



Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
<http://economics.sas.upenn.edu/pier>

PIER Working Paper 11-024

“Are Conditional Cash Transfers Effective in Urban Areas? Evidence from Mexico”

by

Jere R. Behrman, Jorge Gallardo-García,
Susan W. Parker, Petra E. Todd and Viviana Vélez-Grajales

<http://ssrn.com/abstract=1916083>

Are Conditional Cash Transfers Effective in Urban Areas? Evidence from Mexico¹

Jere R. Behrman, Jorge Gallardo-García,
Susan W. Parker, Petra E. Todd and Viviana Vélez-Grajales

August, 2011

¹This paper builds on a paper that Behrman, Gallardo-García, Parker and Todd wrote in 2006 as consultants to the Instituto Nacional de Salud Pública's (INSP) evaluation of Oportunidades under a subcontract to Todd, with additional support from the Mellon Foundation/Population Studies Center (PSC)/University of Pennsylvania grant to Todd (PI). Jere Behrman is W.R. Kenan Jr. Professor of Economics and Sociology and a PSC Research Associate at the University of Pennsylvania. Susan Parker is a Profesora/Investigadora, División de Economía, CIDE, Mexico City. Petra Todd is a Alfred L. Cass Term Professor of Economics, a PSC Research Associate at the University of Pennsylvania and an Associate of the NBER. Jorge Gallardo-García is a manager at Bates-White consulting firm and Viviana Vélez-Grajales is an associate of the Inter-American Development Bank.

Abstract

Conditional cash transfer (CCT) programs have spread worldwide as a new form of social assistance for the poor. Previous evaluations of CCT programs focus mainly on rural settings, and little is known about their effects in urban areas. This paper studies the short-term (one- and two-year) effects of the Mexican *Oportunidades* CCT program on urban children/youth. The program provides financial incentives for children/youth to attend school and for family members to visit health clinics. To participate, families had to sign up for the program and be deemed eligible. Difference-in-difference propensity score matching estimates indicate that the program is successful in increasing school enrollment, schooling attainment and time devoted to homework and in decreasing working rates of boys.

1 Introduction

Conditional Cash Transfer (CCT) programs have spread quickly internationally as a new form of social assistance for poor populations.¹ The programs aim to lessen current poverty and to promote human development by stimulating investments in health, education and nutrition, particularly over the early stages of the life cycle. CCT programs provide cash incentives for families to engage in certain behaviors, such as sending children to school and/or attending health clinics for regular check-ups, prenatal care and to receive infant nutritional supplementation. One of the earliest and largest scale CCT programs is the Mexican *Oportunidades* program (formerly called PROGRESA), in which 30 million people are currently participating. Cash benefits are linked both to regular school attendance of children and to regular clinic visits by beneficiary families. Average household benefits amount to about US\$35 to 40 dollars monthly.

The *Oportunidades* program was first initiated in 1997 in rural areas of Mexico. Impact evaluations based on a randomized social experiment carried out in rural villages demonstrated significant program impacts on increasing schooling enrollment and attainment, reducing child labor and improving health outcomes.² In 2001, the program began to expand at a rapid rate into semiurban and urban areas. Impacts in urban areas might be expected to differ from those in rural areas, both because some features of the program have changed and because access to schooling and health facilities as well as work opportunities differ. Although there have been a number of evaluations of the schooling impacts of *Oportunidades* and other CCT programs in rural areas, relatively little is known about their effectiveness

¹Programs are currently available or under consideration in over 30 countries, primarily in Latin America but also in Asia, Africa and developed countries (e.g., New York City) and have been advocated by international organizations such as the Inter-American Development Bank and the World Bank. See Becker (1999) for an early discussion of how CCT programs reduce the use of child labor and Fiszbein et. al. (2009) for a recent survey of CCT programs that have been implemented in a range of low and middle income countries around the world.

²See, e.g., Schultz (2000, 2004), Behrman, Sengupta and Todd (2005), Parker and Skoufias (2000), Skoufias, E. and B. McClafferty (2001), Todd and Wolpin (2006), Behrman, Parker and Todd (2011), and Parker and Behrman (2008). Also, see Buddelmeyer and Skoufias (2003) and Diez and Handa (2006) for nonexperimental approaches to evaluating the program in rural areas and for comparisons of nonexperimentally and experimentally obtained impact estimates, which are found to be similar.

in urban areas. Evaluations of CCT programs in other Latin American countries in rural areas have shown overall similar impacts to those in rural Mexico (Parker, Rubalcava, and Teruel, 2008).

Although the structure of benefits is identical in rural and urban areas, the procedure by which families become beneficiaries differs. In the rural program, a census of the targeted localities was conducted and all families that met the program eligibility criteria were informed of their eligibility status. For cost reasons, this type of census could not be carried out in urban areas, and an alternative system of temporary sign-up offices was adopted. The location of the sign-up offices was advertised and households that considered themselves potentially eligible had to visit the offices to apply for the program during a two month enrollment period when the office was open. About 40% of households who were eligible did not apply for the program, which contrasts sharply with the very high participation rates in rural areas. About one third of nonparticipating but eligible urban households report not being aware of the program. Another important difference between the initial rural program and the current program is that the schooling subsidies were initially given for attendance in grades 3 through 9, but were extended in 2001 to grades 10 to 12 (high school). We might therefore expect to see impacts of the program on older children as well as potentially larger impacts on younger children than observed under the original rural program.

This paper evaluates the effectiveness of *Oportunidades* in influencing schooling and working behaviors of adolescent youth using observational data. Specifically, it examines how participation in the program for two years affects schooling attainment levels, school enrollment rates, time devoted to homework, working rates and monthly wages. Our analysis focuses on girls and boys who were 6-20 years old in 2002, when the *Oportunidades* program was first introduced in some urban areas. The treated group are households who were eligible for the program and elected to participate during the two month incorporation window. The comparison group consists of eligible households living in localities where the program was not yet available. These localities were preselected to be otherwise similar to the treatment localities in terms of observable average demographics and in terms of other characteristics,

such as regional location, availability of schools and health clinics. Nonetheless, in making comparisons between individuals in the treated and comparison groups, it is important to account for program self-selection to avoid the potential for selection bias. We find that participating households differ from nonparticipating households; for example, households who expect a larger gain from participation in terms of subsidies are more likely to seek incorporation. The methodology that we use to take into account nonrandom program selection is difference-in-difference matching (DIDM). Matching methods are widely used in both statistics and economics to evaluate social programs.³ The DIDM estimator compares the change in outcomes for individuals participating in the program to the change in outcomes for matched nonparticipants. The advantage of using a DIDM estimator instead of the more common cross-sectional matching estimator is that the DIDM estimator allows for time-invariant unobservable differences between the participant and nonparticipant groups that may affect participation decisions and outcomes. Such differences might plausibly arise, for example, from variation across regional markets in earnings opportunities. For comparison, we also report impact estimates based on a standard difference-in-difference estimator.

This paper makes use of data on three groups: (a) households who participated in the program, (b) households who did not participate but were eligible for the program and lived in intervention areas, and (c) households living in areas where the program was not available but who otherwise met the program eligibility criteria. Groups (a) and (b) are used to estimate a model of the program participation decision, and groups (a) and (c) are used for impact evaluation. Importantly, group (b) is not used as a comparison group for the purposes of impact evaluation because of the potential for spillover effects of the program onto nonparticipating households in intervention areas. For example, if the school subsidy component of the program draws large numbers of children participating in the program into school, then it can lead to increased pupil-teacher ratios that represent a deterioration

³See, e.g. Rosenbaum and Rubin (1992), Heckman, Ichimura and Todd (1997), Lechner (2002), Jalan and Ravallion (2003), Behrman, Cheng and Todd (2004), Smith and Todd (2005), Galiani, Gertler, and Schargrodsky (2005).

in school quality for children who are not directly participating.⁴ However, by drawing the comparison group instead from geographic areas that had not yet received the intervention and assuming that the same program participation model applies to those areas, matching estimators can still be applied. Specifically, we apply the participation model estimated from groups (a) and (b) to predict which of the households in (c) would likely participate were the program made available.

Our analysis samples come from the Urban Evaluation Survey data that were gathered in three rounds. One round is baseline data, gathered in the fall of 2002 prior to beneficiary households beginning to receive program benefits. The 2002 survey gathered information pertaining to the 2001-2002 school year. The other two rounds were gathered post-program initiation in the fall of 2003 and fall of 2004 after participating households had experienced one and two years in the program and pertain to the 2002-2003 and 2003-2004 school years. In all rounds, data were gathered on households living in both intervention and nonintervention areas.

The paper proceeds as follows. Section 2 briefly summarizes the main features of the *Oportunidades* program. Section 3 describes the parameters of interest and the estimation approach. Section 4 discusses the data and presents mean comparisons across the various analysis samples. Section 5 describes how we estimate the model for program participation. Section 6 presents estimated program impacts on education and work and Section 7 concludes.

2 Description of the *Oportunidades* Program

The *Oportunidades* program is targeted at poor families as measured by a marginality index summarizing characteristics of the households, such as education, type of housing and assets. Households with index scores below a cutoff point are eligible to participate. The program has two main subsidy components: a health and nutrition subsidy and a schooling subsidy.

⁴For evidence of spillovers in the rural program, see Angelucci and De Giorgi (2009), Bobonis and Finan (2009) and Angelucci et. al. (2010). Angelucci and Attanasio (2009) discuss similar concerns in their study of *Oportunidades*' impact on urban consumption.

To receive the health and nutrition benefit, household members have to attend clinics for regular check-ups and attend informational health talks. To receive the school subsidy, children or youth in participating households have to attend school in one of the subsidy-eligible grade levels (grades 3-12) for at least 85% of school days. They cannot receive a subsidy more than twice for the same grade.

Because of these conditionalities, one would expect to find positive program impacts on school enrollment and schooling attainment. However, the magnitude of these impacts is not predictable, because households can participate to different degrees in the program. For example, households can choose to participate only in the health and nutrition component, so it is possible that the program has no impact on schooling for some households. Second, households can choose to send only a subset of their children to school for the required time. For example, rather than send two children to school fifty percent of the time, households might send their older child to school 85% of the time (to receive the subsidy) and keep the other child at home. Thus, children within the household may be affected by the program in different ways. The goal of this paper is to assess the magnitude of the *Oportunidades* impacts on a variety of school and work-related outcome measures. Some of these outcomes, such as school enrollment, schooling attainment and grade accumulation, are closely tied to program coresponsibilities, while other outcomes, such as time spent doing homework, whether parents help with schoolwork, and employment and earnings, are not.⁵

Table 1 shows the subsidy amounts and how they increase with grade level, a design that attempts to offset the higher opportunity costs of working for older children. The slightly higher subsidies for girls reflect one of the emphases of the program, which is to increase the schooling enrollment and schooling attainment for girls, who in Mexico traditionally have lower enrollment in post-primary grade levels.

⁵The estimates here represent the total impact of the program on schooling; we are unable to isolate, for instance, the impact of the grants linked to schooling from, say, potential health effects on schooling from the regular clinic visits.

3 Parameters of Interest and Estimation Approach

The general evaluation strategy used in this paper compares outcomes for program participants (the treatment group) to those of a comparison group that did not participate in the program. Data were gathered on households living in urban areas served by *Oportunidades* (intervention areas) and on households living in urban areas that had not yet been incorporated into *Oportunidades* (non intervention areas), as described further below in section four. Baseline data (year 2002) are available as well as two years of follow-up data (year 2003 and 2004). We distinguish among the following three groups of households:

(a) Participant Group - Households that applied to the program, were deemed eligible for it and chose to participate.

(b) Eligible Non-Participant Group - Households living in program intervention areas that did not apply to the program but were eligible for it.

(c) Outside Comparison Group – Households living in localities where *Oportunidades* was not yet available but who otherwise meet the eligibility criteria for the program. These localities were planned for incorporation after the 2004 data collection, but the households living in these localities did not know about plans for future incorporation. We refer to group (a) as the treatment group and groups (b) and (c) as the comparison groups.

The main parameter of interest in this paper is the *average impact of treatment on the treated*, for various subgroups of children/youth defined by age, gender and preprogram poverty status. Let Y_1 denote the potential outcome of a child/youth from a household that participates in the program and let Y_0 denote the potential outcome if not exposed to the program. The treatment effect for an individual is:

$$\Delta = Y_1 - Y_0.$$

Because no individual can be both treated and untreated at the same time, the treatment effect is never directly observed. This paper uses matching methods to solve this missing data problem and focuses on recovering mean program impacts.⁶

⁶See Heckman, Lalonde and Smith (1999) for a discussion of other possible parameters of interest in

Let Z denote a set of conditioning variables. Also, let $D = 1$ if an individual is from a participating household, $D = 0$ if not. The average impact of treatment on the treated, conditional on some observables Z , is:

$$E(\Delta|Z, D = 1) = E(Y_1 - Y_0|Z, D = 1) = E(Y_1|Z, D = 1) - E(Y_0|Z, D = 1)$$

We can use the treatment data to estimate $E(Y_1|Z, D = 1)$, but the data required to estimate $E(Y_0|Z, D = 1)$ are missing.

Estimators developed in the evaluation literature provide different ways of imputing $E(Y_0|Z, D = 1)$. Traditional matching estimators pair each program participant with an observably similar nonparticipant and interpret the difference in the outcomes of the matched pairs as the treatment effect (see, e.g., Rosenbaum and Rubin, 1983). Matching estimators are typically justified under the assumption that there exists a set of observable conditioning variables Z for which the nonparticipation outcome Y_0 is independent of treatment status D conditional on Z :

$$Y_0 \perp\!\!\!\perp D|Z. \tag{1}$$

It is also assumed that for all Z there is a positive probability of either participating ($D = 1$) or not participating ($D = 0$), i.e.

$$0 < \Pr(D = 1|Z) < 1, \tag{2}$$

so that a match can be found for all treated ($D = 1$) persons. Under these assumptions, after conditioning on Z , the Y_0 distribution observed for the matched nonparticipant group can be substituted for the missing Y_0 distribution for participants.⁷ If there are some observations for which $\Pr(D = 1|Z) = 1$ or 0, then it is not possible to find matches for them.

evaluating effects of program interventions.

⁷If the parameter of interest is the mean impact of treatment on the treated (TT), then (1) and (2) can be replaced by weaker assumptions:

$$E(Y_0|Z, D = 1) = E(Y_0|Z, D = 0) = E(Y_0|Z),$$

and

$$\Pr(D = 1|Z) < 1.$$

See Heckman, Ichimura and Todd (1997) for discussion on this point.

Matching methods can be difficult to implement when the set of conditioning variables Z is large.⁸ To address this difficulty, Rosenbaum and Rubin (1983) provide a useful theorem that shows that when matching on Z is valid, then matching on the propensity score $\Pr(D = 1|Z)$ is also valid. Specifically, they show that assumption (1) implies

$$Y_0 \perp\!\!\!\perp D | \Pr(D = 1|Z).$$

Provided that the propensity score can be estimated at a rate faster than the nonparametric rate (usually parametrically), the dimensionality of the matching problem is reduced by matching on the univariate propensity score. For this reason, much of the recent evaluation literature focuses on propensity score-based methods.⁹

As described above, cross-sectional matching estimators assume that mean outcomes are conditionally independent of program participation. However, for a variety of reasons, there may be systematic differences between participant and nonparticipant outcomes that remain even after conditioning on observables. Such differences may arise, for example, because of program selectivity on unmeasured characteristics, or because of level differences in outcomes across different geographic locations in which the participants and nonparticipants reside.

A difference-in-differences (DIDM) matching strategy, as proposed and implemented in Heckman, Ichimura and Todd (1997), allows for temporally invariant sources of differences in outcomes between participants and nonparticipants and is therefore less restrictive than cross-sectional matching.¹⁰ It is analogous to the standard DID regression estimator but does not impose the linear functional form restriction in estimating the conditional expectation of the outcome variable and reweights the observations according to the weighting functions used by the matching estimator. Letting t' denote a pre-program time period and

⁸If X is discrete, small (or empty) cell problems may arise. If X is continuous, and the conditional mean $E(Y_0|D = 0, X)$ is estimated nonparametrically, then convergence rates will be slow due to the "curse-of-dimensionality."

⁹Heckman, Ichimura and Todd (1998), Hahn (1998) and Angrist and Hahn (2001) consider whether it is better in terms of efficiency to match on $P(X)$ or on X directly. For the TT parameter considered in this paper, neither is necessarily more efficient than the other (see discussion in Heckman, Ichimura and Todd (1998)).

¹⁰Smith and Todd (2005) find better performance for DIDM estimators than cross-sectional matching estimators in an evaluation of the U.S. National Supported Work Program.

t a post-program time period, the DIDM estimator assumes that

$$Y_{0t} - Y_{0t'} \perp\!\!\!\perp D \mid \Pr(D = 1 \mid Z),$$

or the weaker condition

$$E(Y_{0t} - Y_{0t'} \mid \Pr(D = 1 \mid Z), Z) = E(Y_{0t} - Y_{0t'} \mid D = 0, \Pr(D = 1 \mid Z))$$

along with assumption (2) (the support condition), which must now hold in both periods t and t' (a non-trivial assumption given the attrition present in many panel data sets). The local linear DIDM estimator that we apply in this paper is given by

$$\begin{aligned} \hat{\alpha}_{DD} = & \frac{1}{n_1} \left(\sum_{i \in I_1 \cap S_P} (Y_{1ti} - Y_{0t'i}) \right. \\ & \left. - \sum_{j \in I_0 \cap S_P} W(i, j) (Y_{0tj} - Y_{0t'j}) \right), \end{aligned}$$

where I_1 denotes the set of program participants, I_0 the set of non-participants, S_P the region of common support and n_1 the number of persons in the set $I_1 \cap S_P$. The match for each participant $i \in I_1 \cap S_P$ is constructed as a weighted average over the difference in outcomes of multiple non-participants, where the weights $W(i, j)$ depend on the distance between P_i and P_j . Appendix A gives the expression for the local linear regression weights and also describes the method that we use to impose the common support requirement.¹¹

Estimating the propensity score $P(Z)$ requires first choosing a set of conditioning variables, Z . It is important to restrict the choice of Z variables to variables that are not influenced by the program, because the matching method assumes that the distribution of the Z variables is invariant to treatment. As described below, our Z variables are baseline characteristics of persons or households that are measured prior to the introduction of the program. We estimate the propensity scores by logistic regression.

¹¹Research has demonstrated advantages of local linear estimation over simple kernel regression estimation. These advantages include a faster rate of convergence near boundary points and greater robustness to different data design densities. See Fan (1992a,b). Local linear regression matching was applied in Heckman, Ichimura and Todd (1998), Heckman, Ichimura, Smith and Todd (1998) and Smith and Todd (2005). We do not report results based on nearest neighbor matching estimators, in part because Abadie and Imbens (2005) have shown that bootstrap standard errors are inconsistent for nearest neighbor matching estimators. Their critique does not apply to the local linear matching estimators used here.

Lastly, the propensity score model is estimated from groups (a) and (b) and then applied to groups (a) and (c) to determine the impacts, our matching procedure assumes that the model for program participation decisions is the same for intervention areas as in nonintervention areas. However, the DIDM estimator allows for households in nonintervention localities to have a potentially different distribution of observed characteristics Z .¹²

4 Data Subsamples

We use the Urban Evaluation Survey (Encelurb) baseline carried out in the fall of 2002 and two follow-up rounds carried out in the fall of 2003 and 2004. The baseline was carried out before the program was advertised and before any monetary benefits had been received and collected information pertaining to the 2001-2002 school year. The Encelurb data include rich information on schooling and labor force participation, as well as information on income, consumption, assets and expectations about future earnings. The database contains information on 2972 households in the eligible participant group (group a), 2556 households who were eligible but did not participate (group b) and 3607 households from nonintervention areas who satisfied the eligibility criteria for the program (group c). Appendix B provides further information on the selection of the subsamples used in our analysis. As also described in the appendix, some observations had to be eliminated from our analysis because of inconsistencies in reported ages and schooling levels across the survey rounds. Appendix B details how we constructed variables for our analysis and how we identified and dealt with different kinds of data inconsistencies. Table 2 shows the sample sizes for groups (a), (b) and (c), comprised of boys and girls age 6-20 in 2002 (age 8-22 in 2004).¹³

¹²The assumption that the participation model is the same across intervention and nonintervention areas is untestable, as we cannot observe participation decisions in nonintervention areas.

¹³These are the sample sizes prior to deleting observations for reasons of attrition across rounds and missing data. For details on how the final analysis samples were obtained, see Appendix B.

4.1 Interpretation of the estimator in the presence of attrition

In this study, we focus on children and youth that were observed across all the years of the survey. Table 3 provides information on the number of children in groups (a) and (c) (the groups used for the impact analysis) by age, survey round and control versus treatment status. The percentage of children who are lost between rounds tends to increase with age from 17-20% for 6-7 year-olds and to 35-37% for the oldest age category (19-20 year-olds) for 2002-2004. The higher attrition rate for the oldest group is not surprising given that children typically leave home and get married at earlier ages in Mexico than in higher per capita income countries. Of more relevance for our impact analysis is whether the patterns differ substantially between the treatment and the control groups. Between 2002 and 2004 there are slightly higher attrition rates for most age groups (except for 15-18 year-olds) and on average for the control (22%) than the treatment (19%). But the overall difference (3%) and the differences for each of the age groups (-1% to 4%) are fairly small, and therefore not likely to affect substantially the estimates presented below. Nevertheless, we describe below the extent to which our estimation method accommodates attrition.

As described in the previous section, our impact analysis aims to recover the average effect of treatment on the treated. In our application, the treated correspond to children observed over the three survey rounds. For the DIDM estimator to be valid in the presence of some attrition, we require an additional assumption on the attrition process. Let $A = 1$ if the individual attrited between time t' and t , else $A = 0$ and, for ease of notation, denote $\Delta Y = Y_{0t} - Y_{0t'}$. By Bayes rule,

$$\begin{aligned} \Pr(D = 1|A = 0, \Delta Y, Z) &= \frac{\Pr(A = 0|D = 1, \Delta Y, Z) \Pr(D = 1|\Delta Y, Z)}{\Pr(A = 0|\Delta Y, Z)} \\ &= \frac{\Pr(A = 0|D = 1, \Delta Y, Z) \Pr(D = 1|\Delta Y, Z)}{\Pr(D = 1|\Delta Y, Z) \Pr(A = 0|\Delta Y, Z, D = 1) + \Pr(D = 0|\Delta Y, Z) \Pr(A = 0|\Delta Y, Z, D = 0)} \end{aligned}$$

Under the DIDM assumption,

$$\Pr(D = 1|\Delta Y, Z) = \Pr(D = 1|Z).$$

We invoke an additional assumption that attrition is not selective on the change in potential

outcomes in the no-treatment state conditional on Z and D , then

$$\Pr(A = 0|\Delta Y, D, Z) = \Pr(A = 0|D, Z)$$

we get

$$\Pr(D = 1|\Delta Y, Z, A = 0) = \frac{\Pr(A = 0|D = 1, Z) \Pr(D = 1|Z)}{\Pr(A = 0|Z)} = \Pr(D = 1|Z, A = 0),$$

which would imply that the DIDM estimator is valid for nonattriters. Attrition is allowed to depend on time-invariant unobservables that are eliminated by the differencing, so attrition on unobservables is permitted.¹⁴ However, the estimation approach does rule out certain types of attrition, for example, that individuals with particularly high or low differences in potential earnings or schooling outcomes (that are not explainable by the observed characteristics Z) are the ones who attrit. Also, the application of the DID matching estimator does not require assuming that attrition is symmetric in the treatment and comparison groups. The method would allow, for example, for the case of no attrition in the treated group and substantial attrition in the comparison group, as long as the attrition in the comparison group is not selective on ΔY , conditional on Z .

4.2 Descriptive analysis and comparison of groups at baseline

In addition to the usual school enrollment indicators, the Encelurb includes information on time spent doing homework. The labor module includes characteristics of both the primary and secondary job of all individuals over age 5 for the previous 12 months and, for the household head and spouse, information covering labor history for the past 24 months. Our empirical analysis focuses on schooling attainment, school enrollment, homework time, employment, and monthly earnings.

Figures 1a,b to 6a,b depict the average outcome variables by age and gender of the child/youth for the participant group (a) and the eligible nonintervention group (c). The

¹⁴Also, attrition can be selective to some extent on treatment outcomes because it can depend on Y_{1t} . For example, those who experience the lowest Y_{1t} could be the ones who attrit. In that case, the DIDM estimator would still be valid for the subgroup of nonattriters, but the average treatment effect for nonattriters would not generally be representative of the average treatment effect for the population that started in the program.

outcome variables are measured prior to the program intervention, so they are informative about pre-program group differences and not about impacts of the program. Figure 1a graphs the schooling attainment of boys. For both the participant and nonintervention group, schooling attainment increases steadily with age until around age 15 when the rate of increase slows down. Schooling attainment reaches an average of 8 grades at age 20. The nonintervention group has slightly higher schooling attainment levels at each age until about age 14, after which the averages by group are roughly the same. For girls (Figure 1b), the pattern of increase in schooling attainment by age is similar to that of boys, with the comparison group having higher schooling attainment levels for all of the ages.

The fact that the schooling attainment levels are on average different between groups (a) and (c) in 2002, prior to implementation of the program, shows the importance of taking into account preprogram differences in evaluating program impacts. The DIDM estimator compares the change in schooling attainment for children in group (a), whose families participated in the program, to the change in schooling attainment for children from matched families in group (c).

Figures 2a and 2b graph school enrollment rates in the 2001-2002 baseline school year for the same groups. For both the participant and the nonintervention groups, enrollment rates are near 100% over the age range associated with primary school. After that, enrollment rates steadily decline. For girls, the enrollment rates of participants are lower than those of the nonintervention group over most of the age range, but for boys the enrollment rate for participants is higher at ages older than 12. Our impact evaluation compares the change in enrollment rates (post-program minus pre-program) for children in group (a) to the change in enrollment rates for matched children from the nonintervention group (c).

Figures 3a and 3b graph the percentage of children who receive help at home in doing their homework. The general pattern is that the percentage helped increases at first with age, peaks at about age 7-9 and then declines. For young children, age 6, the nonintervention group reports a higher percentage helped on homework, but for other ages the participants and the nonintervention group are fairly similar in terms of homework help patterns. Figures

4a and 4b show the number of weekly hours spent by boys and girls doing homework. The figures show that adolescent girls spend on average more time on homework than similar aged boys. A comparison of program participants and eligibles from the nonintervention group shows that they are fairly similar preprogram in terms of reported time spent on homework. Figures 5a and 5b compare the employment rates of boys and girls age 12-20 in the program participant and eligible nonintervention group at baseline. A child is defined as employed if he/she reports working for pay. The boys in the two groups report fairly similar rates of employment. However, for girls the employment rates are clearly higher in the participant than in the nonparticipant group. Lastly, Figures 6a and 6b exhibit the average income earned by boys and girls in the two groups. Children who are not working are included in the average as earning zero income. Participant girls have on average higher income than girls in the eligible nonintervention group, consistent with their higher rates of working.

5 Propensity Score Estimation

The propensity score DIDM method requires first estimating $\Pr(D = 1|Z)$, which we estimate using logistic regression applied to data on households who are eligible to participate and live in areas where the program is available, that is groups (a) and (b) described in Section 2 of this paper. Participating in the program ($D=1$) means that a household member attended the module to sign up for the program (see Section 1) and that the household was deemed eligible and elected to participate. We use administrative data to determine which households are or are not participating.

An important question in implementing any evaluation estimator is how to choose the set of conditioning variables Z . We include in Z variables that are designed to capture key determinants of program participation decisions, such as measures of the benefits the family expects to receive from the program (based on the number of children in the family of various ages attending school), variables related to costs associated with signing up and participating in the program (such as poverty status of the family, employment status of family members, availability of schools and health clinics), and variables that potentially affect whether the

family is aware of the program (such as geographic location, population living in the locality, average poverty score of people living in the locality, schooling of household members). All of the variables are measured prior to program initiation and are therefore unlikely to be influenced by the program. We chose the set of regressors to maximize the percentage of observations that are correctly classified under the model.¹⁵

Table B.1 in appendix B presents the estimated coefficients of the program participation model estimation along with their standard errors. 73% of the observations are correctly classified under the estimated model as being either a participant or nonparticipant. In general, most of the coefficients are statistically significant at the 10 percent level. The results show that having a dirt floor or walls or ceilings made of provisional materials (cardboard, plastic, tires, etc.) increases the probability of participating in the program. Having piped water inside the house decreases the probability of participating, although this coefficient is small in magnitude. Having a truck or a refrigerator also decreases the probability of participation in the program, which is consistent with these goods being indicators of higher economic status for a household.

Participating in some other in-kind subsidy program has a positive effect on the probability of participation. The other relevant programs operating under the Mexican social support system are the free tortilla program, the milk subsidy program, the free groceries program, and the free school breakfast program.¹⁶

The number of children of ages that make them eligible for attending primary or secondary school increases the probability of program participation, which is to be expected as

¹⁵This model selection procedure is similar to that used in Heckman, Ichimura and Todd (1997). That paper found that including a larger set of conditioning variables in estimating the propensity score model improved the performance of the matching estimators. We do not use so-called balancing tests as a guide in choosing the propensity score model in part because of the very large number of regressors. Balancing tests cannot be used in choosing which variables to include in the model (see the discussion of the limitations of such tests in Smith and Todd (2007)), but they are sometimes useful in selecting a particular model specification for the conditional probability of program participation, given a set of conditioning variables. For example, they provide guidance as to whether to interact some of the regressors. However, with the large number of variables included in our propensity score model, the set of possible interactions is huge.

¹⁶Participating in the PROCAMPO program decreases the probability of participation, because this program focuses on rural areas and being eligible for this rural program would mean that the house is located far from the *Oportunidades* urban incorporation modules.

their presence increases the household's potential benefits from the school enrollment transfers. Being a female headed household also increases the participation probability. Being poor, as classified by the *Oportunidades* poverty score, is associated with higher rates of program participation. Poorer households may be more likely to attend the sign-up offices, either because they live in high density poverty areas where information about the program is better disseminated or because they are more likely to believe that their application to get *Oportunidades* benefits will be successful.

Figures 7a and 7b show the histograms of the propensity scores that are derived for each household from the logit model estimates. The figures compare propensity scores for participant households (group a) to those for nonintervention households (group c). The histograms for both groups are similar. Most importantly, the figures reveal that there is no problem of nonoverlapping support. For each participant household, we can find close matches from the nonintervention group of households (as measured by similarity of the propensity scores).

6 Program Impacts

We next evaluate the impacts of participating in the program on schooling and work. Given the relatively high school enrollment levels for young children and lower levels for older children prior to the introduction of the program, we expect the impacts on school enrollment to be greater for older children. Because the program requires regular attendance, we also expect to see improvements in grade progression as measured by educational attainment. The school attendance requirements would be expected to increase the amount of time spent in school relative to work, so in addition we expect the program to decrease the working rates of children/youth, with a corresponding decrease in earnings. Our expectations for the one-year impact analysis are guided in part by the results of the short-term rural evaluation, which found no effects on primary school enrollment levels but statistically significant effects

on increasing enrollment at secondary grades.¹⁷

As described previously, our evaluation strategy compares children/youth from households who participate in the program (group a) to observably similar (matched) children/youth from households living in the areas where the program was not yet available (group c). We impute propensity scores for the households in the nonintervention areas using the estimated logistic model, as described in Section 4 and Appendix B. For these households, the estimated propensity scores represent the probability of the household participating in the program had the program been made available in their locality. Because the propensity score model is estimated at the household level, all children within the same household receive the same score. However, the impact estimation procedure allows children of different gender and age groups from a given household to experience different program impacts.

Lastly, the program subsidies start at grade 3 (primary level), so 6 and 7 year-old children do not receive benefits for enrolling in school. However, we include these children in our impact analysis, because they may benefit from the subsidies received by older siblings and/or the expectation of cash payments at higher grades could provide an incentive to send younger children to school, as was found in rural areas (Behrman, Sengupta and Todd 2005).

6.1 Schooling Attainment

Tables 4a and 4b show the estimated program impacts on total grades of schooling attainment for boys and girls who are age 6 to 20 at the time of the 2002 survey (8-22 at the time of the 2004 survey). The columns labeled "Avg Trt 2002" and "Avg Con 2002" give the average schooling attainment at baseline (year 2002) for children in treatment and comparison households, respectively, and show that controls had slightly higher grades of attainment in that year. The program can achieve increases in grades of schooling attainment through earlier matriculation into school, lower rates of grade repetition and lower rates of dropping out. The columns labeled "1 Year DID Match" and "2 Year DID Match" give the one-year

¹⁷See Behrman, Sengupta and Todd (2005), Parker, Behrman and Todd (2004), Parker and Skoufias (2004), and Schultz (2000).

and two-year program impact estimates obtained through DID matching, where we match on propensity scores after exact matching by age and gender. Also, we do not impose any restriction on the estimation that the two-year impacts exceed the one-year impact. The program might over the short run increase school attendance and educational attainment, but over the longer run might lead to lower attendance at older ages if youth attain their targeted schooling levels at earlier ages. For example, suppose that before the program, a child's targeted schooling level was ninth grade and it took eleven years to complete ninth grade (because some grades were repeated). The program might encourage the child to attend school more regularly and to increase the targeted schooling level from ninth to tenth grade. If as a result of the program the child completes ten grades in ten years as opposed to nine grades in eleven years, then the child finishes school at an earlier age with more grades completed. In this sense, schooling attainment could increase without necessarily increasing enrollment at older ages.¹⁸

The columns labeled "1 Year DID" and "2 Year DID" give impact estimates obtained using an unadjusted DID estimator where as the next two columns contain the DID matching estimates. The one-year impacts correspond to the 2003-2004 school year and two-year impacts to the 2004-2005 school year. As seen in the columns labeled "% Obs in Support," the supports for the most part overlap and not many of the treated observations were eliminated from the analysis for support reasons. Each of the tables presents impact estimates obtained using a bandwidth equal to 0.3. The bandwidth controls how many control observations are used in constructing the matched outcomes. As the estimates may depend on bandwidth choice and there are a variety of criteria for choosing bandwidths, it is important to consider a range of choices. We found that the program impact estimates are relatively insensitive to bandwidth choices in the range of 0.2-0.4, so for the sake of brevity we only report results for one bandwidth.¹⁹

For boys (Table 4a), there is strong evidence of an effect of the program on schooling

¹⁸Behrman, Parker and Todd (2011) find for longer program exposure (up to six years) in rural areas that the impacts on schooling attainment are nondecreasing.

¹⁹Results on the full set of bandwidths tried are available upon request.

attainment at ages 8-18, with the two-year impact (2003-2004) being about 40% larger than the one-year impact. For girls (Table 4b), one year and two-year program impacts are statistically significant over ages 6-14. The two-year impacts for girls are generally of similar magnitude to the one-year impacts and are somewhat lower than for boys, despite the higher subsidy rates for girls. For both girls and boys, the largest impacts are observed in the 12 to 14 age range (approximately 0.20 grades, which represents about a 4% increase in schooling attainment). The estimated impacts are largest for boys age 12 to 14, ranging from 0.23 to 0.30 grades. These ages coincide with the transition from primary to secondary school, when preprogram enrollment rates dropped substantially and when the subsidy level increases substantially (See Table 1). The DIDM and the simple DID estimates are often similar. The similarity of the estimates, however, does not imply that self-selection into the program is not an issue. As discussed in Heckman, Ichimura and Todd (1997), selection biases can sometimes fortuitously average to close to zero.

Tables 5a and 5b show the estimated impacts on school enrollment for boys and girls. The third and fourth columns (labeled Avg Trt in 20002 and Avg Con in 2002) give the enrollment rates of the children in the baseline 2001-2002 school year. The enrollment rates are usually over 90% for ages up to age 14, after which there is a significant drop. We observe statistically significant impacts of the program on enrollment rates mainly for boys and girls ages 6-14. Both the one-year and two-year impacts are larger for girls age 6-7 than for similar age boys. The impacts for the age 8-11 group are similar across boys and girls and are roughly of the same magnitude (around 2-3 percentage point increases in enrollment, which represents an increase of about 2-3%). The two-year impact is not necessarily larger than the one-year impact, but this is to be expected given that children are aging over the two years and enrollment rates generally decline with age.

Tables 6a and 6b examine the estimated impacts of the program on the percentage of children whose parents help them on their schoolwork, for the subsample of children reported to be in school at the baseline year. The potential impact of the program on parents helping with homework is theoretically difficult to predict. On the one hand, the

subsidies may provide an incentive for parents to help their children so that they pass grades levels according to the pace required under the program rules. On the other hand, if the subsidies enable children/youth to work less (outside of school) and spend more time on schoolwork, then they may not require as much help. As children/youth increase school enrollment and decrease their time spent working at home or outside the home, parents may also decide to increase their time spent working and so may have less time available to help children on their schoolwork.

As seen in Table 6a, for boys there are statistically significant impacts only at ages 8-11, when parents in households participating in the program have a rate of helping that is lower by around 10 percentage points in the first year and 12 percentage points in the second year. Similar negative impacts on homework help are observed for girls, who also show lower two-year impacts on homework help at ages 8-11, though the estimates for girls are for the most part statistically insignificantly different from zero. The general pattern of a decrease in help on homework is consistent with parents in participating households spending more time on work activities to substitute for their children working less (see discussion of employment impacts below).

Tables 7a and 7b examine the impacts of the program on the amount of time that boys and girls dedicate weekly to their studies, for the subsample of children reported to be in school in the 2001-2002 school year. For boys, the average number of hours devoted to homework at the baseline is increasing with age and is in the range of 5-8 hours. The largest one-year and two-year impacts of the program are observed for boys age 12-14 (in 2002), who devote an additional 0.85-1.3 hours to homework, which represents an increase of about 14- 21% over their baseline hours. We also observe statistically significant one-year impacts for boys age 6-7 and for boys age 8-11, who devote an additional 20 to 50 minutes per week on homework. For girls of all ages, however, the estimated two-year impacts are statistically insignificant.

Tables 8a and 8b show the estimated impacts of the program on the percentage of children working for pay. The tables only show estimates for ages 12-20, because very few children

work at younger ages, probably in part because of legal restrictions on work for children younger than age 14. As seen in the third and fourth columns, showing the baseline work patterns, the percentage of children working for pay increases sharply with age. From age 12-14 to 15-18, the percentage increases from 10% to 44% for boys and 10% to 27% for girls. Participation in the program is associated with a reduction in the percentage of boys age 12-14 working of about 8 percentage points in the first year and 12-14 percentage points in the second year. For girls, the percentage of youth who work is lower at all ages than for boys, and the program does not seem to affect the percentage of girls who work. The only statistically significant impact for girls is observed at ages 15-18, when the program lowers the percentage working by around 11%, but only in the first year. Thus the gender difference in working and in time spent on homework are consistent with significant shifts away from market work to homework for boys but no significant changes in either for girls.

Lastly, Tables 9a and 9b present estimates of the program's impact on monthly earnings of children and youth 12 to 20 years old, where earnings are set to zero for children who report not working. The average earnings combines information on the proportion working and on the average earnings of those who work. We observe an impact of the program on earnings only for boys age 12-14, for whom earnings decline in the first year after the program started. The other estimates are statistically insignificantly different from zero. This result is not necessarily inconsistent with the statistically significant effects on employment noted above, as it is possible that the decrease in the supply of child labor induced by the program increases the wages of the children who remain in the labor market.

7 Conclusions

This paper analyzes the one- and two-year impacts of participation in the *Oportunidades* program on schooling and work of children/youth who were age 6-20 in 2002 (8-22 in 2004) who live in urban areas and whose families were incorporated into *Oportunidades* program in 2002. An important difference between the implementation of the program in urban areas in comparison to rural areas is that families in urban areas had to sign up for the program at a

centralized location during an enrollment period. Only a subset of the households eligible for the program chose to participate. Another important difference is that the data collection in urban areas is nonexperimental, and the estimators used to evaluate program impacts must therefore take into account nonrandomness in the program participation process.

This paper makes use of data on three groups: (a) a group of eligible households who participate in the program, (b) a group of eligible households who do not participate but who live in intervention areas, and (c) a group of eligible households who live in areas where the program was not available. The difference-in-difference matching estimator compares the change in outcomes for individuals participating in the program with the change in outcomes for similar, matched individuals who are not in the program.

Our analysis finds significant positive impacts for both boys and girls on schooling attainment, school enrollment, proportions working, and amount of time children spend doing homework. In addition, we find a statistically significant negative impact of the program on the percentage of children whose parents help them with their homework and no discernible impact on average earnings.

Overall, the education impacts deriving from one and two years of program operation are sizeable. They are generally encouraging, given that the data come from the initial phase of expansion to urban areas, a time period where one might expect operational problems.²⁰ The impacts thus far are consistent with the program, at least in the area of education, being carried over to urban areas in a successful manner. Note that the parameter estimated in this paper was average treatment on the treated. An intent-to-treat estimator, which would include eligibles who did not apply to the program, would indicate smaller size impacts.

One concern in implementing the *Oportunidades* school subsidies in urban areas is that the level of subsidies in urban areas was identical to rural areas, but the opportunity costs of school for children/youth from urban areas are likely to be greater because of greater earnings potential. It is therefore remarkable that the impact estimates for the urban areas

²⁰However, there may also be opposing pioneer effects as discussed in Behrman and King (2008) and King and Behrman (2009).

are for the most part similar to results found in the rural evaluation in the initial phases of the program. A difference, however, in the urban areas compared with the initial rural implementation is that school subsidies are now offered at grades 10-12, which provides incentives for older children to enroll in school and may also provide extra incentives for younger children to reach those grade levels.

With respect to gender, we found fairly similar impacts for boys and girls on schooling outcomes, which suggests that the slightly higher grants given to girls attending grades 7-12 are not generating large differences in impacts by gender (although one cannot say in the current analysis whether impacts on girls would be different if grant amounts were changed to be the same for girls as boys). Prior to the program, enrollment rates for girls in the participant groups started to fall at earlier ages than those for boys (See figures 2a and 2b), which could justify the higher grants to girls for secondary school grades. However, there do not appear to be strong differences in education levels between boys and girls (see figures 1a and 1b). An evaluation of the short-term impacts in rural areas (Behrman, Sengupta and Todd, 2005) demonstrated that girls tend to progress more quickly than boys through the primary grades (i.e. have lower failure rates), which can explain why their completed average schooling is higher than that of boys despite leaving school on average at earlier ages.²¹

Boys report spending more time doing homework as a result of the program, so it is perhaps surprising that they seem to be receiving less help from parents/other relatives with the program than they would without. As described earlier, this may reflect either that they need less help or that other adults are working more in response to the program (perhaps as a response to their children working less) and thus have less time to spend during after school hours with their children. In any case, the impact of the program on the time allocation of different family members seems an important topic for further research.

In addition to program impacts on schooling, we find some evidence that the program

²¹This is a widespread pattern in developing countries. In all regions and almost all countries for which recent Demographic Health Surveys permit comparisons, though girls in some cases have lower enrollment rates, conditional on starting in school they have on average equal or higher schooling attainment (Grant and Behrman, 2010).

affects youth working behavior. Boys age 12-14 in 2002 (14-16 in 2004) experience a significant decrease in working, ranging from 7-13 percentage points. Girls' work behavior is not significantly affected by the program, in part because a lower proportion of girls than boys work for pay.

We noted that there is little scope for an effect of the program on enrollment rates at primary grades when enrollment rates preprogram were already very high. If increasing average schooling levels is a primary aim of the program, then it is worth considering further whether decreasing or eliminating the subsidies at the lower grades and using the resources to increase the level of payments to higher grade levels would be a more effective program design.²² Of course, changing the subsidy schedule in this way would have distributional consequences and would shift resources towards families whose children have higher schooling attainment levels. Further study of the urban data would be useful to evaluate the effectiveness of alternative program designs.

²²Todd and Wolpin (2006) and Attanasio, Meghir and Santiago (2005) study how shifting resources towards higher grades would affect schooling attainment in their study of *Oportunidades* in rural areas.

References

- [1] Abadie, Alberto and Imbens, Guido (2004), "On the Failure of the Bootstrap for Matching Estimators," working paper, University of California at Berkeley.
- [2] Angelucci, Manuela and Orazio Attanasio (2009): "Oportunidades: Program Effects on Consumption, Low Participation and Methodological Issues," in *Economic Development and Cultural Change*, 57:3(April), 479-506.
- [3] Angelucci, Manuela and Giacomo De Giorgi (2009): "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?," *American Economic Review*, 99(1): 486-508.
- [4] Angelucci, Manuela, Giacomo De Giorgi, Marcos A. Rangel and Imran Rasul (2010): "Family Networks and School Enrollment: Evidence from a Randomized Social Experiment," *Journal of Public Economics*, 94:3-4, 197-221.
- [5] Attanasio O, Meghir C and Santiago A. 2005. Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progreso. forthcoming in *Review of Economic Studies*, University College, London.
- [6] Becker, Gary S. (1999): "Bribe Third World Parents to Keep their Kids in School," *Business Week*, November 22nd.
- [7] Behrman, Jere R., Yingmei Cheng and Petra Todd (2004): "Evaluating Preschool Programs when Length of Exposure to the Program Varies: A Nonparametric Approach," *Review of Economics and Statistics* 86:1, 108-132
- [8] Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2009): "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico," *Economic Development and Cultural Change* , 57:3, 439-477.

- [9] Behrman, Jere R., Piyali Sengupta, and Petra Todd (2000): "Final report: The impact of PROGRESA on achievement test scores in the first year," September. Washington, D.C.: International Food Policy Research Institute, Processed. <http://www.ifpri.org/themes/progresa.htm>
- [10] Behrman, Jere R., Piyali Sengupta and Petra Todd (2005): "Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment," *Economic Development and Cultural Change*, 54:1, 237-275.
- [11] Behrman, Jere R. and Elizabeth M. King (2008): "Program Impact and Variation in Duration of Exposure," in Samia Amin, Jishnu Das and Markus Goldstein, eds. *Are You Being Served: New Tools for Measuring Service Delivery*, Washington, DC: World Bank.
- [12] Behrman, Jere R., Susan W. Parker, and Petra E. Todd (2011): "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits?," in *Journal of Human Resources*, Vol. 46: 1 (Winter), 93-122.
- [13] Bobonis, Gustavo J. and Frederico Finan (2009): "Neighborhood Peer Effects in Secondary School Enrollment Decisions," *Review of Economics and Statistics*, 91(4), 695-716.
- [14] Buddelmeyer, Hilke, and Emmanuel Skoufias (2003): "An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA," IZA Discussion Paper No. 827 (July), Institute for the Study of Labor (IZA), Bonn, Germany.
- [15] Coady, D. and S. Parker (2003): "Combining Self-Selection with Proxy Means Targeting in Mexico," Mimeo.
- [16] Diaz, Juan José and Sudhanshu Handa (2006): "An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator," in *Journal of Human Resources*, 41:2, 319-345.

- [17] Efron, Bradley and Rob Tibshirani (1993): *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- [18] Fan, J. (1992a): “Design Adaptive Nonparametric Regression,” *Journal of the American Statistical Association*, 87, 998-1004.
- [19] Fan, J. (1992b): : “Local Linear Regression Smoothers and their Minimax Efficiencies,” *The Annals of Statistics*, 21, 196-216.
- [20] Fan, J. and Gijbels, I. (1996): *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.
- [21] Fiszbein, A., Shady, N., Ferreira, F., Grosh, M., Kelleher, N., Olinto, P. and E. Skoufias. (2009): *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington, DC: The World Bank.
- [22] Heckman, James, Hidehiko Ichimura and Petra Todd (1997): “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies*, 64:4, 605-654.
- [23] Galiani, Sebastian, Gertler, Paul, and Ernesto Schargrotsky (2005): ”Water for Life: The Impact of the Privatization of Water Services on Child Mortality in Argentina,” *Journal of Political Economy*, 113:1, 83-120.
- [24] Grant, Monica J. and Jere R. Behrman, 2010, “Gender Gaps in Educational Attainment in Less Developed Countries,” *Population and Development Review*, 36:1. 71-89 (March).
- [25] Heckman, J., Lalonde, R. and J. Smith (1999): ”The economics and econometrics of active labour market programs,” *Handbook of Labor Economics*, Elsevier.
- [26] Jalan, J. and M. Ravallion (2003): ”Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching,” *Journal of Business and Economic Statistics* 21:1 (January), 19-30.

- [27] King, Elizabeth M. and Jere R. Behrman (2009): "Timing and Duration of Exposure in Evaluations of Social Programs," *World Bank Research Observer*, (February) 24:55-82.
- [28] Lechner, Michael (2002): "Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods," in *Journal of the Royal Statistical Society: Series A*, 165:1, 59-82.
- [29] Parker, Susan W (2004): "Evaluación del impacto de Oportunidades sobre la inscripción, reprobación y abandono," mimeo.
- [30] Parker, Susan and Jere R. Behrman, 2008, "Following Young Adults Who Benefited from Oportunidades for Almost a Decade: Impacts on Education and Achievement Tests," Report for Evaluación Externa del Programa de Desarrollo Humano Oportunidades 2007-2008, Instituto Nacional de Salud Pública (INSP), Cuernavaca, Mexico.
- [31] Parker, Susan W., Jere R. Behrman and Petra E. Todd (2004): "Medium-Term Effects on Education, Work, Marriage and Migration in Rural Areas," Philadelphia: University of Pennsylvania, Technical Document Number 1 on the Evaluation of Oportunidades 2004, mimeo.
- [32] Parker, Susan W. and Emmanuel Skoufias (2000): "The impact of PROGRESA on work, leisure and time allocation," October. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C. <http://www.ifpri.org/themes/progres.htm>;
- [33] Parker, SW, Rubalcava, L. and G. Teruel (2008): "Evaluating conditional schooling and health programs," in *Handbook of Development Economics*, Elsevier, Chapter 62, 3963-4035.
- [34] Rosenbaum, Paul and Rubin, Donald (1982): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70, 41-55.

- [35] Schultz, T. Paul (2000): "Impact of PROGRESA on school attendance rates in the sampled population," February. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, D.C.
- [36] Schultz, T. Paul (2004): "School Subsidies for the Poor: Evaluating the Mexican ProgresA Poverty Program," *Journal of Development Economics*, 74:2, 199-250.
- [37] Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, London: Chapman and Hall.
- [38] Skoufias, E. and B. McClafferty (2001): "Is PROGRESA working? Summary of the results of an Evaluation by IFPRI." Report submitted to PROGRESA. Washington, D.C.: International Food Policy Research Institute, <http://www.ifpri.org/themes/progresA.htm>;
- [39] Smith, Jeffrey and Todd, Petra (2005): "Does Matching Address LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, Vol. 125:1-2 (March-April), 305-353.
- [40] Todd, Petra E. (2004) "Technical Note on Using Matching Estimators to Evaluate the Oportunidades Program For Six Year Follow-up Evaluation of Oportunidades in Rural Areas," Philadelphia: University of Pennsylvania, mimeo.
- [41] Todd, Petra and Kenneth Wolpin (2006): "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility," *American Economic Review*, 96: 5, 1384-1417.

Appendix A. Local Linear Regression and Method for Determining Common Support

This appendix gives expressions for the local linear regression weighting function (as introduced in Fan (1992a,b)) and also describes the method we use to impose the common support requirement of matching estimators. Letting $\tilde{P}_{ik} = P_k - P_i$, the local linear weighting function is given by

$$W(i, j) = \frac{G_{ij} \sum_{k \in I_0} G_{ik} \tilde{P}_{ki}^2 - [G_{ij} \tilde{P}_{ji}][\sum_{k \in I_0} G_{ik} \tilde{P}_{ki}]}{\sum_{j \in I_0} G_{ij} \sum_{k \in I_0} G_{ik} \tilde{P}_{ki}^2 - \left(\sum_{k \in I_0} G_{ik} \tilde{P}_{ki} \right)^2}.$$

where $G(\cdot)$ is a kernel function and a_n is a bandwidth parameter.²³

To implement the matching estimator given above, the region of common support S_P needs to be determined. By definition, the region of support includes only those values of P that have positive density within both the $D = 1$ and $D = 0$ distributions. Following the procedure in Heckman, Ichimura and Todd (1997), we require that the densities exceed zero by a threshold amount determined by a “trimming level” q . That is, after excluding P points for which the estimated density is exactly zero, we exclude an additional q percentage of the remaining P points for which the estimated density is positive but very low. The set of eligible matches are therefore given by

$$\hat{S}_P = \{P \in I_1 \cap \hat{S}_P : \hat{f}(P|D = 1) > c_q \text{ and } \hat{f}(P|D = 0) > c_q\},$$

where c_q is the density cut-off trimming level.²⁴

²³We assume that $G(\cdot)$ integrates to one, has mean zero and that $a_n \rightarrow 0$ as $n \rightarrow \infty$ and $na_n \rightarrow \infty$. In estimation, we use the quartic kernel function, $G(s) = \frac{15}{16}(s^2 - 1)^2$ for $|s| \leq 1$, else $G(s) = 0$.

²⁴The q th quantile c_q is determine by solving for

$$\sup_{c_q} \frac{1}{2J} \sum_{\{i \in I_1 \cap \hat{S}_P\}} \{1(\hat{f}(P|D = 1) < c_q) + 1(\hat{f}(P|D = 0) < c_q)\} \leq q,$$

where J is the number of observed values of P that lie in $I_1 \cap \hat{S}_P$. That is, matches are constructed only for the program participants for which the propensity scores lie in \hat{S}_q . The densities are estimated by the standard kernel density estimator, using Silverman’s rule-of-thumb bandwidth.

Appendix B. Selection of Intervention and Non-intervention Areas, Construction of Variables and the Propensity Score Model

Selection of Intervention and Non-intervention Areas

The intervention areas were selected by first choosing a random sample of communities (localities) where the program was scheduled to be started in 2002. Then, a stratified random sample of 149 blocks was chosen with all households in the blocks included in the sample. The comparison group blocks were chosen by matching the treatment blocks with blocks in eligible but non participating communities using characteristics based on pre-program community-level data from the 2000 Census. The non participating localities were eligible communities who were deferred for incorporation into the Program until 2004. In fact during 2004, these areas did become beneficiary communities Program administrative data were used to define who received benefits and who did not.

Construction of Variables

We next describe the construction of the main variables and the restrictions placed on the sample used in the analysis. It was necessary to exclude observations from the analysis that had missing data on key variables. The initial sample considered had 56,437 observations which included individuals from the groups used in the analysis mentioned before. Because the study is focused on the effects of the program on children's education, only individuals that were between 6 and 20 years old in 2002 were considered (34,161 individuals dropped). From those, 14,427 were in groups (a) or (c). Additional deletions were made on basis of age and other demographic information. For example, individuals with no age information for any of the rounds were deleted as well as individuals with information for all rounds but with age inconsistencies between rounds. Regarding the information on education, those individuals with no information in the baseline were not considered; and, individuals with contradictory information in the follow-up rounds with respect to grades completed and enrollment were not considered. These deletions left a usable sample of 12,160 individuals in group (a) or (c) in the baseline. Due to attrition, there is no information on some of those children in the data gathered in 2003 and 2004. From the 12,160 children, in the 2003

data we have information on 10,541 individuals and in 2004 we have 9,652 observations. Some additional observations were dropped in the impact analysis if they were missing the information required to obtain propensity scores or when information on the outcome variable was missing.

When the variable age was missing in all years, the individual was dropped from the sample. In the cases where the information was only available in 2003 or 2004 but not in 2002, the missing value was imputed using information from other years. The individual was dropped from the sample whenever the reported age difference from 2002 to any other years was negative or greater than 2 years in 2003 or greater than 3 in 2004, because the disparity is indicative of at least one age being reported with error.

The variable grades of schooling takes into account completed grades, not counting Kindergarten. We imposed an upper bound for the number of grades achieved in each degree according to the grades required for each degree by the SEP (the Ministry of Education in Mexico). For example, an individual who reports having a primary degree is considered to have 6 (and no more) grades. Individuals that reported a value that exceeded the upper bound on the reported degree were imputed with the maximum number of grades, according to the degree reported.

The number of years for someone in secondary school degree was counted as 6 years of primary school plus the number of grades reported, up to a maximum of three grades. High school counted as 9 grades (primary plus secondary) plus the number of grades of high school reported, up to a maximum of three grades. The normal school degree is required to become a teacher in the national schooling system and the number of grades spent in normal school differs depending on the grades that the individual is going to teach. For this degree, we assume that finishing secondary level (12 grades of school) is the general prerequisite for normal school. The grades of education for someone with normal school sum to the grades of any previous degrees plus up to a maximum of three grades in normal school. The technical or commercial track has many varieties, making it difficult to assign the grades for individuals who report obtaining technical or commercial degrees. The grades of schooling

for someone with a technical degree sum to the grades of any previous degrees (which were reported as prerequisites) plus the number of years of technical training, up to a maximum of three years. The college (university level) degree starts after high school so the number of grades is counted as 12 grades of school plus the number of grades of college reported, up to a maximum of 5 grades. Those reporting masters or doctoral degrees are assumed to have 17 plus the reported number of grades within the stated degree. Given these rules, an individual could have a schooling attainment of at most 20 grades.

Once the schooling attainment was determined for each individual in each year, we applied some additional restrictions with regard to consistency of the information. Because we assume that grades accumulate from first grade of primary and individuals usually start this degree when they are 6 years old, we imposed the restriction that the total grades of school attainment cannot be higher than the age of the individual plus one grade of school minus 5 (we used 5 instead of 6 to allow for the possibility that some individuals start grade 1 at an early age as in the data). For those exceeding the upper bound restriction, the grades of school attainment were considered as missing. Also, when the difference between completed school grades from 2002 to 2003 was negative or greater than two the information was considered to be missing; this is, we did not allow for decreases in school grades or skipping more than one school grade in a natural year.

School enrollment status was determined by the question: Did you attend school last year? There is information for 2001-2002, 2002-2003 and 2003-2004 school years regarding this variable. We did not allow for discrepancies between school enrollment and number of school grades attained. Individuals that reported non-enrollment on the school year 2002 but reported an increase in school grades were excluded.

The variables pertaining to hours spent doing homework and whether the child receives help with homework were constructed conditional on attending school in the 2001-2002 school year. The variables are set to zero if the individual reports enrollment in the 2001-2002 school year but non-enrollment in the current school year.

An individual is considered employed in either year if reported to have worked the week

before to being surveyed or to have a job but not have worked in the reference week.

The income variable used in our analysis is measured on a monthly basis. It includes the monthly income from both the principal job and any secondary job. Income is set to zero if an individual is not employed. Also, if individuals report income in weekly or annual terms, we convert the reported income to a monthly measure.

The Propensity Score Model

The propensity scores were estimated using a logistic model. A household was considered to be a participant or treated if it was located in an intervention area, was eligible and was incorporated by the program, according to the administrative data. A household was considered non-participant if it was located in an intervention area, was eligible but was not incorporated. Table B.1 presents the list of variables used in the logistic model, along with their definitions. Table B.2 shows the estimated coefficients. The key predictors of program participation are described in the text.