	(7.790)	(13.089)	(21.341)	(0.150)	(0.165)
5 nearest neighbors	0.217***	9.465***	27.454***	-2.922	26.888*	-0.126	-0.005
	(0.061)	(2.539)	(7.159)	(12.648)	(16.337)	(0.137)	(0.129)
Kernel	0.218***	9.512***	26.778***	-2.063	25.634*	-0.139	0.023
	(0.063)	(2.549)	(7.151)	(9.190)	(13.924)	(0.105)	(0.134)
Weighted OLS	0.203***	9.154***	25.015***	3.769	31.049***	-0.113	0.033
	(0.052)	(1.981)	(5.425)	(6.431)	(10.886)	(0.080)	(0.082)
N	1906	1906	1906	1906	1906	1906	1906
Weighted OLS (trimmed)	0.207***	9.861***	27.789***	2.485	33.072***	-0.110	0.053
	(0.046)	(1.722)	(4.721)	(6.584)	(11.280)	(0.088)	(0.088)
N	1885	1885	1885	1885	1885	1885	1885

Note: */**/*** denote significance at a 10/5/1% confidence level. Standard errors for matching estimators are in parentheses and are computed using bootstrap with 100 repetitions. Robust standard errors for the efficient weighted estimators that are clustered at the community-level in parentheses. Number of clusters equals 52.

3.7 Conclusions

Most of the literature on the effectiveness of large scale child care centers in developing countries has concluded that this type of intervention is beneficial to promote the cognitive development of poor children at early ages. In contrast to this literature, the results presented in chapter 2 do not find an effect on cognitive tests for the largest provider of child care services in Ecuador (FODI) using a regression discontinuity design. Although the RD design allows us to estimate an effect that is internally valid, it is identified on a special sample of children in the neighborhood of the discontinuity point.

Considering these conflicting results and the approach used in the previous chapter, I investigate the effects of another provider of child care services in Ecuador, the Child Rescue Program (ORI), which has the same objective and same target population. Using non-experimental methods that have also been applied in previous studies in developing countries, this chapter finds that children exposed to the ORI program do not perform better on a large range of cognitive tests than a group of eligible children that was not exposed to the intervention. Considering that matching methods estimate an average treatment effect, the results on this chapter give support to the findings reported in chapter 2 on the impact effect of child care centers and are in line to the evidence presented for large scale programs in Canada and United States (Baker et al., 2008; Barnett, 2011).

Although the findings on children's cognitive outcomes are relevant for public policy, it is equally important to investigate the potential mechanisms that might explain the lack of effect. This chapter provides evidence of the effect of the program on some of these mechanisms: children's health, quantity and quality of mother's time spent with the children, and household income. The results show that the program has a negative effect on children's health as it increases the probability of a child to be malnourished. It also has a negative effect on the quantity of time mothers spent with their children since it increases mothers' labor participation and mothers' hours of work. Moreover, the program increases household income and does not reduce the labor supply of other household members. Finally, unlike the results reported for FODI, this chapter gives evidence that the program does not have an adverse effect on the quality of mother's time as measured by parenting styles such as the mental health of the mother and the responsiveness of mothers toward their children.

Assuming that the pure effect of the educational component of the program on children's cognitive development is positive, the results presented in this chapter suggest that the potential positive effects of the stimulation received in the center and the higher household income are apparently canceled by the negative effects of the program on children's health outcomes and on the amount of maternal time spent with the children. This leads to a neutral total effect of the program on the cognitive development of the treated children.

These results give additional support to the existence of a trade-off between children's development and mothers' labor market participation as described in chapter 2. Being informed about the existence of this trade-off is especially relevant for policy makers at the moment of designing and implementing efficient policies aiming to enhance children's development at early ages while ensuring women's right to work in developing countries.

Appendix

Table A1: p values of a t test of equality of means by sample type. Matching method: Nearest neighbor and 5 Nearest neighbor

Variable	Sample	Nearest neighbor			5 Nearest neighbor		
		All	≥36	<60	All	≥36	<60
Boy (dummy)	Unmatched	0.726	0.875	0.446	0.726	0.875	0.446
	Matched	0.578	0.914	0.772	0.654	0.975	0.519
Age (months)	Unmatched	0.087*	0.001**	0.000***	0.087*	0.001**	0.000***
	Matched	0.649	0.312	0.779	0.968	0.269	0.365
Household size	Unmatched	0.627	0.741	0.567	0.627	0.741	0.567
	Matched	0.651	0.365	0.240	0.547	0.376	0.294
HH members <6	Unmatched	0.585	0.387	0.879	0.585	0.387	0.879
	Matched	0.940	0.696	0.845	0.950	0.898	0.427
HH members 6-17	Unmatched	0.821	0.536	0.851	0.821	0.536	0.851
	Matched	0.673	0.821	0.192	0.651	0.575	0.109
HH members: adults	Unmatched	0.253	0.885	0.052	0.253	0.885	0.052
	Matched	0.933	0.267	0.413	0.388	0.479	0.561
HH members > 65	Unmatched	0.434	0.804	0.619	0.434	0.804	0.619
	Matched	0.376	0.079	0.389	0.705	0.325	0.636
Cash transfer (dummy)	Unmatched	0.047**	0.111	0.048**	0.047**	0.111	0.048**
	Matched	0.855	0.697	0.844	0.803	0.718	0.943
Urban	Unmatched	0.312	0.362	0.286	0.312	0.362	0.286
	Matched	0.639	0.587	0.795	0.694	0.602	0.628
Mother's age (years)	Unmatched	0.858	0.365	0.682	0.858	0.365	0.682
	Matched	0.637	0.511	0.262	0.954	0.697	0.671
Schooling mother (yrs)	Unmatched	0.760	0.958	0.612	0.760	0.958	0.612
	Matched	0.636	0.238	0.659	0.703	0.416	0.707
Schooling head (yrs)	Unmatched	0.693	0.621	0.911	0.693	0.621	0.911
	Matched	0.975	0.413	0.519	0.539	0.461	0.696
Language score mother	Unmatched	0.459	0.713	0.391	0.459	0.713	0.391
	Matched	0.909	0.841	0.587	1.000	0.614	0.966
Father present	Unmatched	0.064*	0.271	0.008**	0.064*	0.271	0.008**
	Matched	0.631	0.746	0.943	0.658	0.814	0.731
Mother present	Unmatched	0.991	0.323	0.171	0.991	0.323	0.171
	Matched	0.910	0.493	0.505	0.682	0.727	0.777
Poverty index[a]	Unmatched	0.243	0.269	0.194	0.243	0.269	0.194
	Matched	0.691	0.692	0.935	0.626	0.768	0.833
Illiteracy index[a]	Unmatched	0.400	0.347	0.458	0.400	0.347	0.458
	Matched	0.644	0.409	0.736	0.682	0.576	0.758
School supply[a]	Unmatched	0.780	0.822	0.602	0.780	0.822	0.602
	Matched	0.897	0.629	0.981	0.934	0.600	0.962
High school supply[a]	Unmatched	0.788	0.903	0.670	0.788	0.903	0.670
	Matched	0.963	0.698	0.975	0.989	0.654	0.969
Higher education supply[a]	Unmatched	0.701	0.928	0.546	0.701	0.928	0.546
	Matched	0.919	0.651	0.983	0.958	0.621	0.982
Density[a]	Unmatched	0.623	0.505	0.550	0.623	0.505	0.550
	Matched	0.479	0.412	0.591	0.516	0.432	0.665
Outpatient health facilities[a]	Unmatched	0.845	0.890	0.845	0.845	0.890	0.845
	Matched	0.330	0.469	0.863	0.442	0.399	0.941
Inpatient health facilities[a]	Unmatched	0.023**	0.026**	0.022**	0.023**	0.026**	0.022**
	Matched	0.995	0.874	0.924	0.985	0.970	0.952

Note: p values are based on a test for equality of means. */**/*** denote significance at a 10/5/1% confidence level.

[a]Variables at the municipality level

Table A2: p values of a t test of equality of means by sample type. Matching method: Kernel

Variable	Sample	Kernel		
		All	≥36	<60
Boy (dummy)	Unmatched	0.726	0.875	0.446
	Matched	0.732	0.962	0.541
Age (months)	Unmatched	0.087*	0.001**	0.000***
	Matched	0.963	0.345	0.552
Household size	Unmatched	0.627	0.741	0.567
	Matched	0.492	0.629	0.316
HH members <6	Unmatched	0.585	0.387	0.879
	Matched	0.989	0.956	0.591
HH members 6-17	Unmatched	0.821	0.536	0.851
	Matched	0.570	0.798	0.198
HH members: adults	Unmatched	0.253	0.885	0.052
	Matched	0.501	0.290	0.619
HH members > 65	Unmatched	0.434	0.804	0.619
	Matched	0.636	0.716	0.621
Cash transfer (dummy)	Unmatched	0.047**	0.111	0.048**
	Matched	0.883	0.927	0.855
Urban	Unmatched	0.312	0.362	0.286
	Matched	0.677	0.636	0.608
Mother's age (years)	Unmatched	0.858	0.365	0.682
	Matched	0.895	0.519	0.572
Schooling mother (yrs)	Unmatched	0.760	0.958	0.612
	Matched	0.693	0.580	0.600
Schooling head (yrs)	Unmatched	0.693	0.621	0.911
	Matched	0.628	0.672	0.970
Language score mother	Unmatched	0.459	0.713	0.391
	Matched	0.967	0.811	0.768
Father present	Unmatched	0.064*	0.271	0.008**
	Matched	0.737	0.911	0.657
Mother present	Unmatched	0.991	0.323	0.171
	Matched	0.855	0.517	0.799
Poverty index[a]	Unmatched	0.243	0.269	0.194
	Matched	0.663	0.619	0.811
Illiteracy index[a]	Unmatched	0.400	0.347	0.458
	Matched	0.720	0.721	0.797
School supply[a]	Unmatched	0.780	0.822	0.602
	Matched	0.950	0.822	0.936
High school supply[a]	Unmatched	0.788	0.903	0.670
	Matched	0.994	0.885	0.993
Higher education supply[a]	Unmatched	0.701	0.928	0.546
	Matched	0.970	0.845	0.952
Density[a]	Unmatched	0.623	0.505	0.550
	Matched	0.498	0.499	0.590
Outpatient health facilities[a]	Unmatched	0.845	0.890	0.845
	Matched	0.457	0.506	0.807
Inpatient health facilities[a]	Unmatched	0.023**	0.026**	0.022**
	Matched	0.978	0.914	0.983

Note: p values are based on a test for equality of means. */**/*** denote significance at a 10/5/1% confidence level.

[a]Variables at the municipality level