



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

PIER Working Paper 01-034

“Evaluating Preschool Programs when Length of Exposure to the Program Varies:
A Nonparametric Approach”

by

Jere Behrman, Yingmei Cheng and Petra Todd

http://papers.ssrn.com/sol3/papers.cfm?abstract_id=286296

Evaluating Preschool Programs when Length of Exposure to the Program Varies: A Nonparametric Approach*

Jere R. Behrman
Yingmei Cheng
Petra E. Todd

University of Pennsylvania

first version, October 1999
revised, December 2000

*This research is sponsored by the World Bank Research Foundation Project on “Evaluation of the Impact of Investments in Early Childhood Development on Nutrition and Cognitive Development” (P.I. Harold Alderman). This paper was presented at the 2000 World Congress meetings of the Econometric Society. We thank Harold Alderman, John Newman and participants at seminars at the University of Minnesota, Lehigh University, University of Delaware, University of Virginia, University of Pennsylvania, Hebrew University and the NBER for helpful comments. We are also grateful to Elizabeth Peñaranda of the PAN staff in La Paz, Bolivia for help in understanding the details of the program being evaluated and of the data. Todd thanks NSF for support under SBR-9730688. Behrman and Todd are professors of economics at the University of Pennsylvania. Todd is also a member of NBER and a research associate of the Center for Social Program Evaluation, Harris School, University of Chicago. Cheng is a graduate student at the University of Pennsylvania. Please direct correspondence to any of the authors at University of Pennsylvania, Department of Economics, 3718 Locust Walk, Philadelphia, PA 19004-6297 or by email to jbehrman@econ.sas.upenn.edu or petra@athena.sas.upenn.edu.

Abstract

This paper uses a large, nonexperimental data set to evaluate the effects of a preschool enrichment program in a developing country on cognitive, psycho-social, and anthropometric outcomes. Outcomes are shown to be highly dependent on age and duration of exposure to the program. To minimize the impact of distributional assumptions, program impacts are estimated nonparametrically as a function of age and duration. A generalized version of the method of matching is developed and used to control for nonrandom selectivity into the program or into alternative program durations. The estimates obtained using this method reveal a different pattern of program impacts with respect to age and duration than does a parametric model under more restrictive functional form assumptions. The estimates based on matching show a greater dependence of test score impacts on duration of exposure to the program and show larger impacts for the anthropometric measures for a range of durations.

Impact estimates are based on three different comparison groups: children in the same communities in which the program was introduced who were not in the program, children in similar communities in which the program had not yet been introduced, and children who were in the program for a month or less. The average impact estimates for test score outcomes are robust to the three alternative comparison groups. The preschool program is found to increase cognitive and psycho-social test scores, but only for children who participated in the program for at least seven months. The anthropometric results for weight differ substantially depending on which comparison group is used, suggesting that estimates based on the first two comparisons are contaminated by important unobserved characteristics related to program entry. The preferred estimates based on the third comparison group indicate that the program tends to improve the anthropometric outcomes, again with initially increasing effects as the duration of program participation increases. Cost-benefit analysis based on these estimates and other assumptions indicate fairly high rates of return for this program.

Keywords: program evaluation; matching; selection bias; propensity scores; local regression; human capital; nutrition; child development; anthropometrics; preschool programs; cognitive tests.

JEL classification: C14, C51, I12, I21, O15

1. Introduction

There is a growing recognition that human capital investments made in early childhood are important determinants of school performance and lifetime productivity.¹ Much accumulated evidence supports a strong association between cognitive and psycho-social skills measured at young ages and educational attainment, earnings, and employment outcomes.² These findings suggest that policies that promote the development of early skills can be effective in alleviating poverty.

In developing countries, low levels of investment in human capital are seen as a major barrier to growth as well as a source of poverty. Among the factors contributing to low levels are that children enroll late in elementary school relative to their U.S. counterparts, repeat grades more frequently, and drop out of school at earlier ages – often before completing elementary school. Recent research demonstrates that nutrition is an important factor in explaining delayed school enrollments and lower educational attainment levels.³ To combat such problems, governments in several developing countries have introduced subsidized preschool programs with the twofold goals of improving child nutrition and providing environments conducive to learning (e.g. see Myers, 1995). In recent years, the World Bank and other international agencies have devoted increasing resources to support such efforts.

Very little is known about the effectiveness of early preschool interventions in developing country settings, but several studies have evaluated the effects of preschool programs in the U.S. that are targeted at children from impoverished families. These programs include the Perry Preschool, Head-Start and Follow-through programs, the Milwaukee and Abecedarian Projects, and HIPPY (Home Instruction Program for Preschool Youngsters) programs, which are described in Table 1.⁴ The Perry Preschool Program is perhaps the best known of the

¹This view is expressed, for example, in the United States' Congress's 1994 stated goal to send every child to school "ready to learn." *Goals 2000: Education America Act*.

²See e.g. Currie and Thomas (1999), Neal and Johnson (1996), Herrnstein and Murray (1994), Murnane, Willett and Levy (1995).

³See Glewwe and Jacoby (1995), Alderman et. al. (1997), Glewwe, Jacoby and King (1998) and Martorell (1999) for evidence for Ghana, Pakistan, the Phillipines, and Guatemala.

⁴See Barnett (1992) for a survey of the findings from evaluations of many different U.S. programs.

U.S. programs in the evaluation literature. An experimental evaluation of this program found that children who participated in the half-day, center-based program scored higher on cognitive test scores for a few years after the program. Over time the test score gains disappeared, but the program nevertheless appeared to have enduring effects on many other outcome measures, some measured as late as age 27. Among the effects documented are increased educational attainment and earnings, lower welfare participation rates, lower out-of-wedlock birth rates, and lower crime rates.⁵ Not all evaluations of preschool programs find that effects on test scores fade. Experimental evaluations of the Milwaukee project found effects on cognitive skills tested at age 14, and an evaluation of the Abecedarian project documented modest effects on test scores measured at age 21. (Ramey, Campbell and Blair, 1998). The positive impacts consistently found for interventions aimed at very young children are in sharp contrast to the relatively weak impacts often found in evaluations of U.S. job training programs targeted at adolescent youth or adults.⁶

While the promising results from U.S. preschool program evaluations might lead to high expectations about similar programs in other settings, the results from U.S. experiences may not be generalizable to developing countries. Both the preschool programs and the families and children they aim to help differ in some possibly important respects. For example, program expenditure per child in developing countries is usually lower, although as a fraction of the family's income it may be higher. Lower levels of expenditure do not necessarily imply low impacts, however, because diminishing marginal returns to investment could lead to higher impact per unit of investment in countries with low levels of investment. Another difference in developing countries is that preschool providers are often less well trained. Lastly, in terms of the target population, children frequently suffer from protein and energy malnutrition and micronutrient deficiencies, which is why preschool programs in developing countries tend to put greater emphasis on nutrition. All of these differences in the characteristics of the programs and the contexts in which they operate could affect the extent and type of benefit from the intervention.

⁵In considering these impacts, though, it is important to keep in mind that the Perry Preschool Program spent significantly more per pupil than is typically spent on preschool interventions (\$7252/year). It spent over a third more than the Head-Start Program, for example. Most of this expenditure for the Perry Preschool Program went to teacher pay; the teachers tended to be highly trained professionals (Sweinhart and Weikart, 1998).

⁶See e.g. Bloom et. al. (1993) for discussion of U.S. job training programs.

In this paper, we analyze the effectiveness of an early childhood development and nutrition program in Bolivia called PIDI (*Proyecto Integral de Desarrollo Infantil*). This program provides day-care, nutrition and educational services to children between the ages of six months and six years who live in poor, predominantly urban areas.⁷ The goals of the program are to improve health and early cognitive/social development by providing children with better nutrition, adequate supervision and a stimulating environment. It is hoped that the program will also ease the transition to elementary school, improve progression through elementary grades, and raise school performance, all of which are expected to increase post-school productivity.

Through PIDI, children attend full-time child-care centers located in the homes of women living in low-income areas targeted by the program. These women are given training in child-care and loans and grants (up to \$500) to upgrade facilities in their homes. Each PIDI center has up to 15 children and approximately one staff member per five children.⁸ The program provides food to supply 70% of the children’s nutritional needs as well as health and nutrition monitoring and educational activity programs. The cost of the program has been estimated to be about \$43 per beneficiary per month, which is substantial in a country where per capita GDP is \$800 in exchange-rate-converted-pesos or \$2540 in purchasing power parity terms, with about 40% of the expenditure goes to the nutritional component of the program.⁹ In terms of its target population, the PIDI program is highly similar to the Abecedarian and Milwaukee projects, accepting children over a broad range of ages: 6 weeks to 5 years. The other programs described in table 1 are targeted at children in the older part of this range.

In this paper, we use a large nonexperimental dataset to assess the impact of the PIDI program on multiple child outcome measures related to health, cognitive development, and psychosocial skill development. As measures of health, we consider standard anthropometric measures such as height-for-age, weight-for-age, and weight-for-height. In the nutrition literature, height-for-age serves as a measure of ‘stunting’ and weight-for-height as a measure of ‘wasting.’ To measure cognitive and psycho-social development, we use children’s scores on a battery of tests of bulk motor skills, fine motor skills, language and auditory skills, and psycho-social skills.

As shown in Table 1, several of the existing studies of U.S. preschool programs use experi-

⁷There has been considerable expansion in rural areas subsequent to the collection of the data currently available for analysis.

⁸Additional staff are provided when there is a larger proportion of infant children.

⁹The cost estimate comes from Ruiz (1996) and the per capita GDP figure from World Bank (1997, Table 1, p. 214).

mental data. While experimental studies have an advantage over nonexperimental ones in that, under ideal conditions, randomization guarantees comparability between the comparison and treatment groups, they often have some disadvantages. Many of the experimental studies are based on relatively small samples drawn from just a few program sites.¹⁰ For example, the evaluation of the widely acclaimed Perry Preschool Program is based on a sample of 123 children, and the more recent experimental evaluation of the Abecedarian Program is based on 111 children. Even if the relatively small sample provides enough power for detecting the impact of the program, it is difficult to generalize results based on the experiences at a single program site to a larger population of interest. Also, the sample size would need to be increased to provide sufficient power to detect smaller size program impacts that might be expected from programs that spend far less on children than did the Perry Preschool and Abecedarian Programs. The only U.S. program evaluated using a large, representative data set is the Head Start Program, which was studied by Currie and Thomas (1995) using a large nonexperimental sample of children of mothers in the National Longitudinal Survey of Youth (NLSY) dataset.

In our study of the PIDI program, the sample size is about ten times larger than the sizes typically observed in experimental evaluations, and the data set is representative of the entire population of program recipients. Our data are nonexperimental in the sense that treatment is not randomly assigned, and families self-select into the program. In any evaluation based on nonexperimental data, noncomparability between the comparison group sample and the treatment group sample poses a threat to the validity of the results. Although our comparison group data sets were chosen by a sampling scheme designed to increase comparability with the participant group, we still find some important differences between the treatment and comparison group samples. For example, families with children in the program tend to have lower parental education levels and incomes, a difference that would likely downward bias the estimated program impacts if not taken into account. This bias is partly offset by the fact that program participants tend to be older than nonparticipants, which increases their average test scores and anthropometric outcomes. Our analysis shows the importance of carefully taking into account age differences between program participants and nonparticipants.

To control for potential bias due to nonrandom program selectivity, we develop a generalized matching estimator that allows the impact of the program to depend in a flexible way on the the age (in months) of the child and on the length of time in the program. The methodological

¹⁰For a discussion of other problems associated with randomized social experiments, such as randomization bias, substitution bias and contamination bias, see Heckman (1992).

contribution to the previous literature on matching is that we allow for a continuous dose of treatment (in our case, time spent in the program), while the literature usually assumes a discrete, finite number of treatment categories. Some of the matching estimators that we use require the assumption that selection into the program is “on observables,” i.e. that it can be taken into account by conditioning on observed family and child characteristics. An alternative estimator that we develop allows selection into the program to be based on unobservables, but assumes that conditional upon having selected into the program, selection into alternative program durations is on observables. The latter estimator is used to assess the marginal effects of increasing duration of exposure to the program. A key advantage of the marginal effect estimator is that it uses only data on the treatment group and is thus feasible in cases where no comparison group data are available.

We find that impact estimates for cognitive achievement and psycho-social test scores are relatively robust to the use of alternative comparison groups and estimators. However, impact estimates for the anthropometric measures for weight are highly sensitive to the choice of comparison group and estimator. Estimates based on the identifying assumption that outcomes are independent of the program participation decision after conditioning on observables are not credible in the sense that they imply, for example, substantial negative impacts on weight immediately upon program entry. Estimates of the marginal effects of longer durations in the program are credible for weight and indicate moderate and, in some cases, fairly large positive impacts of the program on weight. The estimates based on our matching methods are larger, particularly at longer durations and for toddlers age 18-36 months, than those obtained using standard methods with stronger functional form assumptions. Cost-benefit analysis based on our estimates and other assumptions, in part based on wage studies for developing countries, indicate that the PIDI Program may have fairly high rates of return.

1.1. Organization of the paper

In section 2, we develop a model of enrollment in preschool that gives an economic interpretation for the statistical parameters that we estimate by the method of matching. Section 3 describes the identifying assumptions required to justify the various matching estimators that we use. First, we discuss how we generalize matching estimators to accommodate a continuous treatment dose and to allow for impact heterogeneity with respect to children’s ages. Section 3 also describes the marginal effect estimators that can be implemented using only data on program

participants. Section 4 provides additional information about the PIDI program, analyzes the determinants of program participation, and presents the program impact estimates obtained by matching and, for comparison, by more standard regression methods. Section 5 performs a cost-benefit analysis based on our preferred marginal effect estimates and other explicit assumptions regarding subsequent schooling and wage effects and section 6 concludes.

2. A Model of the Preschool Participation Decision and Parameters of Interest

We next develop a model of the mother’s decision to enroll her child in preschool at a particular age and calendar time, which provides a way of interpreting the treatment effects that will be estimated by the matching method and gives some insight into which variables should be used in the matching procedure. Our framework is dynamic and assumes that the mother maximizes a time separable utility function that depends on her own consumption and on the quality of her child. We assume that there is a quality production function that depends on the mother’s time and child consumption of household monetary resources and on whether the child is in preschool. Although the program we are considering in this paper is free of charge, we assume there is a fixed cost K to the mother of transporting the child to the program site.

To focus only on the most relevant aspects, we abstract from certain considerations. We assume that the father’s only role is to contribute to the asset income $A(t)$ of the family, which is consumed in full every time period. We also assume that there is only one child for whom the mother is making decisions and avoid considerations of fertility decisions. Let the quality of the child at time t be represented by $q(t)$. Let $C^m(t)$ represent the consumption of the mother and $C^c(t)$ the consumption of the child at home (there may be consumption in preschool). Let $s^c(t)$ represent the time the mother spends with the child. Time not spent with the child is assumed to be spent working at wage $w(t)$. $a(t)$ represents the age of the child at date t and $D_p^*(t)$ is an indicator that takes a value one if the mother would choose to enroll the child in the preschool program were the child eligible, $e(t)$ is an indicator that takes a value one if the family is eligible for the program. $D_p(t)$ is an indicator for whether the child is actually enrolled, which depends both on the mother’s decision and on eligibility status. Let “ \sim ” above any variable denote the entire history vector.

The mother’s problem can be expressed as a dynamic programming problem, where the choices at any point in time are whether to enroll the child in preschool (only a viable alternative

if the family is eligible), how much time to invest in the child and how much consumption to allocate to the child:

$$V(D_p^*(t), s^c(t), C^c(t) | \tilde{A}(t-1), \tilde{w}^m(t-1), \tilde{e}(t-1), q(t-1)) =$$

$$\max_{\{C^m(t), s^c(t), D_p^*(t)\}} U(C^m(t), q(t)) +$$

$$\beta E_{t+1}(V(D_p^*(t+1), s^c(t+1), C^c(t+1) | \tilde{A}(t), \tilde{w}^m(t), \tilde{e}(t), q(t)))$$

subject to the constraints:

$$q(t) = q(s^c(t), C^c(t), D_p(t), a(t), q(t-1)) \quad (2.1)$$

$$C^m(t) + C^c(t) + D_p(t)K \leq (1 - s^c(t))w^m(t) + A(t) \quad (2.2)$$

$$D_p(t) = e(t)D_p^*(t) \quad (2.3)$$

Equation (2.1) describes the production technology for child quality, where we are assuming that previous period quality is a sufficient statistic for prior inputs and where we allow for the possibility that the productivity of the inputs depends on the child's age. For example, preschool could be highly productive for a toddler but not for a six month old infant. Equation (2.2) represents the budget constraint and (2.3) describes the constraint that preschool is only an option for eligible families. We assume that mothers do not try to influence their children's eligibility (for example, by making their child appear undernourished).

The total effect on child quality from participating in preschool from time period t to t' (the "treatment" effect of switching from state $D_P = 0$ to $D_P(t) = 1$ for $t..t'$) can be expressed in terms of the parameters of the model. It does not, however, correspond to any single parameter of the model. The total effect of treatment includes both the direct effect that participation has on quality in the current period as well as indirect effects that could occur, for example, if the mother reduces the child's consumption at home knowing that he/she receives meals at school. The total effect of preschool participation on current period quality for a particular child who starts off at quality level $q_{t-1}=\bar{q}$ is given by

$$\frac{dq(t)}{dD_p(t)} = \left\{ \frac{\partial q(v)}{\partial s^c(v)} \frac{\partial s^c(v)}{\partial D_p(v)} + \frac{\partial q(v)}{\partial s^c(v)} \frac{\partial C^c(v)}{\partial D_p(v)} + \frac{\partial q(v)}{\partial D_p(v)} \right\} \Big|_{q(t-1)=\bar{q}}$$

In addition, any change in current period quality affects future quality levels. If a child starts off period $t+1$ at a high quality level, he/she may be better able to take advantage of consumption and time investments. The effect of increasing quality due to program enrollment at time t on future quality levels (at time $t' > t$) is given by

$$\frac{dq(t')}{dq(t'-1)} \frac{dq(t'-1)}{dq(t'-2)} \cdots \frac{dq(t+1)}{dq(t)} \frac{dq(t)}{dD_p(t)},$$

where $\frac{dq(t+1)}{dq(t)} = \frac{\partial q(t+1)}{\partial c^c(t+1)} \frac{\partial c^c(t+1)}{\partial q_t} + \frac{\partial q(t+1)}{\partial s^c(t+1)} \frac{\partial s^c(t+1)}{\partial q(t)} + \frac{\partial q(t+1)}{\partial q(t)}$.

Thus the cumulative effect of being enrolled in preschool in periods t through t' is given by

$$\Delta_{t,T} = \sum_{v=t}^{t'} \left\{ \frac{dq(v)}{dD_p(v)} + \sum_{w=v+1}^T \frac{dq(w)}{dq(w-1)} \cdots \frac{dq(v+1)}{dq(v)} \frac{dq(v)}{dD_p(v)} \right\}$$

The first term captures the current period impact and the second term the impact on future quality levels up until some end period T .¹¹ In our study we do not observe children over the entire time period (up to T), so we can only estimate the cumulative effect of the preschool treatment up until the time of observation t^o . We must also assume that the empirical test score and anthropometric measures available in the dataset capture aspects of child quality.¹² In terms of the notation from section 3, the outcome in the no-program state and in the treated state correspond to

$$Y(a, 0, t) = q(t)|_{D_P(t')=0 \text{ for all } t' \leq t} \text{ and } Y(a, l, t) = q(t)|_{D_P(t')=0 \text{ for all } t' \leq t} + \Delta_{t,t+l}$$

where we now add a time dimension to Y to denote the calendar time of observation. The mean treatment effect parameter we estimate (conditional on age and duration of time in the program) corresponds to $E(\Delta_{t,t^o} | \text{age} = a, t^o - t = l, D_p = 1)$. It depends on the production function for quality, on the utility function determining other input levels and on the distribution of asset and wage income among persons. Knowledge of the parameters of the production function and/or the utility function is not enough to infer the effect of the program. Also, knowing the effect of the program tells us very little about the quality production technology, since we estimate only the net effect of the program. Only under the very strong assumption of no other effects ($\frac{\partial s^c(t+1)}{\partial q(t)} = 0, \frac{\partial c^c(t+1)}{\partial q_t} = 0, \frac{\partial s^c(v)}{\partial D_p(v)} = 0, \frac{\partial C^c(v)}{\partial D_p(v)} = 0$) could we conclude that the treatment effect corresponds to a feature of the production technology. (See Todd and Wolpin, 2000)

Now that we have defined one parameter of interest in terms of the model of the preschool participation decision, we consider the question of how to choose the set of matching variables. Let $s_0^c(t)$ and $C_0^c(t)$ be the values that solve the dynamic programming problem under the constraint of no program participation up until date t ($D_p(t) = 0$) and let $q_o(t) = q(t)|_{D_P(t')=0 \text{ for all } t' \leq t}$ be

¹¹In writing the treatment effect solely as a function of effects on current and future quality, we are also assuming that there are no effects of anticipation of the program on quality levels prior to the program entry date.

¹²Preschool investments could increase the amount learned in school and lead to higher quality in elementary school years, but these benefits will not be captured by our estimation approach due to the data limitation posed by not observing children at these later ages. In the cost-benefit analysis of section 6, we will briefly consider these other sources of benefits.

the quality level. Also, let $s_1^c(t)$, $C_1^c(t)$ and $q_1(t)$ be the values that solve the problem if the mother chooses to enroll the child in the program at date t (and possibly thereafter) given that the child is eligible (i.e. for the case $D_p^*(t) = 1$ and $e(t) = 1$). The decision to enroll the child at any time period implies that at that date, the expected future utility from participating is higher than the utility from not participating:

$$U(C_0^m(t), q_0(t)) + \beta E_{t+1}(V(D_p^*(t+1), s^c(t+1), C^c(t+1)|\tilde{A}(t-1), \tilde{w}^m(t-1), \tilde{e}(t-1), q(t-1))) < U(C_1^m(t), q_1(t)) + \beta E_{t+1}(V(D_p^*(t+1), s^c(t+1), C^c(t+1)|\tilde{A}(t-1), \tilde{w}^m(t-1), \tilde{e}(t-1), q(t-1))).$$

It matters at what date the preschool program becomes available to the mother. Some mothers faced with the option of a new program would have liked to enroll their children at an earlier age had the program been available. Therefore, as a new preschool program matures, the age distribution of the children enrolled would naturally tend to get younger as mothers are able to choose over a larger set of possible enrollment ages. This implies that treatment impacts estimated for newly available preschool programs are likely to understate the benefits of the program because the mothers may not be enrolling their children at optimal ages.

Suppose the econometrician has access to all the variables that describe the state variables for the enrollment decision at time t ($\tilde{A}(t-1), \tilde{w}^m(t-1), \tilde{e}(t-1), q(t-1)$). Let $\tilde{Z}(t)$ denote this set of variables. The cumulative matching estimator introduced in the next section that compares children in the program to observably similar children out of the program assumes that

$$E(Y(a, 0, t)|\tilde{Z}(t), D_p^*(t)) = E(Y(a, 0, t)|\tilde{Z}(t)).$$

If $\tilde{Z}(t)$ contains all the information relevant to the $D_p^*(t)$ decision, then this assumption is justified since $\tilde{Z}(t)$ is a sufficient statistic for $D_p^*(t)$. However, in this case the econometrician could also perfectly predict whether mothers put their child in preschool or not, so matching would also be impossible (the support condition discussed in the next section would be violated). It is more realistic to assume that the econometrician has access only to a partial set of conditioning variables, denoted by $\tilde{Z}^*(t)$, so that observably similar persons make different choices. The application of matching is then only justified under the restriction that

$$E(Y(a, 0, t)|\tilde{Z}^*(t), D_p^*(t)) = E(Y(a, 0, t)|\tilde{Z}^*(t)).$$

From consideration of the model, we conclude that the types of variables that should be included in $\tilde{Z}^*(t)$ are any variables relevant to the decision to enroll in the program, including variables that capture eligibility, asset and wage history, as well as variables that measure the age of

the child and the quality of the child at the time of the enrollment decision. Although not an explicit feature of the model described above, other variables such as family structure and number of dependents would probably also factor into the participation decision process by affecting resources that the mother allocates to the child on which the model focuses. We include all these kinds of variables in implementing the matching estimators that are described in section 3.

Lastly, it is important *not* to use in matching variables that themselves are affected by the program. This is because the matching estimator integrates over $f(\tilde{Z}^*(t)|D_p^*(t) = 1)$ (the average over the treatment group is self-weighting by this density). To estimate correctly the mean no-treatment outcomes, we require that the density of the matching variables not change depending on whether the treatment group receives the treatment. For this reason, we use for matching the following kinds of variables: (a) variables observed prior to the enrollment decision (under the assumptions that the density of these variables does not change due to anticipation about the program), (b) variables that we expect to be stable over the time period of observation (such as mother’s and father’s education, family structure, characteristics of the household), or (c) variables that are deterministic with respect to time (such as child’s age). We do not include variables that directly relate to children’s physical , mental and social development.

3. Generalized Cumulative and Marginal Matching Estimators

We now describe the estimation methods that we will use to control for potential evaluation bias arising from differences between the treated group and the comparison group. We go beyond the previous literature in two respects: by extending matching methods to permit conditioning on children’s age and on the duration of time in the program and by developing a marginal matching procedure that can be implemented when only data on program participants are available. The latter method allows selectivity into the program to be based on unobservables.

First we need to introduce some notation to define the parameters of interest in our evaluation. Let $Y(a, l)$ denote the outcome (test score or anthropometric measure) for a child of age a who participated in the program for length of time l .¹³ For nonparticipants, $l = 0$. Also define $D_P = 1$ if $l > 0$, and otherwise $D_P = 0$.

¹³We assume participation takes place in consecutive time periods. Our data gives us the date of entry at the time of the observation. Even though there may have been some periodic absences from the program, we cannot determine in our data if and when they occurred.

One treatment effect of interest is the cumulative effect of participating in the program l time periods relative to not participating at all. For a child of age a , it is given by $\Delta(a, l, 0) = Y(a, l) - Y(a, 0)$. Another parameter of interest is the marginal effect of participating in the program l_1 time periods relative to l_0 time periods: $\Delta(a, l_1, l_0) = Y(a, l_1) - Y(a, l_0)$. Neither of these program effects are directly observable, because every child in the program is observed for a single duration at each age and no child is observed simultaneously in and out of the program at the same age. Because of this missing data problem, we do not attempt to estimate the full distribution of treatment impacts. We focus instead, firstly, on the problem of estimating average treatment impacts, and, secondly, on the problem of estimating marginal treatment impacts, in both cases, conditional on age and duration of exposure to the program. The average program impact for children of age $a \in A$ who participated l_1 time periods as opposed to l_0 (where l_0 could equal 0) is given by

$$\bar{\Delta}(A, l_1, l_0) = \frac{\int_{a \in A} [Y(a, l_1) - Y(a, l_0)] f_a(a|l = l_1) da}{\int_{a \in A} f_a(a|l = l_1) da}.$$

where $f_a(a|l = l_0)$ is the conditional density of ages and A could be a singleton set or correspond to a range of ages.

Integrating over the density of observed program durations gives the overall impact of the program:

$$\bar{\Delta}(A, L, l_0) = \frac{\int_{l \in L} \int_{a \in A} [Y(a, l) - Y(a, l_0)] f_{a,l}(a, l) da}{\int_{l \in L} \int_{a \in A} f_{a,l}(a, l) da}$$

where $L = \{l : l > 0\}$. $\bar{\Delta}(A, L, 0)$ gives the average impact of participating in the program relative to not participating for the $D_P = 1$ group, commonly known as the *average impact of treatment on the treated*. This is a key parameter of interest in many evaluations. A comparison of a monetary valuation of this parameter with average program costs is informative on whether the program has a positive net benefit. (See section 5 below)

3.1. Estimators of program impacts

Information on $Y(a, l)$ for $l > 0$ is observed from data on participant children, but $Y(a, l')$ is unobserved. Below we describe the identifying conditions that justify the method of matching as a way of imputing the missing data. We also discuss how the matching estimators we use relate to the more traditional estimators of program impacts based on linear regression.¹⁴

¹⁴Matching methods are developed and applied to the evaluation of training programs in Heckman, Ichimura, and Todd (1997), Heckman, Ichimura, Smith and Todd (1998) and Dehejia and Wahba (1998).

3.1.1. Estimating Program Impacts when the Counterfactual is No Treatment

The matching method estimates no-program outcomes for program participants by taking weighted averages over outcomes for observably similar persons who did not participate in the program. The degree of similarity between two persons is determined by the distance, according to some metric, between a set of their characteristics that constitute the “matching variables.” By matching on the characteristics of the treatment group, the method effectively aligns the distribution of observables of the comparison group to that of the treated group. In performing this alignment, it emulates a feature of a randomized experiment in which the characteristics of the treatment and comparison group are aligned by virtue of randomization.

The identifying assumption that justifies the matching estimator we use to estimate $\bar{\Delta}(A, l, 0)$ is that there exist a set of conditioning variables X such that

$$E(Y(a, 0)|l, X) = E(Y(a, 0)|l = 0, X) \quad (3.1)$$

and

$$0 < f(l, a|D_P = 0) < 1 \quad (3.2)$$

The first set of conditions imply that after conditioning on a set of observed characteristics $\{a, X\}$, no-treatment outcomes for children who have participated for duration l for children who have not participated ($D_P = 0$) are the same on average. This allows us to use outcomes for nonparticipants to impute no-treatment outcomes for participants. The second condition ensures that for each child in the participant group there is positive probability of finding a match from the nonparticipant group.¹⁵ Let $S_P = \{(a, x) : f(a, x|D_P = 1) > 0 \text{ and } f(a, x|D_P = 0) > 0\}$ denote the region of the support of (a, x) that satisfies (2), called the region of overlapping support.¹⁶

Under the above conditions, $\bar{\Delta}(A, L, 0)$ can be estimated by

$$\widehat{\Delta}(A, L, 0) = \frac{1}{n} \sum_{\substack{i \in \{D_P=1\} \cap \{a_i \in A\} \cap \\ \{l_i \in L\} \cap \{(a_i, l_i) \in S_P\}}} \{\widehat{E}(Y(a_i, l_i)|X_i, D_{P_i} = 1) - \widehat{E}(Y(a_i, 0)|X_i, D_{P_i} = 0)\}, \quad (3.3)$$

¹⁵See Rosenbaum and Rubin (1982). Under both conditions, treatment is termed *strictly ignorable*. If there are some (a, X) values for which the second support condition fails, then treatment impacts cannot be estimated by the method of matching for individuals with those characteristics.

¹⁶See Heckman, Ichimura and Todd (1997) for related discussion.

where n is the cardinality of the set $\{D_P = 1\} \cap \{a_i \in A\} \cap \{l_i \in L\} \cap \{(a_i, l_i) \in S_P\}$ and $\hat{E}(Y(a_i, l_i)|X_i, D_{Pi} = 1)$ and $\hat{E}(Y(a, 0)|X_i, D_{Pi} = 0)$ are consistent estimators of the conditional expectations of the outcomes with and without treatment.

We use local nonparametric regression methods to estimate the conditional expectations. The estimator $\hat{E}(Y(a_i, 0)|X_i, D_P = 0)$ can be written compactly as

$$\sum_{k \in \{D_P=0\}}^{n_0} Y_k(a_k, 0)W_k(\|a_k - a_i\|, \|X_k - X_i\|), \quad (3.4)$$

where $W_k(\|a_k - a_i\|, \|X_k - X_i\|)$ are weights that sum to one and that depend on the Euclidean distance between (a_i, X_i) and (a_k, X_k) . For standard kernel weighting functions, observations that are close according to the distance metric receive greater weight. The nonparametric estimators we use are local linear regression estimators that have been developed and studied in Cleveland (1979), Fan (1992), and Fan and Gijbels (1996). The details of local linear regression estimators are described in Appendix B where we also discuss how the weights are calculated.¹⁷

The analogous nonparametric estimator for $\hat{E}(Y(a_i, l_i)|X_i, D_{Pi} = 1)$ in (3.3) is

$$\hat{E}(Y(a, l)) = \sum_{k \in \{D_P=1\}} Y_k(a_k, l_k)W_k(\|l_k - l_i\|, \|a_k - a_i\|, \|X_k - X_i\|), \quad (3.5)$$

where the weights now additionally depend on the distance between l_k and l_i (allowing the impact of the program to depend on the duration of time in the program).¹⁸ Note that in (3.3) averaging is performed in two stages, once in obtaining the nonparametric estimates and again in averaging over the set $\{D_P = 1\} \cap \{a_i \in A\} \cap \{l_i \in L\} \cap \{(a_i, l_i) \in S_P\}$. Because of the second averaging, the average impact estimators over ranges of age and duration values converge at a faster rate than the pointwise in a and l estimators. The asymptotic theory developed in Heckman, Ichimura and Todd (1998) is general enough to accommodate the estimators we use. In the empirical work, we evaluate the variation of the estimators using bootstrap methods.

¹⁷The number of observations used in constructing the averages are determined by the choice of bandwidth or smoothing parameter. We use least squares cross-validation to choose these parameters as discussed in section 4.

¹⁸An alternative approach would be to construct the weighted averages in (4) over the set of observations $k \in \{D_P = 1\} \cap \{l = l_i\}$. Instead, we do local averaging over durations l because there may not be many observations at any individual duration value.

3.1.2. Estimating Marginal Program Impacts

Instead of or in addition to the impact of the program against the benchmark of no program, we may be interested in the marginal treatment effect of increasing duration in the program from l_1 to l_0 : $\bar{\Delta}(a, l_1, l_0) = E(Y(a, l_1) | D_P = 1, X) - E(Y(a, l_0) | D_P = 1, X)$, where $l_1, l_0 > 0$. As we next describe, a potential advantage of a marginal effect estimator is that one can allow for the possibility that program participation selection is on unobservables related to program outcomes.

There are two different ways of estimating marginal effects. One is to estimate them from the cumulative program effects. The other is to only use data on program participants, drawing comparisons between program participants who have taken part in the program for different lengths of time.

(i) Marginal impact estimator based on difference in cumulative effects An estimator of the marginal effect from participating in the program for l_1 time periods as opposed to l_0 time periods ($l_1 > l_0$) can be obtained as the difference between the two cumulative program effects.

$$\hat{\Delta}(a, l_1, l_0) = \hat{\Delta}(a, l_1, 0) - \hat{\Delta}(a, l_0, 0).$$

Estimating marginal effects in this way assumes that the group of children observed participating in the program l_0 periods provide an appropriate comparison group for the children observed participating l_1 time periods - an assumption that may not be justified if children are systematically entering or dropping out from the program at different ages. Partly for this reason we prefer the approach described next, which is the one we take in our empirical work.

(ii) Marginal impact estimator that only uses data on program participants An alternative estimation strategy only uses data on program participants and compares outcomes for children of similar ages with different durations. An advantage of this approach over those discussed above is that it does not require assumptions on the process governing selection into the program and allows for the possibility that selection into the program is based on unobserved characteristics. However, here we are faced with a different type of possibly nonrandom selection—the process governing selection into alternative program durations. For example, four year-olds who have taken part in the program for three years may be systematically different from four year-olds who just recently entered the program. Again, matching methods can be

used to solve the selection problem relating to the choice of program duration—this time under the assumption that children who have taken part in the program for different lengths of time can be made comparable by conditioning on observed child and parental characteristics.

The identifying assumption we invoke is that there exists a set of conditioning variables X such that

$$E(Y(a, l_0)|l = l_1, X) = E(Y(a, l_0)|l = l_0, X)$$

and

$$0 < f(a, X|l = l_0) < 1.$$

Under these assumptions, an estimator for $\bar{\Delta}(a, l_1, l_0)$ is given by

$$\widehat{\bar{\Delta}}(a, l_1, l_0) = \frac{1}{n} \sum_{i \in \{l_i = l_1\}} \hat{E}(Y(a, l_1)|X_i) - \hat{E}(Y(a, l_0)|X_i).$$

The conditional expectations are estimated by the same local regression method as described in relation to (3.5).

$\widehat{\bar{\Delta}}(a, l_1, l_0)$ gives the effect of increasing duration in the program from l_1 time periods to l_0 time periods *for the set of children who participated length of time l_1* . Obtaining the effect of increasing duration from l_0 to l_1 *for the set of children who participated l_0 time periods* would require summing over $i \in \{l_i = l_0\}$ values.

3.1.3. Reducing the dimension of the conditioning problem

The above estimation strategy is difficult to implement for a large set of conditioning variables X . To reduce the dimensionality of the conditioning problem, we can use the insights of Rosenbaum and Rubin (1983) who observed that for random variables Y and X and a discrete random variable D denoting assignment to a binary treatment,

$$E(D|Y, P(D = 1|X)) = E(E(D|Y, X)|Y, \Pr(D = 1|X)),$$

so that $E(D|Y, X) = E(D|X)$ implies $E(D|Y, \Pr(D = 1|X)) = E(D|\Pr(D = 1|X))$. This shows that when no-treatment outcomes are independent of program participation conditional on X they are also independent of participation conditional on $P(X) = \Pr(D = 1|X)$. Matching on the probability of participation instead of on X directly provides a way of reducing the

dimensionality of the conditioning problem when $P(X)$ can be estimated parametrically or semiparametrically at a rate faster than the nonparametric rate.

In our context, by imposing stronger conditional independence assumptions, we could apply the same reasoning to $Y(a, 0)$. However, for the purpose of estimating average effects of treatment, the assumption of conditional independence of outcomes and participation status is stronger than necessary.¹⁹ Instead, we will assume directly that (1) holds when we replace X by $P(X) = \Pr(D_P = 1|a, X)$. The conditional expectations can then be estimated by three- and two-dimensional nonparametric regressions:

$$\begin{aligned}\hat{E}(Y(a, l)|P(X), D_{P_i} = 1) &= \sum_{k \in \{D_P=1\}} Y_k(a_k, l_k) W_k(\|l_k - l\|, \|a_k - a\|, \|P(X_k) - P(X)\|) \\ \hat{E}(Y(a, 0)|P(X), D_{P_i} = 0) &= \sum_{k \in \{D_P=0\}} Y_k(a_k, 0) W_k(\|a_k - a\|, \|P(X_k) - P(X)\|).\end{aligned}\quad (3.6)$$

In our empirical work, we estimate the conditional probabilities $P(X)$ by logistic regression.²⁰

Modifications Required to Accomodate Choice-based Sampled Data In evaluation settings, data are often choice-based sampled, meaning that program participants are oversampled relative to their frequency in a random population. In this kind of sampling situation, weights are required to consistently estimate the probabilities of program participation.²¹ However, the required weights are often unknown. Heckman and Todd (1995) show that with a slight modification, matching methods can still be applied to choice-based sampled data when weights are unknown. They show that the odds ratio $P(X)/(1 - P(X))$ estimated using the wrong weights (*i.e.* ignoring the fact that the data is choice-based sampled) is a scalar multiple of the true odds ratio, which is a monotonic transformation of the propensity scores. Therefore, matching can proceed on the (misweighted) estimate of the odds ratio (or on the log odds ratio). In our empirical work, the data are choice-based sampled and the sampling weights are unknown, so we match on the odds ratio.

¹⁹See the discussion in Heckman, Ichimura and Todd (1997,1998)

²⁰Heckman, Ichimura, and Todd (1998) and Hahn (1998) consider whether it is better in terms of efficiency to match on $P(X)$ or on X directly. For the treatment on the treated parameter, Heckman et. al. (1998) show that neither is necessarily more efficient than the other. If the treatment effect is constant, then it is more efficient to condition on the propensity score; but in the general case the answer depends on the mean of the conditional variance relative to the variance of the conditional mean.

²¹See e.g. Manski and Lerman (1977) for discussion of weighting for logistic regressions.

3.2. Comparison between matching methods and traditional regression-based methods of evaluating program impacts

Because matching estimators are seldomly used in economic applications, we next briefly compare them to more conventional evaluation estimators. Assume outcome measures Y_{1it} and Y_{0it} , where i denotes the individual and t the time period of observation, can be represented by

$$\begin{aligned} Y_{1it} &= \varphi_1(X_{it}) + U_{1it} \\ Y_{0it} &= \varphi_0(X_{it}) + U_{0it}, \end{aligned} \quad (3.7)$$

where U_{1it} and U_{0it} are distributed independently across persons and satisfy $E(U_{1it}) = 0$ and $E(U_{0it}) = 0$.²² The observed outcome during any post-program time period is $Y_{it} = D_{P_i}Y_{1it} + (1 - D_{P_i})Y_{0it}$, which can be written as

$$Y_{it} = \varphi_0(X_{it}) + D_{P_i}\alpha^* + [U_{0it} + D_{P_i}\{U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_{P_i} = 1, X_{it})\}] \quad (3.8)$$

where $\alpha^* = \varphi_1(X_{it}) - \varphi_0(X_{it}) + E(U_{1it} - U_{0it}|D_P = 1, X_{it})$ corresponds to the average-impact-of-treatment-on-the-treated parameter described earlier. This is a random coefficient model because the conditional-on- X_{it} treatment impact varies across persons. Assuming $U_{1it} = U_{0it}$ gives a fixed coefficient model.

If we wish to estimate impacts conditional on duration or age, the conventional approach writes the impact parameter as a parametric function of age of the child and duration $\alpha^*(X, a, l)$.

A *cross-section* estimator uses data on $D_P = 0$ persons in a single time period to impute the outcomes for $D_P = 1$ persons in the same time period. For notational ease of exposition, assume $\varphi_0(X_{it}) = \varphi_1(X_{it})$. Define $\hat{\alpha}_{CS}$ as the least squares solution to α^* in

$$Y_{it} = \varphi_0(X_{it}) + D_{P_i}\alpha^* + U_{0it} + D_{P_i}\{U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_{P_i} = 1, X_{it})\}$$

For this model, $\hat{\alpha}_{CS}$ is unbiased for α^* if $E(U_{0it}|D_{P_i}, X_{it}) = 0$. The component of the error in brackets is not a source of bias for α^* , because it has conditional mean zero by construction.

A *difference-in-difference* (DID) estimator measures the impact of the program by the difference in the pre-program–post-program gain between participants and nonparticipants. It requires data for at least two time periods, t and t' , for $D_P = 1$ and $D_P = 0$ persons. The DID estimator $\hat{\alpha}_D$ corresponds to the least squares solution for α^* in

$$Y_{it} - Y_{it'} = \varphi_0(X_{it}) - \varphi_0(X_{it'}) + D_{P_i}\alpha^* + \{U_{0it} - U_{0it'}\} + D_{P_i}\{U_{1it} - U_{0it} - E(U_{1it} - U_{0it}|D_{P_i} = 1, X_{it})\}$$

²²This does not impose any restrictions because the $\varphi_1(X_{it})$ and $\varphi_0(X_{it})$ functions can be conditional means.

The application of the DID estimator is justified if $E(U_{0it} - U_{0it'} | D_{P_i}, X_{it}, X_{it'}) = 0$.²³

An *instrumental variable* approach is feasible when there is an exclusion restriction—a variable Z_i correlated with program participation, D_P , that can be excluded from the outcome equation. To apply the instrumental variables estimator, it is usually assumed that $E(U_{0it} | Z_i) = 0$. Also, for the parameter α^* it is necessary to assume $E(U_{1it} - U_{0it} - E(U_{1it} - U_{0it} | D_{P_i} = 1, X_{it}) | D_{P_i} = 1, X_{it}, Z_i) = 0$.²⁴ A drawback of the instrumental variables approach for our purposes is that it does not accommodate choice-based sampled data.

3.2.1. Comparison with Matching Methods

In many evaluations, some matching is carried out implicitly prior to applying regression methods in selecting the comparison group to have certain features in common with the treatment group. Individuals may be required to meet age, race, geographic location, income or other criteria for inclusion in the sample. The method of matching aims to increase the comparability between the treatment and comparison group samples by aligning the distribution of observed covariates of comparison group members to that of the treatment group. To see how realignment occurs, write the mean no-treatment outcome for program participants $E(Y(a, 0) | D_P = 1) = E_{X|D_P=1} \{E_Y(Y(a, 0) | D_P = 1, X)\}$ as

$$\begin{aligned} &= \int_{\mathcal{X}} E_Y(Y(a, 0) | D_P = 1, X) f(X | D_P = 1) dX \\ &= \int_{\mathcal{X}} E_Y(Y(a, 0) | D_P = 0, X) f(X | D_P = 1) dX \\ &= \int_{\mathcal{X}} E_Y(Y(a, 0) | D_P = 0, X) f(X | D_P = 0) \left\{ \frac{f(X | D_P = 1)}{f(X | D_P = 0)} \right\} dX, \end{aligned}$$

where the second equality follows under the assumptions that would justify the application of matching. The last line shows that matching can be seen as a reweighting method, where comparison group observations are reweighted by $\frac{f(X|D=1)}{f(X|D=0)}$. The reweighting balances observed characteristics of the treatment and comparison groups. Such a balancing would also occur in an experiment where treatment is randomized. Randomization would balance both observables and unobservables while matching methods can only balance the observables. This means that they perform best when aligning observations in terms of observables also aligns them closely in

²³Usually it is justified by assuming that U_{0it} follows a fixed effect error structure, $U_{0it} = f_i + \varepsilon_{it}$, where ε_{it} is a purely random innovation.

²⁴As discussed in Heckman (1992), this assumption requires that individuals not select into the program based on anticipated gains that can be partly predicted on the basis of the value of the instrument.

terms of unobservables. It is conceivable, however, that aligning people in terms of observables could make them more dissimilar in terms of unobservables, in which case applying the matching method to construct a comparison group may exacerbate bias.

Traditional regression-based estimators, such as the cross-sectional, difference-in-difference or iv estimators, do not attempt to emulate the balancing feature of randomized experiments. Instead, they control for differences in observables between groups in a conditional mean sense by specifying a functional form for how the observables affect the conditional mean of Y . A typical assumption would be that the conditional mean is linear in X , *age* and *duration*. Another difference between matching and regression estimators is how they deal with the problem of non-overlapping support. The matching estimator is only defined over the support of X where $f(X|D_P = 1) > 0$ and $f(X|D_P = 0) > 0$, and it assigns zero weight in estimation to comparison group observations for which $f(X|D_P = 1) = 0$ but $f(X|D_P = 0) > 0$. In contrast, a regression estimator typically uses all the observations in estimation and uses functional form assumptions to extrapolate over any regions of X where the supports do not overlap.

3.2.2. Selection on Unobservables

The estimators for cumulative program effects described above assume that after conditioning on a set of observable characteristics, mean outcomes are conditionally mean independent of program participation. If the program participation equation can be described by the model

$$D_P = 1(\psi(Z) - V > 0),$$

the matching estimator assumes that $E(Y(a, 0)|X, V < \psi(Z)) = E(Y(a, 0)|X)$. This assumption is not likely to be satisfied when unobservables are an important determinant of program selection.

One option in this case is to use a difference-in-difference (DID) matching strategy that allows for time-invariant unobservable differences in the outcomes between participants and nonparticipants. (See Heckman et. al. 1997) However, our data do not allow application of this estimator, because program participants are only observed after they already entered the program. As we show below, lack of preprogram data is a definite limitation in the data for our study and makes it difficult to estimate reliably the cumulative effects of the program. However, we can estimate the marginal impact of short verses long durations using the estimators described in section 3.1 that allow selection into the program to be based on unobservables.

4. Empirical Results

4.1. Description of the Data sets

The PIDI evaluation data sets consist of longitudinal and cross-sectional data collected in two rounds on three different subsamples:

- (1) a participating subsample (P) of children selected randomly from children in the PIDI program. Families of children in the P sample were required to satisfy the previously described eligibility criteria.
- (2) a comparison group subsample (A) selected from a stratified random sample of households with children in the age range served by PIDI living in poor urban communities comparable to those in which PIDI had been established, but in which PIDI programs had not yet been established as of the time of the survey.²⁵
- (3) a nonparticipant subsample (B) selected from a stratified random sample of households with at least one child in the age range served by PIDI living in poor urban communities in which PIDI had been established but without any children attending PIDI. This sample was chosen by its geographic proximity to PIDI sites.

We use both samples A and B as comparison groups for our estimates of average program impact. Sample B has an advantage over A in being drawn from the same area as the participant sample P, which controls for unobserved residence or local community effects that may affect children's outcomes. However, sample B families elected not to participate in the program, so the outcomes observed for B children may not be directly comparable to those for A children. Sample A combines data on families that would have participated in the program had the program been available as well as data on families that would not have participated. Finally, to estimate marginal program impacts (as described below in section 3.1.2), we compare children in the participating sample P who had been in PIDI for two or more months with children in the participating sample P children who had been in PIDI for one month or less.

All the children in sample P meet the eligibility criteria summarized in section 2.1.²⁶ Children in samples A and B do not necessarily meet the eligibility criteria. In our analysis, we impose

²⁵Stratification is based on information given in the 1992 Bolivian Census.

²⁶Once they were determined eligible, they could not become ineligible for the program even if some of their characteristics changed over time.

eligibility on the comparison group sample because children who are not eligible cannot choose to participate in the program.²⁷ However, we use the current eligibility criteria rather than the original ones because the original ones included subjective aspects, the application of which we can not duplicate with much confidence. In addition, the first and most important (at least in the lexicographical ordering sense) of the original criteria is a child characteristic - being malnourished - that the program is attempting to affect directly. Because we do not have baseline data on children in P, we cannot infer their pre-program nutritional status and, in particular, if they were malnourished at the time of entry into the program. Thus, we are aware of at least one important omitted variable that likely affects both program entry and program outcomes, particularly the anthropometric outcomes.

The data were collected in two rounds. Table 2 shows the sample sizes in groups P, A and B for the first and second rounds with and without imposing eligibility on the samples. The first round of data consists of 1198 participant (P) children, 1227 group A children and 628 group B children interviewed between November 1995 and May 1996. The second round consists of 2420 participant children, 2205 group A children and 1732 group B children interviewed between November 1997 and May 1998. Imposing eligibility criteria leads to a substantial reduction in the comparison group sample sizes - roughly cutting them in half. A subset of the children are observed in both rounds: 364 participants, 745 group A and 392 group B children.

4.2. Eligibility Criteria

To participate in PIDI, families are also required to meet eligibility criteria. The initial eligibility requirements (in effect at the time of the collection of the data that we use for analysis) were that candidates would be taken who were 6-72 months of age living in the poor urban communities selected by the program according to whether they met the following criteria (in order): (1) malnourished children, (2) children with working parents at risk of lack of supervision, (3) children who had been maltreated, (4) children who lived with only one parent or another relative, (5) children with four or more siblings, and (6) younger children. These criteria were in part subjective (particularly the first and third) and were supposed to be applied lexicographically. They subsequently were replaced by a more objective eligibility index that awards one point if

²⁷Equivalently, one could not impose eligibility on the sample and include everybody in the program participation model with an indicator variable for whether persons are eligible for the program. Ineligible persons would have a predicted probability of participating in the program equal to 0 and would therefore be excluded in the matching analysis by the support restriction.

the family has (a) no running water in the household, (b) no sewer system, (c) no more than two rooms in addition to the bathroom and kitchen in the house, (d) no bathroom or latrine in the household, (e) no separate kitchen, (f) more than four children, (g) the mother has five grades or less of schooling and (h) the father is unemployed. Two points are awarded if (a) the family has only a mother or a father or (b) if the mother of the family worked outside the household. A total of six points are required to be eligible for the program. The second index has fixed weights rather than the lexicographical one used initially. It also focuses more on household characteristics and does not include the more subjective aspects of the previous one – i.e. children being malnourished or maltreated. Nevertheless, in some general sense both the original and the current criteria attempt to identify children from poor socioeconomic families with limited provision of home childcare.²⁸

4.3. Variables

The PIDI Data Sets provide detailed information on parental, household and child characteristics. There is information, for example, on income sources, educational attainment and occupations of parents, fertility and reproductive histories, family structure, housing characteristics and possession of durable goods. For all children in the sample households, between six and 72 months of age, there are data on cognitive, psycho-social and anthropometric test score measures. The specific outcome measures that we examine in this paper are the following: (i) height-for-age percentile, (ii) height z-score below a threshold (<-3)²⁹, (iii) weight-for-age percentile, (iv) weight z-score below a threshold (<-2), (v) weight-for-height, (vi) bulk motor skills (BMS), (vii) fine motor skills (FMS), (viii) language-auditory skills (LAS) and (ix) psycho-social skills (PSS).

The first five are anthropometric measures, the next three are measures of cognitive skills and the last one is a psycho-social outcome measure.³⁰ The test score outcomes (vi,vii,viii,and ix)

²⁸Because of the latter concern about home supervision, children are more likely to become eligible for the program if their mothers work even if that implies more family income, *ceteris paribus*.

²⁹Z-scores give the number of standard deviations from the mean. They are widely used in the nutrition literature to characterize the degree of malnutrition (i.e., <-2 indicating moderate malnutrition and a value <-3 indicating severe malnutrition). Z-scores are increasingly used in the economic literature on the determinants of and impact of malnutrition (e.g. see some of the studies surveyed in Strauss and Thomas, 1998).

³⁰The cognitive outcomes and psycho-social outcomes are measured by a battery of 32 questions, but our analysis focuses on the summary scores. Appendix A provides additional information on the variables contained in the data sets.

are significantly highly correlated with each other, with a statistically significant Kendall-Tau coefficient of 0.8-0.9 for each of the pairwise correlations. Height and weight percentile measures are weakly positively correlated (Kendall-Tau=0.43). Height-for-age percentile is only slightly positively correlated with the test score outcome measures (with a Kendall-Tau coefficient equal to 0.06 for each of the test score measures). The pairwise correlations between weight-for-age and the test score outcomes are all insignificantly different from zero.

4.4. Comparison of Group Mean Characteristics

In Tables 3 and 4, we compare the characteristics of the parents and of the households for children participating in PIDI (group P) with those of nonparticipating children (in groups A and B), with and without imposing eligibility on the A and B samples and with group P subdivided by duration of program participation between one month or less and two months or more.

Table 3a compares characteristics of the mothers. About 8% of mothers in the PIDI group have no education and cannot read or write, which does not differ much from the values for the eligible and total sample B but is slightly higher than for the eligible and total sample A.³¹ For labor force participation there is a much more notable difference: PIDI mothers are much more likely to participate in the workplace (87% in group P compared to 55% in group A and 54% in group B). Much of this difference is eliminated by imposing eligibility criteria for the program on samples A and B; the labor force participation patterns become much more similar to those in sample P - 81% and 80% respectively. A comparison of the incomes of working mothers shows that PIDI mothers have significantly lower incomes, even though they work on average more hours per day. Imposing eligibility criteria lowers the incomes of group A and B mothers and brings them closer, but still not to the level of the PIDI mothers. PIDI mothers are also more likely to work in clerical jobs, while mothers from groups A and B are more likely to be self-employed. Among PIDI mothers in the two duration subsamples (shown in the last two columns), there are no significant differences.

Table 3b compares characteristics for the fathers. Paternal differences are not as striking as the maternal ones. Fathers' education levels are also lower in the PIDI group than in group A (24% with basic or no education compared to 16%) but about the same as in group B (26%).

³¹44% of PIDI mothers have only basic or no education, compared with 34% of group A and 46% of mothers in group B (47% and 61% if A and B are limited to those eligible).

For the eligible A group the percentage is close to that for the PIDI group (24%) but that for the eligible B group is higher (36%). The literacy rates among fathers are above 98% for all the groups, and are somewhat higher than for mothers in each of the groups (comparing Tables 3a and 3b). Fathers in the PIDI group appear to have less stable employment and are more likely to be employed in occasional work than are the fathers in the other groups; but if the eligibility criteria are applied, if anything there is a reversal in these comparisons. PIDI fathers also have a somewhat lower rate of unemployment than fathers in the A and B groups. Average number of hours worked are comparable across groups. Income levels are lower for PIDI fathers, but imposing eligibility on groups A and B makes the average income levels similar across groups. Within the P group there are not significant differences for the two subsamples defined by program duration.

Table 4 compares other characteristics of the household and reveals differences across groups in terms of family structure. PIDI households are less likely to have both parents residing in the household (78% in P versus 86% in A and 87% in B)- a difference mostly attributable to a greater number of PIDI households with only a mother present (20% of group P families do not have a father in the household). Group differences are reduced substantially if the eligibility criteria are applied to groups A and B. Average total household income and per capita income is lowest among PIDI families, though again imposing eligibility on the comparison groups helps to reduce the difference. The average number of persons per household is 5.2 in PIDI families, which is slightly lower than group A and B eligibles. Group B families have a significantly higher rate of home ownership than groups P and A (that persists even if the eligibility is imposed). PIDI households tend to be more active participants in outside organizations than are households in group B, which are somewhat more active than those in group A. Again, there are generally not significant differences between the two P subsamples defined by duration with the exception that those who have participated longer also tend to participate significantly more in outside organizations.

In summary, in terms of the observed mothers', fathers' and other household characteristics that are summarized in Tables 3 and 4, the total A and B samples tend to be better off than the P sample.³² If the eligibility criteria are applied, the differences are reduced and in a few cases reversed. Applying the eligibility criteria, therefore, makes the comparison samples based

³²There has been considerable concern in the policy-oriented development literature about how well social programs are successfully targeted to the poor (e.g., van de Walle and Nead, 1995). These comparisons suggest some success in targeting PIDI towards the poorer households in poor communities.

on A and B much more similar to group P, though groups A and B still probably on the whole have more resources. Subdividing the P sample into subsamples for one month and less versus two months or greater duration leads to no significant differences in the subsample means, with the single exception of greater participation in outside organizations by households with greater duration.

4.4.1. Child characteristics

Figure 1 compares the age distribution for PIDI participating children to those of children in eligible groups A and B and between the two P subsamples. It reveals major differences, with program participants tending to be more concentrated in the middle age ranges. As might be expected, those with longer program durations tend to be older than those in A and B. Few children in the program are under 20 months of age, and the mode of the age distribution is around 40 months.

Figure 2 compares children in the eligible B, eligible A and P groups in terms of weight, height and four test score outcome measures, conditioning here only on the age in months of the child. From the figure, it is apparent that PIDI children are short for their age at ages above 12 months. For weight, there are no discernable group differences. The test score comparisons do not show any distinct advantage or disadvantage for children in the PIDI group. Of course, these findings could be consistent with a positive effect of the program because PIDI families tend to have lower incomes, lower parental education levels and less stable family structure, which are all characteristics that we might associate with inferior child nutrition and test score outcomes.

If we divide the P sample by length of time spent in the program, the results are suggestive of a positive impact of the program for children who have been in the program for some time. Figure 3 plots the outcome measures for group A-eligible, B-eligible children and for P children who have participated at least 13 months. The PIDI group appears to do better on average in terms of cognitive test scores than children from the other groups, but this difference is not necessarily attributable to the program as it may be due to preexisting differences between program participants and nonparticipants.

4.5. Determinants of program participation

The probability of program participation plays an important role in estimating program effects by the matching method as described in section 3.1. The mean comparisons in Tables 3 and 4 indicate that groups A, B and P differ along several dimensions, each of which could be relevant to the program participation decision. As described in section 3, we estimate a logistic model for the probability of participating in the program using group P and the eligibles in group B, the two groups that selected into and out of the program. We select the set of regressors from those shown in Tables 3 and 4 to maximize the percentage correctly classified by the hit-or-miss criterion. Under the resulting model, 79% of the observations are correctly classified.³³ The regressors in the model are listed in Appendix C. The most useful predictors of participation include (i) presence of mother in the household, (ii) education level of the mother, (iii) number of children, and (iv) education level and monthly income of the father.

For group A, it is impossible to know which families would have elected to participate in the program had the program been available to them. However, under the assumption that the same participation process governs group A decisions as for group B, we can impute probabilities of participation for group A families using the coefficients from the participation model estimated on groups P and B.³⁴

The first column of Figure 4 plots the log-odds ratio for participating children and for eligible children in the B and A groups. For groups P and A the supports of the log-odds ratios overlap, but if group B is used as a comparison group, there are some high values of the log-odds ratio observed for program participants for which no matching values can be found for children in the B group. This limits the range of values over which treatment impacts can be estimated.

Our estimates of the marginal effects of longer durations in the program are based on the survival probability corresponding to the probability that duration in the program is two months or more.³⁵ Appendix C lists of set of regressors we used for this model (also chosen using the hit-or-miss criterion with a correct classification rate of 76%.) The log odds ratio of the survival probabilities is plotted in the second column of figure 4.³⁶

³³77% of participants and 84% of the controls are correctly classified.

³⁴This assumes that a similar model for participation would apply to group A, which requires assuming that there are no significant program site effects. Site effects were not found to be an important predictor of participation in the group P and B logistic model.

³⁵This is a version of the estimator described in section 2.1.2 that integrates over the observed program durations greater than or equal to 2.

³⁶When we use the survival probability calculated only using the data on program participants, there is no

4.6. Impacts Estimated by Traditional Regression Methods

Before presenting impact estimates based on the matching estimators, we first report for comparison the estimates that are obtained by simple regression estimators that are described in section 3. First, we estimate a cross-sectional regression model for the three cognitive development tests, the psycho-social ability test, and three anthropometric indicators, based on the combined P and eligible B samples. As discussed in Section 3.2, regression estimators make strong functional form assumptions on the relationship between outcomes, duration, age and other variables, relative to our matching estimators. The specification we estimate includes as independent variables a dichotomous variable for participation in PIDI, a cubic in duration in PIDI, a cubic in the child’s age, the child’s sex and some family background characteristics related to parents’ presence in the household, education, ages, fathers’ job and income, and the number of children in the household.³⁷ Figure 5 plots the estimated program impact as a function of duration in the program.

We can see that estimated program impacts on test scores are mostly positive and on the order of one additional question correct (out of a possible 32). As will be shown below, estimates obtained by matching indicate larger impacts of the program on test scores for children with longer durations of exposure. For the anthropometric outcomes, we find the counterintuitive re-

need to use the odds ratio in matching as there is no choice-based sampling problem. However, for convenience we also match on the log-odds ratio for this group.

³⁷For brevity, the regression estimates are shown in Appendix E, Table E.1, which is available on request from the authors. The conventional interpretations of the estimates are: (a) these relations explain considerable shares of the variance in the four test scores (84% or more) but much less of the sample variance in the anthropometric indicators (about 4%); (b) family background characteristics are significant determinants of all the child outcomes (the family background variables are highly significant and F-tests reject the null that they are insignificant at conventional significance levels); (c) being in PIDI (controlling for family background and duration) is negatively associated with six of the seven outcome measures (significantly so at least at the 10% level in four cases), a surprising result that is consistent with at least initial program impact being negative and/or the program having successfully targeted children who were relatively limited in their pre-program development; (d) for the test score outcome measures, child individual characteristics are significant with ages in the middle range (for which the quadratic term dominates) experiencing the highest test scores and males doing significantly better on the bulk motor ability test but significantly worse on two of the three anthropometric indicators; and (e) the linear and cubic effects of duration in the program have positive coefficients (with the linear effects significantly nonzero at the 1% level for the test scores and the cubic ones significantly nonzero for two of the test scores) but the negative coefficient estimates (significantly nonzero for three of the test scores) for the quadratic terms suggest some tendencies for diminishing returns over a range.

sult of a negative impact of the program on weight and on height that is increasing in duration of time spent in the program. We do not find these estimates to be credible because large, negative impacts of the program on anthropometrics immediately upon program entry (as indicated by the estimated negative impact of PIDI participation on the intercept) are extremely unlikely.

Finally, we consider estimates based on the difference-in-difference (DID) estimator (discussed in Section 3.2) for the much smaller subset of children who are observed in both sample rounds in P and in eligible group B (see Table 2 for sample sizes and Appendix E, Table E.2 for the coefficient estimates—available on request from the authors). In this case, the estimates suggest that the effects of the program are negative for all but the language and auditory test score measure (See Figure 6). However, these estimates are quite imprecise due to the substantially reduced sample sizes.

4.7. Cumulative impacts estimated by the method of matching

We next describe estimated cumulative program impacts based on the matching estimators developed in section 3, first conditional on age only and then conditional on both age and duration of time in the program. Then we present results on the marginal impacts. In implementing the matching estimators, we choose bandwidth values by the least squares cross-validation (LSCV) method, which searches over a grid of possible bandwidth values and chooses the values that minimize the integrated-squared-error of the nonparametric estimators.³⁸

Table 5 compares the conditional-on-age difference in raw means of the outcome measures with the mean program impacts estimated by the cross-sectional matching estimator given in (3.3).³⁹ The column ‘Mean Diff’ shows the difference in raw mean outcomes, the column ‘Impact, mag’ shows the estimated program impact obtained by the matching method, and the column ‘Impact %’ gives the estimated program impact as a percentage of the average outcome measure for the comparison group children in the relevant age range. In parentheses, we report bootstrapped standard errors of the estimates.

³⁸The grid is three dimensional for estimating the conditional $D_P = 1$ expectation and two dimensional for estimating the conditional on $D_P = 0$ expectation. The values over which we searched are 1.0 through 16.0 with a stepsize of one for the log odds ratio, 1.0 through 28.0 for age with a step size of one month and 1.0 through 28.0 for duration with a step size of one month. See Jones et. al. (1996) for a discussion of the LSCV method and of alternative methods of choosing bandwidths in nonparametric regressions.

³⁹For the sake of brevity, we report only the estimates based only on group B in the text. Estimates based on the A comparison group are shown in Appendix E that is available on request.

The test score impacts are almost all positive for children age 37-54 months. For this age group, the program is estimated to increase test scores by roughly one additional correct item, which is about 5% of the average score within age classes of the untreated group. Although this impact may seem modest in magnitude, it is worth noting that the recently evaluated Tennessee class size experiment, which has widely been acclaimed in the United States as a successful program, found an increase in test score outcomes of only 6 percentile points (Krueger, 1998).

In terms of height percentiles, the estimated program impacts are positive at ages up to 36 months but mostly negative at older ages. For weight percentiles, estimated impacts are negative at ages younger than 36 months. These results, particularly for weight, might suggest that the program had initially a detrimental effect on the anthropometric outcomes. It is possible, however, that children in the program are simply shorter and weigh less to begin with for reasons that are not being fully captured by our model of program participation. This would be consistent with the primacy given to malnourishment in the initial program eligibility criteria if that were interpreted primarily in terms of weight (see section 2.1) and with the negative estimated impact of PIDI on child anthropometric outcomes in the conventional cross-sectional regressions.⁴⁰ There is support for a positive program impact on the ‘Height>-3’ indicator, which implies that the program helps keep children out of the extreme lower tail of the height distribution. However, the bootstrapped standard errors show that impacts on the anthropometric outcomes are less precisely measured than those on the test score outcomes and few are statistically significantly different from zero.

The top four panels of Table 6 report estimated impacts for the four test scores, now conditional on both age and duration. The estimates are obtained by first estimating mean impacts at each age and duration value observed in the data and then taking averages over the individual impacts within each age-duration class.⁴¹ The patterns of estimates strongly suggest that average impacts increase as length of exposure to treatment increases. Impacts are almost always positive for children who have participated in the program for at least 13 months (with only two exceptions both for children age 31-36 months who have participated 25+ months) and roughly twice the order of magnitude of the overall average impacts reported in Table 5. They are generally larger than those found under the cubic specification of section 4.3. For children with less than seven months exposure to the program, the estimated impacts are frequently

⁴⁰Though in the cross-sectional regressions (for which the coefficient estimates are reported in Appendix E) there also are negative coefficient estimates for the four test score outcomes.

⁴¹Averages are self-weighting by the duration density.

negative.

The bottom two panels of Table 6 show results for the anthropometric measures. We see that the negative weight percentile impacts observed in Table 5 are mostly confined to children who have participated in the program for less than seven months. For children who have participated longer, the estimates show a positive impact of the program on weight percentile. For height percentile, however, we still observe some large negative impacts for children older than 36 months as well as some fairly large positive impacts for younger children with short durations. It is highly unlikely that the program could either increase or decrease children’s height over a short time interval, or initially reduce children’s weight by as much as estimated. It is therefore likely that the estimated impacts for anthropometrics in part are due to unobservables that are not controlled by the cross-sectional matching procedure. Given that there is no pre-program baseline survey, an important unobservable is pre-program nutritional status on which, as noted in section 2.1, the initial program eligibility criteria placed primary emphasis.

We have carried out a similar analysis using as a source of comparison group data the sample of children living in a geographic area not served by the program (group A described in section 2). The implied impact estimates for test scores and for height are similar to those obtained for group B. The impact estimates for weight are more negative than those observed for group B over a wide range of duration values.⁴²

4.8. Marginal program impacts estimated by the method of matching

Because preprogram nutritional status represents an important unobservable that seems particularly relevant to the anthropometric outcomes, the preferred estimates are those for the marginal impact of the program for different durations of participation obtained using only data for program participants P. These estimates use participants with shorter durations as the comparison group for participants with longer durations and use matching to control for differences in child characteristics that affect the program duration rather than the participation decision. As described in section 3.1, the marginal estimator allows selectivity in the program to be based on unobservables, but assumes that children with longer durations can be made comparable to children with shorter durations by conditioning on observables.

Table 7 presents marginal impact estimates in a parallel format to Table 5. The test scores impacts now are generally positive for children of age 19 months+, rather than 37-54 months

⁴²For brevity, these results are reported in Appendix E, Tables E.3-E.4, available on request from the authors.

as in Table 5. The marginal estimates indicate not only generally somewhat larger effects than do the average estimates, but also the prevalence of positive benefits at younger ages. For the anthropometric indicators, the marginal program effects on mean weight-for-age percentile and mean height-for-age percentile are positive for a little over half the age ranges; although the standard errors on the anthropometric outcomes remain large. There again appears to be some support for an effect of the program on keeping children out of the lower tail of the height distribution (Height > -3).

Table 8 is parallel to Table 6, with estimated impacts conditional on age and duration in the program. The patterns on test scores generally show that marginal impacts increase as length of exposure to the program increases. They are most often positive and larger than the overall average marginal impacts in Table 8 for children who have participated in the program for at least six months. For children aged 19-36 months (there are relatively few participants younger than 19 months - see Figure 1), the estimated impacts on height and weight percentiles are generally positive for different durations, while for children older than 36 months they often are negative. For height percentiles, the impacts at shorter durations are still surprisingly large.⁴³ For weight percentiles the marginal effect estimates are more credible than the cumulative estimates (Table 6). These comparisons suggest that the first criterion for selecting children into the program (malnourishment) focused on low weight and not low stature.

5. Cost-Benefit Analysis

So far we have focussed only on the problem of estimating the benefits of the program. Next we consider whether the benefits outweigh the costs, which have been estimated to be about \$43/month per child enrolled (= \$516/year) by Ruiz (1996). We focus here exclusively on benefits in terms of earnings. There are four channels that we consider by which the preschool program can affect lifetime earnings: (1) by increasing cognitive skills as an adult (conditional on grades completed) that directly affects earnings, (2) by increasing physical stature as an adult that directly affects earnings, (3) by increasing the number of grades completed that

⁴³A possible explanation for this result, which unfortunately the lack of preprogram data makes it difficult for us to explore, is that parents tended to enroll their young children only when they considered them to be sufficiently mature and that their assessment of their child's maturity was based on criteria correlated with child's height. This could explain why among children enrolled in the program, those enrolled for longer durations tend to be taller.

directly affects earnings and the age of school completion a and (4) by decreasing the age of school completion without changing the number of grades completed. For the program to have an impact through channels (3) and (4), we are assuming that improved cognitive skills and nutrition as a child facilitates earlier entry into school, lessens repetition rates, and leads to more grades completed. Appendix D summarizes the empirical evidence on the importance of these four channels from the experiences of developing countries.

As our data does not provide information on how higher cognitive skills and better nutrition affects adult earnings and we are unaware of any such estimates for Bolivia, we draw on estimates from previous studies on other developing countries. One is a study by Stauss and Thomas (1997) that analyzes the relationship between adult earnings and height, body mass index (BMI), caloric consumption, protein consumption and education for male workers in a neighboring South American country, Brazil. It finds that a 1% increase in height leads to a 2.4% increase in adult male earnings, in a regression of log hourly wages on height and years of education.⁴⁴ To our knowledge, there has been no research on the cognitive skills-earnings relationship specifically for South American workers, so we base our cost-benefit analysis on a study by Alderman, Behrman, Ross and Sabot (1996) of the cognitive skills-earnings relationship for male workers in Pakistan. Their study has an advantage over some other studies in the literature in that it controls for the potential endogeneity of cognitive ability in the wage equation. As we only observe the children in our study at a very young age, we assume for the cost-benefit analysis that increases in height and cognitive ability as a child have a persistent effect and translate into equiproportional increases as an adult.^{45, 46}

The present discounted value of earnings associated with a 1% increase in height is calculated as follows. Let $y(s, c, h)$ be the annual earnings of individuals with grades completed s , cognitive ability c , and height h . Let a be the age of completing school. We draw a distinction between grades completed and rate of progression through grades because a number of students in Bolivia both start school late and repeat grades. Estimates from the 1990 third round of the *Encuesta Integrada de Hogares* that covers the ten most populous urban areas in Bolivia indicate that for

⁴⁴Their study uses a normal bias correction to control for selectivity into employment.

⁴⁵Measures of intelligence have been found to be highly correlated across ages. For example, the Berkeley Growth Study found a correlation of 0.71 between test scores measured at ages 4 and 17.

⁴⁶We use height for our illustration rather than BMI because this assumption is more dubious for BMI than for height. But, as noted below, we consider a relatively small percentage increase in height in comparison with those obtained for some of the estimators in section 4 because we expect that parents may have selected taller children for consideration for the program.

16 year olds in urban Bolivia the gap in grades completed due to factors such as late starting and grade repetition was 10-16% of the grades actually completed, or between 0.9 and 1.4 grades in comparison with a mean of 8.6 grades actually completed. Let r be an externally determined real rate of interest and T the length of working life, assumed not to depend on $s, c, a,$ or h . In Bolivia, recent life expectancies at birth are approximately 60 years. The present discounted value of earnings for a given (s, a, h, c) vector is $V(s, a, h, c) = \int_a^{60} y(s, h, c)e^{-rt} dt$.⁴⁷ This yields a present discounted value of earnings equal to

$$V(s, a, h, c) = \frac{y(s, c, h)}{r}(e^{-ra} - e^{-r60})$$

The expected impact of a 2% increase in height is:

$$\bar{y}(s)(1 + 0.024)(0.02)\frac{(e^{-ra} - e^{-r60})}{r},$$

where $\bar{y}(s)$ is the average earnings for men with s grades completed and we use the results from the Strauss and Thomas study indicating that a 1% increase in height, controlling for schooling, leads to a 2.4% increase in earnings. We calculate the effect of an increase in adult cognitive skills analogously using Alderman et. al.'s (1996) reported finding that a one percent increase in cognitive skills increases earnings by 0.23%.⁴⁸

The earnings gain that would result solely from a decrease in the school completion age from a_1 to a_2 without changing the level of school attainment is given by

$$\bar{y}(s)\frac{[e^{-ra_2} - e^{-ra_1}]}{r}.$$

An increase in the level of attainment from s_1 to s_2 has two possibly partially offsetting effects (as in Mincer, 1958). It increases earnings capacity but also potentially decreases the amount of time available for work operating through a . To denote the dependence of a on s write $a(s)$. The benefit of increasing schooling from s_1 to s_2 is given by

$$\frac{\bar{y}(s_2)}{r}(e^{-ra(s_2)} - e^{-r60}) - \frac{\bar{y}(s_1)}{r}(e^{-ra(s_1)} - e^{-r60})$$

⁴⁷We assume for simplicity that the earnings path is flat over the life-cycle (i.e. $y(s, h, c)$ does not depend on $t - a$ after controlling for s).

⁴⁸Their study finds that a 7.3% increase in cognitive skills, evaluated at the mean, leads to a 1.3% increase in earnings, conditional on years of schooling. We have converted their estimates to the gain expected from a 1% increase in skills.

On the cost side, the cost for participating in the program for four years between ages two and five is given by

$$\text{cost} = \$516 \int_2^5 e^{-rt} dt = \frac{\$516}{r} (e^{-2r} - e^{-r5}).$$

Table 9 reports the cost-benefit estimates under alternative hypothetical program impacts that are in the range of some of the impacts observed in the impact analysis of section 4.5 (Table 8) and for average male earnings levels associated with three different education levels: 8 years, 11 years and 14 years of education.⁴⁹ The top panel of the table considers an impact of 2% on height and the second panel an impact of 5% on cognitive skills. The third panel considers a hypothetical impact of a one-year decrease in age of school completion and the fourth panel a one-year increase in grades completed and a corresponding one year increase in the age of school completion. The bottom panel considers all of these impacts jointly. For each case, we give estimates for two values of the discount rate, $r = 3\%$ or $r = 5\%$.

As seen in the table, when the impacts are examined one at a time the benefit/cost ratio is frequently less than one. The single impact that has the largest effect among the ones considered in the first four panels is increasing the number of grades completed (under the assumption that there is a corresponding one-year increase in the age of completion), which generates a benefit-cost ratio greater than one for both discount rates and both education levels. When multiple types of program impacts are considered together, however, the benefit/cost ratios are quite high and range from 1.7 to 3.7. (See the bottom panel.)

6. Conclusions

In this paper we analyze the impact of a preschool program in a developing country using a relatively large, nonexperimental dataset. To do so, we generalize matching methods by allowing the impact of the program to vary with the duration of time spent in the program, allowing impacts to depend in a flexible way on the age of the child, and by developing a new marginal treatment effect estimator that uses only data on program participants and does not require assumptions on the process governing selection into the program. Another advantage of the marginal effect estimator is that it is not demanding in terms of its data requirements, requiring

⁴⁹Mean earnings are calculated from the sample of adult males in the group A comparison group data (This group was chosen because it is not self-selected on program participation). The mode number of years of education for these males is 8 years.

only data on program participants. However, the estimator assumes that program participants with differing duration levels can be made comparable by conditioning on observed child and family characteristics.

We applied the estimators to study the effectiveness of a preschool program in Bolivia aimed at improving early cognitive skills and nutrition based on a dynamic model that we developed for the decision to enroll children that provided an interpretation for the treatment impact estimates and guided our selection of matching variables. We considered three sources of comparison group data: children living in regions where the program was available who did not participate in it, children from regions where the program was not available, and children with short program durations. Impact estimates based on cross-sectional regression-based estimators suggest that the program has a positive effect on test scores and a negative effect on height and weight percentiles. Our cumulative matching estimators show that test scores gains depend strongly on duration of exposure to the program, with positive effects observed for children who participated at least 18 months and increasing effects observed with longer durations.

However, cumulative impact estimates for anthropometric measures strongly suggest that the identifying assumptions that justify a matching-on-observables approach are not satisfied in our data, as it is extremely unlikely that the program would have large, negative impacts on weight immediately upon program entry. Estimates of the marginal effects, based on the alternative marginal effect estimator are more plausible and indicate a positive effect of the program on weight percentiles, again for children who have participated in the program for an extended period of time. There is also support for an effect of the program in preventing children from falling into the bottom of the height distribution. Marginal impact estimates for test scores also indicate the prevalence of positive benefits at shorter durations, usually after 7 months, and at younger ages. These patterns differ substantially from those suggested by standard regression estimators.

Our cost-benefit analysis considered a few different channels by which the program might be expected to have an effect on lifetime earnings, including a direct effect of the program on earnings operating through greater physical stature and cognitive skills and indirect effects operating through less time spent in school to achieve a given level of education and/or higher educational attainment levels. When all the channels are combined, the expected benefit of the program outweighs the costs by a fair amount under the assumptions of our simulations.

References

- [1] Alderman, Harold, Jere R. Behrman, Victor Levy and Rekha Menon (1997): "Child Nutrition, Child Health, and School Enrollment: A Longitudinal Analysis," Washington, D.C.: World Bank, Policy Research Working Paper #1700.
- [2] Alderman, Harold, Jere R. Behrman, David Ross and Richard Sabot (1996): "The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market," *Oxford Bulletin of Economics and Statistics*, Vol. 58, No. 1, 29-56.
- [3] Baker, Amy J. L., Piotrkowski, Chaya S. and Brooks-Gunn, Jeanne (1998): "Effects of the Home Instruction Program for Preschool Youngsters (HIPPY) on Children's School Performance at the End of the Program and One Year Later," *Early Childhood Research Quarterly*, 13, No. 4, 571-588.
- [4] Barnett, Stephen (1992): "The Benefits of Compulsory Preschool Education," *Journal of Human Resources*, Vol. 27, #2, 279-312.
- [5] Behrman, Jere R. (1993): "The Economic Rationale for Investing in Nutrition in Developing Countries," *World Development*, Vol. 21, No.11, 1749-1771.
- [6] Behrman, Jere R. and Anil B. Deolalikar (1989): "Wages and Labor Supply in Rural India: The Role of Health, Nutrition and Seasonality," *Causes and Implications of Seasonal Variability in Household Food Security*, ed. David E. Sahn, 107-18. Baltimore, MD: The Johns Hopkins University Press.
- [7] Bloom, H., L. Orr, G. Dave, S.H. Bell, and F. Doolittle (1993): *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months*, Bethesda, Maryland: Abt. Associates.
- [8] Boissiere, Maurice, John B. Knight, and Richard H. Sabot (1985): "Earnings, Schooling, Ability and Cognitive Skills" *American Economic Review*, 75, 1016-30.
- [9] Cleveland, W. (1979): "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, 74, 829-836.
- [10] Currie, Janet and Duncan Thomas (1995): "Does Head Start Make a Difference?" *American Economic Review*, Vol. 85, 3, p.341.

- [11] _____(1996): “Does Head Start Help Hispanic Children?” NBER working paper.
- [12] _____(1999): “Early Test Scores, Socioeconomic Status, and Future Outcomes” NBER working paper #W6943.
- [13] Dehejia, R. H. and S. Wahba (1998): “Causal Inference in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs” forthcoming *JASA*.
- [14] Deolalikar, Anil B. (1988): “Nutrition and Labor Market Productivity in Agriculture: Estimates for Rural South India,” *Review of Economics and Statistics*, 70:3, August, 406-13.
- [15] Fan, J. (1992): “Design Adaptive Nonparametric Regression,” *Journal of the American Statistical Association*, 87, 998-1004.
- [16] Fan, J. and Gijbels, I. (1996): *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.
- [17] Garber, H. L. (1988): *The Milwaukee Project: Preventing Mental Retardation in Children at Risk*. Washington, DC: American Association on Mental Retardation.
- [18] Glewwe, Paul (1996): “The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment,” *Journal of Development Economics*, 51:2, December: 267-90.
- [19] Glewwe, Paul and Jacoby, Hanan G. (1995): “An Economic Analysis of Delayed Primary School Enrollment in a Low-Income Country: The Role of Early Childhood Nutrition,” *The Review of Economics and Statistics*, p. 156-159.
- [20] Glewwe, Paul and Hanan G. Jacoby, and Elizabeth King (1998) “Early Childhood Nutrition and Academic Achievement: Analysis Using Longitudinal Data,” Washington, DC: The World Bank, mimeo.
- [21] Grantham-McGregor, S., C. Walker, S. Chang, and C. Powell (1997): “Effects of Early Childhood Supplementation with and without Stimulation on Later Development in Stunted Jamaican Children” *American Journal of Clinical Nutrition* 66:247-453.

- [22] Haddad, Lawrence, and Howarth Bouis (1991): "The Impact of Nutritional Status on Agricultural Productivity: Wage Evidence from the Philippines," *Oxford Bulletin of Economics and Statistics*, 53:1, no. February: 45-68.
- [23] Haas, J., S. Murdoch, J. Rivera, and R. Martorell (1996): "Early Nutrition and Later Physical Work Capacity," *Nutrition Reviews*, 54: S41-S48.
- [24] Heckman, J. J. (1997): "Randomization as an Instrumental Variables Estimator: A Study of Implicit Behavioral Assumptions in One Widely-used Estimator," *Journal of Human Resources*, 32, 442-462.
- [25] Heckman, J. J. (1992): "Randomization and Social Program Evaluation," in C. Manski and I Garfinkle, eds. (Cambridge, Mass.: Harvard University Press).
- [26] Heckman, J. J. and P. Todd (1995): "Adapting Propensity Score Matching and Selection Models to Choice-based Samples," unpublished manuscript, University of Chicago.
- [27] Heckman, J. J., H. Ichimura, J. Smith and P. Todd (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica* , Vol. 66, 1017-1098.
- [28] Heckman, J.J., H. Ichimura and P. Todd (1997): "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, Vol. 64(4).
- [29] Heckman, J. J., H. Ichimura and P. Todd (1998), "Matching As An Econometric Evaluation Estimator," *Review of Economic Studies*, Vol. 65, 261-294.
- [30] Heckman, J. J., Lalonde, R. and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs" in *Handbook of Labor Economics, Volume III*, eds. O. Ashenfelter and D. Card (Elsevier: Amsterdam), forthcoming.
- [31] Herrnstein, Richard and Charles Murray (1994): *The Bell Curve*. New York: Free Press.
- [32] Jamison, Dean T. (1986): "Child Nutrition and School Performance in China," *Journal of Development Economics*, Vol. 20, No. 2, pp. 299-310.
- [33] Johnston, F., S. Low, Y. de Baessa and R. MacVean (1987): Interaction of Nutritional and Socioeconomic Status as Determinants of Cognitive Achievement in Disadvantaged Urban Guatemalan Children," *American Journal of Physical Anthropology*, 73: 501-506.

- [34] Krueger, Alan (1998): "Reassessing the view that American Schools are Broken," *Federal Reserve Bank of New York Economic Policy Review*, Vol. 4, No. 1, 29-43.
- [35] Lavy, Victor, Jennifer Spratt, and Nathalie Leboucher (1997): "Patterns of Incidence and Change in Moroccan Literacy," *Comparative Education Review*, 41:2, no. May.
- [36] Manski, C. and S. Lerman (1977): "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, Vol. 45, 1977-1988.
- [37] Martorell, R. (1995): "Results and Implications of the INCAP Follow-up Study," *Journal of Nutrition*, 125 (Suppl): 1127S-1138S.
- [38] Martorell, R. (1999): "The Nature of Child Malnutrition and its Long-term Implications," *Food and Nutrition Bulletin*, Vol. 20, 288-292.
- [39] Martorell, R., K. L. Kahn and D. G. Schroeder (1994): "Reversibility of Stunting: Epidemiological Findings in Children from Developing Countries," *European Journal of Clinical Nutrition*, Vol. 48 (Suppl): S45-S57.
- [40] Martorell, Reynaldo, Juan Rivera, and Haley Kaplowitz (1989): "Consequences of Stunting in Early Childhood for Adult Body Size in Rural Guatemala," Stanford, CA: Food Policy Research, Stanford University, Mimeo.
- [41] Mincer, J. (1958): "Investment in Human Capital and Personal Income Distribution" *Journal of Political Economy*, 281-302.
- [42] Moock, Peter R. and Joanne Leslie (1986): "Childhood Malnutrition and Schooling in the Terai Region of Nepal," *Journal of Development Economics*, Vol. 20, No.1, 33-52.
- [43] Murnane, R., Willett, J. B. and F. Levy (1995): "The Growing Importance of Cognitive Skills in Wage Determination," *Review of Economics and Statistics*, 77: 2, 251-266.
- [44] Myers, Robert (1995): *The Twelve Who Survive: Strengthening Programmes of Early Childhood Development in the Third World*. High Scope Press: Michigan.
- [45] Neal, D. and W. Johnson (1996): "The Role of Pre-Market Factors in Black-White Wage Differences" *Journal of Political Economy*, Vol. 104, No. 5, Oct, 869-895.
- [46] Pollitt, Ernesto (1990): *Malnutrition and Infection in the Classroom*, Paris: UNESCO.

- [47] Pollitt, Ernesto, Kathleen S. Gorman, Patrice L. Engle, Reynaldo Martorell, and Juan Rivera (1993): “Early Supplemental Feeding and Cognition,” *Monographs of the Society for Research on Child Development*, Serial No. 235, 58:7.
- [48] Psacharopoulos, George (1994): “Returns to Investment in Education: A Global Update,” *World Development* 22:9, no. September: 1325-44.
- [49] Ramey, Craig T., Campbell, Frances A. and Clancy Blair (1998): “Enhancing the Life Course for High-Risk Children” *Social Programs that Work*, ed. Jonathon Crane, Russell Sage Foundation: New York., p. 184-199.
- [50] Reynolds, A. J. (1998): “The Chicago Child-Parent Center and Expansion Program: A Study of Extended Early Childhood Intervention,” *Social Programs that Work*, ed. Jonathon Crane, Russell Sage Foundation: New York., p. 110-147.
- [51] Rosenbaum, P. and D. Rubin (1982), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41-55.
- [52] Rosenzweig, Mark R. (1995): “Why Are There Returns to Schooling?” *American Economic Review*, 85:2, no. May: 153-58.
- [53] Ruiz, Fernando (1996): “Estudio de Costos del Proyecto Integral de Desarrollo Infantil (PIDI),” La Paz, Bolivia: Ruiz Guissani Consultores, mimeo.
- [54] Strauss, John (1986): “Does Better Nutrition Raise Farm Productivity?” *Journal of Political Economy* 94, no. April: 297-320.
- [55] Strauss, John and Duncan Thomas (1998): “Health, Nutrition and Economic Development,” *Journal of Economic Literature*, Vol. 36, No. 2, 766-817.
- [56] ——— (1997): “Health and Wages: Evidence on Men and Women in Urban Brazil,” *Journal of Econometrics*, 1997, Vol. 77, No. 1, 159-185.
- [57] ——— (1995): “Human Resources: Empirical Modeling of Household and Family Decisions,” in J. Behrman and T.N. Srinivasan, eds. *Handbook of Development Economics*, Volume 3A, Amsterdam: North Holland Publishing Company.

- [58] Sweinhart, Lawrence J. and Weikart, David P. (1998): “High/Scope Perry Preschool Program Effects at Age Twenty-Seven,” *Social Programs that Work*, ed. Jonathon Crane, Russell Sage Foundation: New York., p. 148-162.
- [59] Todd, Petra E. and Kenneth I. Wolpin (2000): “On the Specification and Estimation of the Production Function for Cognitive Achievement,” manuscript under submission.
- [60] van de Walle, Dominique and Kimberly Nead (1995): *Public Spending and the Poor*, Johns Hopkins University Press: Baltimore, MD.
- [61] World Bank (1997): *World Development Report, 1997, The State in a Changing World*, Oxford University Press for the World Bank: New York and Oxford.

Appendix A: Data Appendix

The PIDI survey consists of five modules: two about the household, one about women in the household, one about the children and, for PIDI families, a module about the PIDI center supervisor. The first module gathers socioeconomic data for all household members, including information about parents' educational attainment levels, income sources, father's and mother's occupations, and family structure. The second module gathers information on fertility and reproductive histories for all females in the household between the ages of 13 and 49. The third module gathers a variety of information on the children in the household including anthropometric measures, test scores on cognitive and psycho-social tests, information on vaccination records, recent illnesses and some qualitative data on parent-child interactions. The fourth module gathers information on household living conditions, information on whether the family possesses certain types of durable goods, data on the households' interaction with local community groups, and qualitative data on the parents' opinions of the PIDI program. The fifth module provides information on the characteristics of the PIDI center coordinators.

Appendix B: Technical Appendix on local linear regression

In implementing the nonparametric matching estimators, we estimate conditional expectations by local linear regression methods (LLR).⁵⁰ The local linear estimator for $E[y_i | z_i = z_0]$ can be computed from the minimization problem

$$\min_{a,b} \sum_{i=1}^n (y_i - a - b_1(z_i - z_0))^2 K\left(\frac{z_i - z_0}{h_n}\right),$$

where $K(\cdot)$ is a kernel function and $h_n > 0$ is a bandwidth which converges to zero as $n \rightarrow \infty$. The estimator of the conditional mean is \hat{a} . If b_1 were constrained to equal one, then \hat{a} would give the standard kernel regression estimator. Thus, kernel regression can be viewed as a special case of local linear regression (LLR).

Fan (1992) shows that the local linear estimator has the same variance as the kernel estimator but has a lower order bias at boundary points.⁵¹ The smaller bias associated with the LLR estimator implies that it is more rate-efficient than the kernel estimator. Another advantage of emphasized by Fan is that the bias of the LLR estimator does not depend on the design density of the data. Because of these advantages, local linear methods are usually a better choice than standard kernel methods for nonparametric regression. The local linear estimator is asymptotically normal with a rate of convergence equal to $\sqrt{nh_n^k}$, where k is the dimension of z . In our application, the estimators have $k = 2$ or $k = 3$.

The kernel function we use in the empirical work is the biweight kernel (sometimes also called a quartic kernel). Bandwidth values are selected by least squares cross-validation as described in the text.

⁵⁰Local polynomial estimators were developed in the early statistics literature by Cleveland (1979) and Stone (1977). They are further developed in Fan (1992) and have more recently considered in the econometrics literature by Heckman, Ichimura, Smith and Todd (1998).

⁵¹The advantage stems from the fact that local linear regression imposes an orthogonality condition between the regressors and the residuals that is not imposed under kernel regression. See Fan (1992).

Appendix C: List of Variables Included in Program Participation Model

In this appendix, we list the variables that were included in the discrete choice models for program participation and for the probability of experiencing a duration that exceeds one month (used in comparing groups with durations >1 and durations ≤ 1). The following list gives the variables included in the models. The subset of variables and interactions were selected from a larger set of variables available in the dataset to maximize the percentage of observations correctly classified under the model.

Variables included in the model for program participation: age in months of child, sex of child, indicators for whether mother and father reside in the household, education level of mother, job type of father, monthly income of father, number of siblings, number of rooms in the house, indicator for whether family owns house, indicator for whether house has running water, indicator for whether house has a bathroom, indicator for whether house has a television set, interaction terms between the number of rooms in the house and the age of the child, interaction between employment status of the father and age of child, interaction between number of siblings and age of child, interaction between monthly income of the father and number of siblings, interaction between education level of the mother and age of child.

Variables in the model for the probability of experiencing a duration that exceeds one month: age of the child, sex of the child, indicator for whether the family participates in outside organizations, indicators for whether mother and father reside in the household, education level of the mother, employment status of the mother, job type of the mother, age of the father, education of the father, monthly income of the father, number of siblings, number of rooms in the house, indicator for whether family owns the house, indicator for whether house has running water, indicator for whether bathroom in the house, indicator for whether the household has a television set, interaction between the number of rooms and the age of the child, interaction between employment status of the father and age of child, interaction between employment status of the mother and the age of the child, interaction between number of siblings and age of child, interaction between age of father and number of siblings, interaction between monthly income of the father and number of siblings, interaction between education level of the mother and age of child.

Appendix D: Empirical Evidence on Impact of Preschool Child Nutrition and Cognitive Development on Post-Schooling Earnings

In order to simulate benefits of improved preschool child nutrition and cognitive development on adult earnings, a number of channels must be considered as noted in the text: (1) increasing cognitive skills as an adult (conditional on grades of schooling completed) that directly affects earnings (2) increasing physical stature as an adult that directly affects earnings, (3) increasing the number of grades completed that directly affects earnings and the age of school completion a and (4) only changing the age of school completion without changing the number of grades completed. There is piecemeal empirical evidence of significant effects through all four of the channels for developing countries:

Evidence on (1) – Alderman et. al. (1996) for rural Pakistan; Boissiere, Knight and Sabot (1985) for urban Kenya and Tanzania; Glewwe (1996) for Ghana; and Lavy, Spratt and Leboucher (1997) for Morocco;

Evidence on (2) – Behrman and Deolalikar (1989) and Deolalikar (1988) for rural India; Haddad and Bouis (1991) for rural Phillippines; Strauss (1986) for Cote d'Ivoire; Thomas and Strauss (1997) for Brazil; and Behrman (1993) for the more general experience in developing countries.

Evidence on (1) and (2) – Grantham-McGregor et. al. (1997) for Jamaica; Martorell (1995) and Martorell, Rivera and Kaplowitz (1989) for rural Guatemala; and Haas, et. al. (1996), Martorell (1999) and Martorell, Kahn and Shroeder (1994) for the more general experience in developing countries.

Evidence on (3) – hundreds of studies, many of which are surveyed in Psacharopoulos (1994) and Rosenzweig (1995).

Evidence on (3) and (4) – Jamison (1986) for China; Moock and Leslie (1986) for Nepal; and Behrman (1993) and Pollitt (1990) for the more general experience in developing countries.

Evidence on (4) – Alderman et. al. (1997) for rural Pakistan; Glewwe and Jacoby (1995) for Ghana; and Glewwe, Jacoby and King (1998) for the Phillippines.

For our illustrative simulations, we use estimates from Alderman et. al. (1996) for (1) and Thomas and Strauss for (2) under the assumption in both cases that there is a strong persistence

of changes in preschool child anthropometric and cognitive development so that the percentage changes for adults equal those we estimate for children. We also use the estimate in the latter study for the impact of grades completed in schooling on earnings in (3). The studies on the impact of child nutrition on progression rates through school and total schooling in (3) and (4) indicate significant effects but do not yield parameters that are useful for our simulations because they do not correct for censoring for completed schooling, so we consider illustrative magnitudes for these possible effects.

Appendix E: Supplementary Tables

The tables included in this appendix would be made available only on request and are not intended for publication with the paper.

Table E.1 shows coefficient estimates from the cross-sectional regression specification that is described in section 4.3. The program impact estimates based on this specification and sample are plotted in Figure 5 and the results are summarized in section 4.3. The second-to-last line of the Table shows p-values from an F-test of the joint significance of the family background variables, which strongly rejects the null that family background variables are unimportant in determining outcomes at conventional significance levels. Table E.3 shows the coefficient estimates for a difference-in-difference specification that uses in estimation only the subset of program participants and group B children that are observed at two different time periods. Impact estimates are plotted in Figure 6.

Table E.3-E.4 show the impact estimates based on the cumulative effect matching estimator that uses the group A as the comparison group (group B results are shown in the text). The matches are based on imputed probabilities of program participation to each child in the group A sample based on their characteristics and on the estimates from the logit model described in the text.