EVALUATING PRESCHOOL PROGRAMS WHEN LENGTH OF EXPOSURE TO THE PROGRAM VARIES: A NONPARAMETRIC APPROACH

Jere R. Behrman, Yingmei Cheng, and Petra E. Todd*

Abstract—Nonexperimental data are used to evaluate impacts of a Bolivian preschool program on cognitive, psychosocial, and anthropometric outcomes. Impacts are shown to be highly dependent on age and exposure duration. To minimize the effect of distributional assumptions, program impacts are estimated as nonparametric functions of age and duration. A generalized matching estimator is developed and used to control for nonrandom selectivity into the program and into exposure durations. Comparisons with three groups—children in the feeder area not in the program, children in the program for ≤1 month, and children living in similar areas without the program—indicate that estimates are robust for significant positive effects of the program on cognitive and psychosocial outcomes with ≥7 months' exposure, although the age patterns of effects differ slightly by comparison group.

I. Introduction

There is growing recognition that human capital investments made in early childhood are important determinants of school performance and lifetime productivity.[1] Previous studies suggest strong associations between (1) cognitive and psychosocial skills measured at young ages and (2) educational attainment, earnings, and employment outcomes.[2]

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. Lower levels than in developed countries reflect the facts that children enroll later in elementary school, repeat grades more frequently, and drop out of school at earlier ages. 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, often supported by international agencies, have introduced subsidized preschool programs with the twofold goals of improving child nutrition and providing environments that are conducive to learning (Myers, 1995). In this paper, we evaluate the effectiveness of one such program, an early childhood development program in Bolivia called PIDI (*Proyecto Integral de Desarrollo Infantil*).

There has been little research on preschool interventions in developing-country settings. However, a large literature evaluates the effects of preschool programs in the United States that are targeted at children from impoverished families.[4] The Perry Preschool Program is perhaps the best known of the U.S. programs in the evaluation literature. An experimental evaluation of this program found that children who participated in it scored higher on cognitive tests, although the gains tended to disappear within a few years. Long-lasting effects were found on other outcome measures, such as educational attainment, earnings, welfare participation rates, out-of-wedlock birth rates, and crime rates.[5] Evaluations of two other early intervention programs, the Milwaukee and Abecedarian projects, document long-lasting effects on test scores (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 (for example, Bloom et al., 1993).

Although 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. experience 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. 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. Such differences in program

Received for publication November 19, 2002. Revision accepted for publication September 4, 2003.

* University of Pennsylvania; Florida State University; and University of Pennsylvania and NBER, respectively.

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, Alfonso Flores-Lagunes, Judith McGuire, John Newman, Steven Stern, Edward Vytlacil, and participants at seminars at the University of Minnesota, Lehigh University, University of Delaware, University of Virginia, University of Pennsylvania, Hebrew University, Ohio State 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. We thank an anonymous referee and the editor Robert Moffit for many useful suggestions. Todd thanks the NSF for support under SBR-9730688.

[1] This 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.*

[2] See for example Currie and Thomas (1999), Neal and Johnson (1996).

[3] See Glewwe and Jacoby (1995), Alderman et al. (2001), Glewwe, Jacoby, and King (2001), and Martorell (1999) for evidence for Ghana, Pakistan, the Philippines, and Guatemala.

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

[5] The Perry Preschool Program spent significantly more per pupil than is typically spent on preschool interventions ($7252/year, over a third more than the Head Start program, for example). Most of this expenditure went to teacher pay; the teachers tended to be highly trained professionals (Sweinhart & Weikart, 1998).

The Review of Economics and Statistics, February 2004, 86(1): 108–132

characteristics and the contexts in which they operate could affect the extent and type of benefit from the intervention.

The PIDI program analyzed in this paper provides day-care, nutritional, and educational services to children between the ages of 6 months and 72 months who live in poor, predominantly urban areas. The goals are to improve health and early cognitive/social development by providing children with better nutrition, adequate supervision, and stimulating environments. 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 postschool 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, with additional staff provided when there is a larger proportion of infants. 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 program cost has been estimated by Ruiz (1996) to be approximately $43 per beneficiary per month, which is substantial in a country where per capita annual GDP is $800 in exchange-rate-converted pesos, or $2540 in purchasing power parity terms. Approximately 40% of the expenditure goes to the nutritional component of the program (World Bank, 1997).

This paper uses a large nonexperimental data set 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: height for age and weight for age. To measure cognitive and psychosocial development, we use children's scores on a battery of tests of bulk motor skills, fine motor skills, language and auditory skills, and psychosocial skills.

For our study of the PIDI program, the sample size is approximately 10 times larger than the sizes typically observed in experimental evaluations, and the data set is representative of the entire population of program recipients. However, there is self-selection among eligible children into the program, which poses a threat to the validity of the results. Although the comparison group data sets that we use were chosen by a sampling scheme designed to increase comparability with the families in the program, we still find some important differences between the treatment and comparison group families. For example, families with children in the program tend to have lower parental education levels and incomes, a difference that would likely bias the estimated program impacts downward if not taken into account. This source of 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 and family background differences in analyzing the effects of the program.

We use matching methods to control for potential bias due to nonrandom selectivity into the program. One methodological contribution this paper makes to the previous literature on matching is to allow for a continuous dose of treatment (corresponding to the number of months spent in the program), whereas most of the existing literature assumes that treatment is binary or belongs to a discrete set of treatment types. Two of the matching estimators that we use are justified under the assumption that selection into the program is *on observables*, that is, that it can be taken into account by conditioning on observed family and child characteristics. We also develop an alternative *marginal* matching estimator that 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. An advantage of the marginal estimator is that it only requires data on the treatment group and is thus implementable when no comparison group data are available.

The results show that the program significantly increases cognitive achievement and psychosocial test scores, especially for children who participated in the program for at least 7 months. The impact estimates are fairly robust to the use of alternative comparison groups and estimators. Estimates obtained by the marginal matching estimator tend to be larger, particularly at longer durations and for children aged 6–36 months, than those obtained using traditional econometric estimators that impose stronger functional-form assumptions. Cost-benefit analysis based on our estimates and on evidence from wage studies for developing countries indicates that the PIDI program may have fairly high rates of return.

In section II of the paper, we develop a model of enrollment in preschool that gives an economic interpretation for the average treatment effects that we estimate. Section III describes how we generalize existing matching estimators to accommodate a varying treatment dose as well as impact heterogeneity with respect to children's ages, and introduces the marginal effect estimators. Section IV provides additional information about the PIDI program and data sets, analyzes the determinants of program participation, and presents the impact estimates obtained by matching and, for comparison, by more standard regression methods. Section V performs a cost-benefit analysis based on our preferred marginal effect estimates and other explicit assumptions regarding subsequent schooling and wage effects. Section VI concludes.

II. A Model of the Preschool Participation Decision and Treatment Effects

We develop a model of the mother's decision to enroll her child in preschool, which provides a way of interpreting the treatment effects that will be estimated later in the paper and gives some insight into which conditioning variables should be used in the matching procedure. Our framework assumes that the mother maximizes a time-separable utility function that depends on her own consumption (C_t^m) and leisure (s_t^l) and on the quality of her child (q_t). There is a child quality production function that depends on the mother's time allocated to child production (s_t^c), on the child's consumption of household monetary resources (C_t^c), on whether the child is in preschool, and on stochastic elements. Time not spent in leisure or child quality production is assumed to be spent working at wage w_t^m.

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, and that there is only one child of age a_t for whom the mother is making decisions. $D^*_{p,t}$ is an indicator that takes a value 1 if the mother would choose to enroll the child in the preschool program were the child eligible, e_t is an indicator that equals 1 if the child is eligible. $D_{p,t}$ is an indicator for whether the child is actually enrolled. We assume a fixed cost K to the mother of enrolling in the preschool program and transporting the child to the program site.

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, how much time to invest in the child, how much time to spend in leisure, and how much consumption to allocate to the child. The set of period t state variables, denoted by Ω_t, consists of the child's age, previous-period child quality, mother's wage, father's income, and program participation cost.[6] The random shocks in the model (ε_t^t, ε_t^l, ε_t^q) are shocks to the value of mother's consumption and leisure and to the child quality production function. The mother solves

$$V_t(\Omega_t) = \max_{\{C_t^m, s_t^c, D^*_{P,t}, s_t^l\}} U(C_t^m, q_t, s_t^l; \varepsilon_t^t, \varepsilon_t^l, \varepsilon_t^t) + \beta E(V_{t+1}(\Omega_{t+1} | C_t^m, s_t^c, D^*_{P,t}, s_t^l, \Omega_t))$$

subject to constraints

$$q_t = q(s_t^c, C_t^c, D_{P,t}, a_t, q_{t-1}) + \varepsilon_t^q, \tag{1}$$

$$C_t^m + C_t^c + D_{p,t}K \leq (1 - s_t^c - s_t^l)w_t^m + A_t, \tag{2}$$

$$D_{P,t} = e_t D^*_{P,t}. \tag{3}$$

Equation (1) describes the production technology for child quality, where we are assuming that previous-period quality is a sufficient statistic for prior inputs,[7] that ε_t^q represents a shock realized after input decisions, and that the productivity of inputs may depend 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) is the budget constraint, and (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 changing their labor force participation or by making their children 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 $D_{P,t} = 0$ to $D_{P,t} = 1$ for $t \in \{t, t'\}$) can be expressed in terms of the model. It includes 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. Let $s_{1,t}^c$, $s_{1,t}^l$, $C_{1,t}^c$ denote the values that solve the dynamic programming problem when $D_{P,t} = 1$, and let $s_{0,t}^c$, $s_{0,t}^l$, $C_{0,t}^c$ denote the corresponding values when $D_{P,t} = 0$.

The total effect of preschool participation on current-period quality for a particular child of age a_t who starts off at quality level $q_{t-1} = \bar{q}$ is given by

$$\frac{\Delta q_t}{\Delta D_{P,t}} = q(s_{1,t}^c, C_{1,t}^c, 1, a_t, \bar{q}) - q(s_{0,t}^c, C_{0,t}^c, 0, a_t, \bar{q}).$$

The total program effect is inclusive of any compensating changes in the mother's allocation of time and consumption to the child. In addition, a change in current-period quality levels potentially affects future quality levels. For example, if a child starts off period $t + 1$ at a high quality level, he or 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

$$\frac{\Delta q_{t'}}{\Delta q_{t'-1}} \frac{\Delta q_{t'-1}}{\Delta q_{t'-2}} \cdots \frac{\Delta q_{t+1}}{\Delta q_t} \frac{\Delta q_t}{\Delta D_{p,t}}.$$

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

$$\Delta_{t,T} = \sum_{v=t}^{t'} \left\{ \frac{\Delta q_v}{\Delta D_{P,v}} + \sum_{w=v+1}^{T} \frac{\Delta q_w}{\Delta q_{w-1}} \cdots \frac{\Delta q_{v+1}}{\Delta q_v} \frac{\Delta q_v}{\Delta D_{P,v}} \right\}, \tag{4}$$

where the first term captures the current-period impact and the second term the impact on future quality levels up until some end period T.[8]

In the data analyzed in sections IV and V, we do not observe children over the entire time period of their

[6] For simplicity, there is no uncertainty about the wage, father's income, or participation cost.

[7] We make this assumption for simplicity. The assumption that previous-period quality is a sufficient statistic could be relaxed to allow quality to depend on the history of inputs over the child's lifetime.

[8] In 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.

participation in the program (up to T), so we can only estimate the cumulative effect of the preschool treatment up until the time of observation, t^o. We assume that the empirical test scores and anthropometric measures available in the data set capture aspects of child quality.[9] The mean treatment effect we estimate (conditional on age and duration of time in the program) using the matching estimators described in section III is equal to $E(\Delta_{t,t^o}|\text{age} = a,\ t^o - t = l,\ D_P = 1)$, where $D_P = 1$ denotes participating in the program. The treatment effect 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 families.[10]

Next, we consider the question of how to choose the set of conditioning variables used in matching. Let q_{0t} be the quality level when the child does not participate in the program at time t, and q_{1t} the quality level when the child does participate. The decision to enroll the child at any time period (which is only relevant for eligible families) implies that at that date, the current utility plus the expected future utility from participating is higher than from not participating:

$$\begin{aligned} &U(C^m_{0,t},\ q_{0,t},\ s^l_{0,t};\ \varepsilon^l_t,\ \varepsilon^t_t) \\ &\quad + \beta E(V_{t+1}(\Omega_{t+1}|C^m_{0,t},\ s^c_{0,t},\ s^l_{0,t},\ D^*_{P,t},\ \Omega_t)) \\ &\quad < U(C^m_{1,t},\ q_{1,t},\ s^l_{1,t};\ \varepsilon^l_t,\ \varepsilon^t_t) \\ &\quad + \beta E(V_{t+1}(\Omega_{t+1}|C^m_{1,t},\ s^c_{1,t},\ s^l_{1,t},\ D^*_{P,t},\ \Omega_t)). \end{aligned}$$

Define the outcomes in the no-program state and in the program participation state as

$$Y_t(a,\ 0) = q_t|_{D_{P,t}=0 \text{ for all } t' \leq t} \quad \text{and}$$
$$Y_t(a,\ l) = Y_t(a,\ 0) + \Delta_{t,t+l},$$

and suppose that there is available a set of conditioning variables $\tilde{Z}_t$. The cumulative matching estimator described in the next section assumes that

$$\begin{aligned} &E(Y_t(a,\ 0)|\tilde{Z}_t,\ D^*_{P,t} = 1,\ t \in t_0 \cdots t) \\ &\quad = E(Y_t(a,\ 0)|\tilde{Z}_t,\ e_t = 1,\ D^*_{P,t} = 0,\ t \in t_0 \cdots t). \end{aligned}$$

The estimator also requires that $\Pr(D^*_{P,t} = 0|\tilde{Z}_t,\ e_t = 1) > 0$, so that there is a positive probability of observing both program participants and nonparticipants with the same characteristics $\tilde{Z}_t$. This requirement implies that, conditional on $\tilde{Z}_t$, there must be other variables affecting program participation and that these variables not be correlated with child outcomes in the no-treatment state. For example, suppose that distance to the program site is a determinant of participation and that the placement of program sites can be considered random with respect to child outcomes in the no-treatment state conditional on $\tilde{Z}_t$. If $\tilde{Z}_t$ contains all the elements of the state space except distance, then the above exogeneity condition can be satisfied.

For reasons described later in the paper, it is important not to match on variables that themselves are affected by the program, such as the mother's labor supply in the model. This is because the matching estimator (described below in section III) integrates over $f(\tilde{Z}_t|D^*_{P,t} = 1)$. To estimate correctly the mean no-treatment outcomes, we require that the density of the matching variables do not change with treatment. For this reason, we match on the following: (a) variables observed prior to the enrollment decision (under the assumption that the density of these variables does not change due to anticipation of the program), (b) variables that we expect to be stable over the time period of observation (such as the mother's and father's education, the family structure, and the characteristics of the household), and (c) variables that are deterministic with respect to time (such as the child's age). We do not include variables that directly relate to children's physical, mental, and social development.

III. Cumulative and Marginal Matching Estimators

As discussed in the previous section, we are interested in estimating the treatment effect Δ_{t,t^0} [defined in equation (4)], which gives the total effect of the preschool program on child quality for a child that participates in the program for a duration $l = t - t^0$. We next describe the estimators we use and the assumptions required to justify their application. We go beyond the previous literature on matching by allowing for a continuous dose of treatment (given by the duration of time spent in the program), by permitting impacts to depend in a flexible way on children's ages, and by developing a marginal matching estimator that can be implemented if data on program participants are the only data available.

Let $Y(a,\ l)$ denote the outcome measure intended to capture an aspect of child quality (test score or anthropometric measure) for a child of age a who participated in the program for length of time l.[11] For nonparticipants, $l = 0$. Also define $D_P = 1$ if $l > 0$, $D_P = 0$ otherwise. For a child of age a, the cumulative impact of participating in the program l time periods relative to not participating is given by $\Delta(a,\ l,\ 0) = Y(a,\ l) - Y(a,\ 0)$. Also of interest is the marginal effect of participating in the program l_1 time

[9] Preschool 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 V, we will briefly consider these other sources of benefits.

[10] Knowing the treatment effect of the program does not allow recovery of the parameters of the production technology. Only under the strong assumption that parents do not alter their time or resource allocations when their child participates in the program would the treatment effect correspond to a feature of the production technology (see related discussion in Todd & Wolpin, 2003).

[11] We assume participation takes place in consecutive time periods, as it does in our data.

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 impacts is 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 $f_a(a|l = l_1)$ is the conditional density of ages and A can be a singleton set or a range of ages.

Integrating over the joint density of observed ages and program durations gives the overall impact of the program relative to the counterfactual of participating for length of time equal to l_0:

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

where $L = \{l : l > 0\}$. Thus $\bar{\Delta}(A, L, 0)$ 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.* A comparison of a monetary valuation of this impact with average program costs is informative on whether the program has a positive net benefit (see section V).

A. Estimators

We now give the identifying conditions that justify the method of matching as a way of imputing the missing data required to estimate the average treatment effects defined above, which relate to the treatment impacts defined in terms of the economic model of section II.[12]

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 with that of the treated group.[13]

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

$$E(Y(a, 0)|a, l_i = l, x) = E(Y(a, 0)|a, l_i = 0, x) \tag{5}$$

and

$$0 < f(a, x|D_P = 0). \tag{6}$$

As discussed in section II, the first condition implies that after conditioning on a set of observed characteristics $\{a, x\}$, no-treatment outcomes for children who have participated for duration l will be on average the same as those observed for children who have not participated ($D_P = 0$). The second condition ensures that for each child in the participant group there is positive probability of finding a match from the nonparticipant group.[14] Let $S_P = \{(a, x) : f(a, x|D_P = 1) > 0$ and $f(a, x|D_P = 0) > 0\}$ denote the region of the support of (a, x) that satisfies equation (6), called the *region of overlapping support.*[15]

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

$$\hat{\bar{\Delta}}(A, L, 0) = \frac{1}{n} \sum_{\substack{i\in\{D_P=1\}\cap\{a_i\in A\}\\ \cap\{l_i\in L\}\cap\{(a_i,l_i)\in S_P\}}} \{\hat{E}(Y(a_i, l_i)|x_i, D_{P_i} = 1) - \hat{E}(Y(a_i, 0)|x_i, D_{P_i} = 0)\}, \tag{7}$$

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.

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

$$\sum_{k\in\{D_P=0\}}^{n_0} Y_k(a_k, 0) W_k(\|a_k - a_i\|, \|X_k - X_i\|), \tag{8}$$

where $W_k(\|a_k - a_i\|, \|x_k - x_i\|)$ are weights that sum to 1 and that depend on the Euclidean distance between (a_i, x_i) and (a_k, x_k). For standard kernel weighting functions,

[12] Matching methods have been developed and applied to the evaluation of training programs by Heckman, Ichimura, and Todd (1997), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), Smith and Todd (2001, 2004), and others.

[13] In performing this alignment, it emulates a feature of a randomized experiment in which the characteristics of the treatment and comparison groups are aligned by virtue of randomization.

[14] See Rosenbaum and Rubin (1983). 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.

[15] See Heckman, Ichimura, and Todd (1997) for discussion of support conditions required in matching.

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) and Fan (1992). The details of local linear regression estimators are described in appendix B.[16]

The analogous nonparametric estimator for $\hat{E}(Y(a_i, l_i)|x_i, D_{Pi} = 1)$ in equation (7) is

$$\hat{E}(Y(a, l)) = \sum_{k \in \{D_P = 1\}} Y_k(a_k, l_k) W_k(\|l_k - l_i\|, \|a_k - a_i\|, \|x_k - x_i\|), \tag{9}$$

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).[17] Note that in equation (7) 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 of Heckman, Ichimura, and Todd (1998) is general enough to accommodate the estimators. In the empirical work, however, we evaluate the variation of the estimators using bootstrap methods rather than variance estimators based on asymptotic formulas.

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_0 to l_1: $\bar{\Delta}(a, l_1, l_0) = E(Y(a, l_1)|D_P = 1, x) - E(Y(a, l_0)|D_P = 1, x)$, where $l_1, l_0 > 0$. There are two different ways of estimating marginal effects. One is to first estimate cumulative effects at different duration levels and then take the difference. The other way is to use only data on program participants, drawing comparisons between program participants who have taken part in the program for different lengths of time.

1. *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: $\bar{\Delta}(a, l_1, l_0) = \hat{\Delta}(a, l_1, 0) - \hat{\Delta}(a, l_0, 0)$. 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 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.
2. *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 the previous one 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 potential source of 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—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.

In our empirical application to the analysis of the PIDI program, a major determinant of duration in the program is the time at which the program first became available to children, which differs across children depending on the child's place of residence and child's age at the time the local PIDI site began its operation. Two-thirds of the children in the PIDI evaluation sample began participating in the program as soon as it became available; on average, the delay between the time the local PIDI site opens and the time children begin participating is 3.2 months. The variation in duration of time spent in the program that arises from variation in when the program became available to children is therefore arguably exogenous with respect to program outcomes, conditional on observed child and parent characteristics.

Formally, the identifying assumption the marginal estimator invokes is that there exists a set of conditioning variables x such that

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

and $0 < f(a, x|l = l_0)$. Under these assumptions, an estimator for $\bar{\Delta}(a, l_1, l_0)$ is given by

[16] The numbers 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 described in section IV.

[17] An alternative approach would be to construct the weighted averages in equation (8) 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.

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

The conditional expectations are estimated by the same local regression method as described in relation to equation (9).

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

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 strong conditional independence assumptions, we could apply the above 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 (see Heckman, Ichimura, & Todd, 1997). Instead, we assume directly that equation (5) 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:

$$\begin{aligned} \hat{E}(Y(a, l)|P(x), D_P = 1) &= \sum_{k \in \{D_P = 1\}} Y_k(a_k, l_k) \\ &\quad \times W_k(\|l_k - l\|, \|a_k - a\|, \|P(x_k) - P(x)\|), \end{aligned} \tag{10}$$

$$\begin{aligned} \hat{E}(Y(a, 0)|P(x), D_P = 0) &= \sum_{k \in \{D_P = 0\}} Y_k(a_k, 0) \\ &\quad \times W_k(\|\|a_k - a\|, \|P(x_k) - P(x)\|). \end{aligned}$$

In our empirical work, we estimate the conditional probabilities $P(x)$ by logistic regression.[18]

Modifications Required to Accommodate 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 and weights are required to consistently estimate the probabilities of program participation. 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 with unknown weights. They show that the odds ratio $P(x)/[1 - P(x)]$ that is estimated using the wrong weights (ignoring the fact that the data are 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.

Comparison with Regression-based 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. Matching methods aim to increase the comparability between the treatment and comparison group samples by aligning the distribution of observed covariates of comparison group members with 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

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

where the second equality follows under the assumptions that would justify the application of matching. The last line

[18] Heckman, 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, Ichimura, and Todd (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.

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 accomplished through matching (or through a weighted regression) balances observed characteristics of the treatment and comparison groups. Such a balancing would also occur in a randomized experiment. Traditional regression-based estimators do not attempt to emulate the balancing feature of randomized experiments, but instead control for observable differences between groups by assuming the conditional mean function is correctly specified by the regression equation.[19]

Selection on Unobservables: The estimators for cumulative program effects described above assume that outcomes are mean-independent of program participation conditional on a set of observables. If the program participation equation can be described by the index model $D_P = 1(\psi(Z) - V > 0)$, then 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 if unobservables that are related to program outcomes are important determinants 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, Ichimura, and Todd, 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 (baseline) data is a 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 versus long durations using the estimators, described earlier in this subsection under "Estimating Marginal Program Impacts," that allow selection into the program to be based on unobservables.

IV. Empirical Results

A. The Data

The PIDI evaluation data sets consist of repeated cross-section data collected in two rounds on three different subsamples: (i) a participating subsample (P) of children selected randomly from children in the PIDI program, (ii) 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 within a 3-block radius of a PIDI site but without any children attending PIDI, and (iii) 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 at the time of the survey.[20] As noted above in section IIIA under "Reducing the Dimension of the Conditioning Problem," the data are choice-based sampled with unknown population weights. For this reason, we do not know the participation rate among all eligibles. Fortunately, this information is not needed for our estimation strategy, but it would be required to implement some other common evaluation approaches.[21]

We estimate program impacts using both the comparison group samples A and B. Sample B has an advantage over A in being drawn from the same area as the participant sample P, which controls for unobserved 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 with those for P 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, we compare children in the participating sample P who had been in PIDI for two or more months with children in P who had been in PIDI for one month or less.

All the children in sample P meet the eligibility criteria that are summarized below in section IVB, but children in the comparison samples A and B do not necessarily meet the criteria. In our application of the matching estimators, we only use subsamples of children from the samples A and B who satisfy the eligibility criteria.[22] As described below, there was a change in the eligibility criteria over time. We use the later criteria rather than the earlier ones, because the earlier ones included subjective aspects, the application of which we cannot duplicate with much confidence. 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

[19] Another difference between matching and regression estimators is how they deal with the problem of nonoverlapping 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, regression estimators typically use all the observations in estimation and use functional-form assumptions to extrapolate over any regions of x where the supports do not overlap.

[20] Stratification is based on information given in the 1992 Bolivian Census.

[21] For example, one alternative approach would be to compare sites that received the program with sites where the program was unavailable, viewing program placement as exogenous. The matching strategy we adopt takes into account differences in family and locality characteristics and therefore does not assume that program placement is exogenous. For a discussion of the implication of endogenous program placement, see Rosenzweig and Wolpin (1986) and Pitt, Rosenzweig, and Gibbons (1993).

[22] Equivalently, one could refrain from imposing 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.

TABLE 1.—SAMPLE SIZES OF GROUPS P, A, AND B: ROUNDS 1 AND 2

Round	Participant Sample (P)	Comparison Sample (A): Without Imposing Eligibility	Comparison Sample (A): Imposing Eligibility	Comparison Sample (B): Without Imposing Eligibility	Comparison Sample (B): Imposing Eligibility	Participants with Duration*: =0, 1 mo.	Participants with Duration*: ≥2 mo.
1	1198	1227	558	628	333	472	708
2	2420	2205	987	1732	963	237	1252
Both	364	745	415	392	247	0	268

* The numbers in the last two columns do not sum to the number in the first column, because some observations are missing duration data.

in P, we cannot infer their preprogram nutritional status and, in particular, whether 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.

Table 1 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 A children, and 628 B children interviewed between November 1995 and May 1996.[23] The second round consists of a follow-up sample from the first round and, in addition, a larger sample of new households that were not visited in the first round. The second round includes 2420 participant children, 2205 group A children, and 1732 group B children who were interviewed between November 1997 and May 1998. Imposing the eligibility criteria on the comparison group samples leads to a substantial reduction in the sample sizes—roughly cutting them in half. The numbers of children observed in both rounds are 364 participants in group P, and 745 group A and 392 group B children.[24]

B. Eligibility Criteria

To participate in PIDI, families are required to meet eligibility criteria.[25] The initial eligibility requirements 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 supposed to be applied lexicographically and were in part subjective (particularly the first and third), which introduces a random element in who participates in the program, even after conditioning on observed characteristics. The initial criteria subsequently were replaced by a more objective eligibility index that awards one point if 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) a mother with five grades or less of schooling, and (h) an unemployed father. Two points are awarded if (a) the family has only a mother or a father or (b) the mother of the family works 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—such as 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 child care.[26]

C. Variables

The PIDI data sets provide detailed information on parental, household, and child characteristics. There is information, for example, on income sources, educational attainment, parental occupations, fertility and reproductive histories, family structure, and possession of durable goods. For all children in the sample households between 6 and 72 months of age, there are data on cognitive, psychosocial, and anthropometric test score measures. The outcome measures that we examine in this paper are the following: (i) bulk motor skills, (ii) fine motor skills, (iii) language-auditory skills, (iv) psychosocial skills, (v) height-for-age percentile, and (vi) weight-for-age percentile.[27]

[23] The sampling frame was a stratified random sample. First PIDI sites were randomly sampled, and then children within the sites were selected randomly.

[24] In the first round, there are 1198 children in PIDI. Because of difficulties in relocating some of the families, only 739 of these children were followed up in the second round. Of these children, 364 were still participating in PIDI at the time of the second-round data collection, 268 were too old for PIDI (had graduated from the program), and 104 were no longer participating in the program. Thus, we estimate the program dropout rate among the children who were followed in the second round to be approximately 23%.

[25] Once they were determined to be eligible, they could not become ineligible for the program even if some of their characteristics changed over time.

[26] Because of the 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.

[27] We also explore whether these are effects on the lower tails of the anthropometric distributions—explicitly, on a height Z score below a threshold of 3 and a weight Z score below a threshold of 2, where the different thresholds reflect the relative severity of the nutritional problems in this population. (Z scores give the number of standard deviations from the mean. They are widely used in the nutrition literature to characterize

The first three are measures of cognitive skills, the fourth is a psychosocial outcome, and the last two are anthropometric measures.[28] The test score outcomes (i), (ii), (iii), and (iv) are highly significantly 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 less strongly 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 the weight-for-age percentile and the test score outcomes are all insignificantly different from 0.

D. *Comparison of Group Mean Characteristics*

In table 2, 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 1 month or less and 2 months or more.

Panel A of the table compares characteristics of the mothers. Approximately 8% of mothers in the PIDI group have no education and cannot read or write, which is similar to the rate for mothers in the *B* sample (eligible and total) but slightly higher than for the *A* sample (eligible and total).[29] PIDI mothers are also more likely to participate in the labor force, but much of this difference is eliminated by imposing program eligibility criteria on samples *A* and *B*. A comparison of the incomes shows that PIDI mothers have lower incomes even though they work on average more hours per day. Among PIDI mothers in the two duration subsamples (shown in the last two columns) there are no significant differences.

Panel B compares characteristics for the fathers. Fathers' educational 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%). Fathers in the PIDI group have less stable employment and are more likely to be employed in occasional work than are fathers in the other groups; but if the eligibility criteria are applied, there is a reversal in these comparisons. Imposing eligibility also makes the average income levels similar across groups. Within the *P* group there are no significant differences for the two subsamples defined by program duration (last two columns).

Panel C compares other characteristics of the household and reveals differences in family structure across groups: PIDI households are less likely to have both parents residing in the household, and they have lower total household and per capita income. Group differences are reduced substantially when the eligibility criteria are imposed.

In summary, in terms of the observed mothers', fathers', and other household characteristics, the total *A* and *B* samples tend to be economically better off than the *P* sample.[30] Applying the eligibility criteria makes the comparison samples based 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 1 month and less versus 2 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.

Child Characteristics: A comparison of the age distribution for PIDI participating children with children in groups *A* and *B* reveals major differences, with PIDI participants tending to be much more concentrated in middle age ranges (30–55 months). Figure 1 compares children in the eligible *B*, eligible *A*, and *P* groups with respect to weight, height, and four test-score outcome measures, conditioning only on age. From the figure, it is apparent that PIDI children older than 12 months are short for their age. 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 2 plots the outcome measures for group *A* and group *B* eligible children and for *P* children who have participated at least 13 months. The PIDI group appears to do better on average in cognitive test scores than children from the other groups, but this difference is not necessarily attributable to the program; it may be due to preexisting

---

the degree of malnutrition, with a Z score < -2 indicating moderate malnutrition and < -3 indicating severe malnutrition.) Z scores are increasingly used in the economic literature on the determinants of and impact of malnutrition (see the survey in Strauss & Thomas, 1998). We report these estimates in the text, but for the sake of brevity do not present them in tables.

[28] The cognitive outcomes and psychosocial 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.

[29] 44% of PIDI mothers have only basic or no education, compared with 34% of group *A* and 46% of group *B* mothers (47% and 61% if *A* and *B* are limited to those eligible).

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

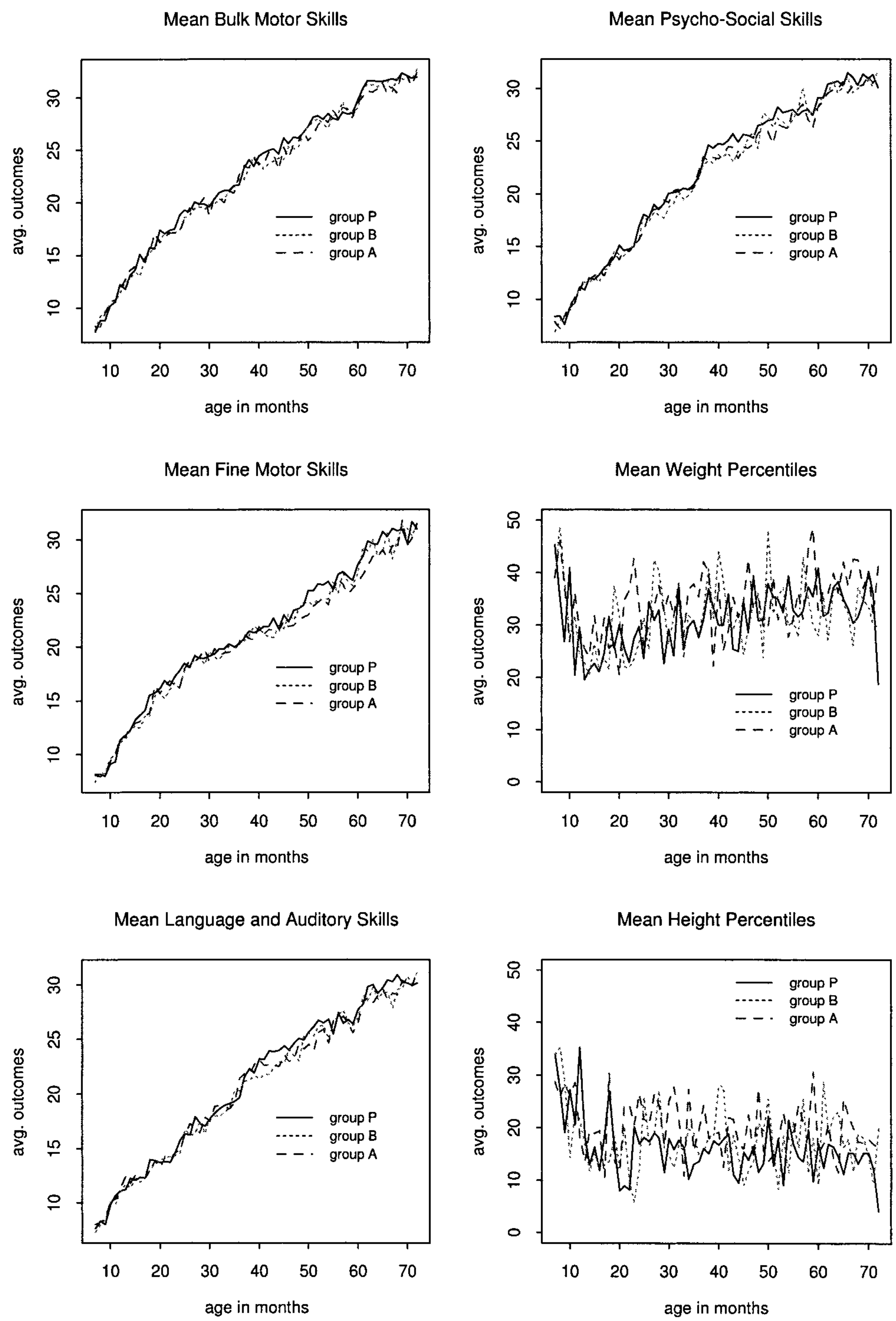
TABLE 2.—COMPARISON OF GROUP MEAN CHARACTERISTICS

Characteristic	Participants' Sample (P)	Eligible Nonpart. Sample		Eligibility Not Imposed		Participants with Duration	
		A	B	A	B	>1 mo.	≤ 1 mo.
A. Comparison of Mothers' Characteristics in Participant and Eligible Nonparticipant Samples*							
Education:							
None	8.48	5.49	8.17	0.88	1.60	8.37	8.81
Preschool	0.04	0.00	0.00	0.04	0.25	0	0.15
Basic	35.53	41.45	52.99	15.45	24.37	35.74	34.93
Middle school	21.48	24.25	20.13	18.89	21.08	21.11	22.54
Secondary	28.40	22.23	14.15	42.89	37.44	29.63	24.93
Normal school	1.56	1.83	2.20	2.34	3.88	1.42	1.94
University	1.98	2.56	1.57	13.55	6.66	1.68	2.84
Technical	2.53	2.10	0.79	4.28	3.20	2.05	3.88
Other or no ans.	0.00	0.09	0.00	1.68	1.52	0.00	0.00
Age:							
Mean	29.04	30.28	30.97	33.17	33.84	29.16	28.71
(s.d.)	(6.34)	(6.92)	(7.34)	(7.80)	(8.86)	(6.43)	(6.07)
Literacy:							
Can read/write	92.02	94.78	91.82	99.34	98.40	92.00	92.09
Cannot read/write	7.98	5.22	8.18	0.66	1.60	8.00	7.91
Type of work:							
Permanent	67.16	59.10	59.91	77.05	75.97	67.47	66.27
Occasional	19.65	22.14	20.44	19.68	11.98	19.95	18.81
No job	13.19	18.76	19.65	3.27	4.05	12.58	14.92
Kind of job:							
Worker	7.04	5.74	7.63	24.95	27.77	6.98	7.19
Clerical	55.45	23.76	19.77	41.01	35.68	57.19	50.35
Self-employed	34.33	61.60	65.95	29.65	33.22	32.81	38.77
Employer	0.09	0.79	0.78	4.01	2.90	0.07	0.35
Family business	2.82	8.11	5.68	0.27	0.44	2.71	3.16
Other	0.27	0.00	0.20	0.09	0.00	0.30	0.18
Hours worked/day:							
Mean	8.70	7.70	7.88	9.19	9.28	8.72	8.63
(s.d.)	(2.62)	(3.18)	(3.29)	(2.43)	(2.51)	(2.56)	(2.79)
Income/month:							
Mean	328.50	443.14	496.20	907.26	858.29	329.78	324.73
(s.d.)	(427.32)	(518.02)	(689.96)	(948.54)	(934.46)	(461.15)	(306.77)
B. Comparison of Fathers' Characteristics in Participant and Eligible Nonparticipant Samples*							
Education:							
None	1.97	1.45	1.94	0.88	1.60	2.14	1.48
Preschool	0.00	1.21	0.39	0.04	0.25	0.00	0.00
Basic	21.91	21.23	33.33	15.45	24.37	21.18	23.99
Middle school	22.44	27.14	25.19	18.89	21.08	22.47	22.32
Secondary	42.81	41.74	31.59	42.89	37.44	43.65	40.41
Normal school	1.77	1.45	2.91	2.34	3.88	1.94	1.29
University	5.13	3.38	2.71	13.55	6.66	4.47	7.01
Technical	3.02	2.29	1.74	4.28	3.20	3.11	2.77
Other or no ans.	0.96	1.33	0.20	1.68	1.52	1.04	0.74
Literacy:							
Can read/write	98.32	99.03	98.06	99.34	98.40	98.12	98.89
Cannot read/write	1.68	0.97	1.94	0.66	1.60	1.88	1.11
Age:							
Mean	32.52	33.78	34.78	33.17	33.84	32.56	32.39
(s.d.)	(7.79)	(8.01)	(8.93)	(7.80)	(8.86)	(7.80)	(7.76)
Type of work:							
Permanent	71.33	72.50	73.06	77.05	75.97	71.05	72.14
Occasional	25.36	23.28	20.35	19.68	11.98	25.91	23.80
No job	3.31	4.22	6.59	3.27	4.05	3.04	4.06
Kind of job:							
Worker	37.48	29.97	28.63	24.95	27.77	36.41	40.58
Clerical	33.52	32.49	27.80	41.01	35.68	34.74	30.00
Self-employed	26.18	33.75	40.25	29.65	33.22	26.25	25.96
Employer	2.58	3.78	2.28	4.01	2.90	2.34	3.27
Family business	0.10	0.00	1.04	0.27	0.44	0.13	0.00
Other	0.15	0.00	0.00	0.09	0.00	0.13	0.19
Hours worked/day:							
Mean	9.51	9.57	9.54	9.19	9.28	9.51	9.52
(s.d.)	(2.35)	(2.47)	(2.62)	(2.43)	(2.51)	(2.36)	(2.31)
Income/month:							
Mean	713.40	745.21	737.95	907.26	858.29	711.82	717.99
(s.d.)	(906.67)	(662.84)	(556.31)	(948.54)	(934.46)	(1011.26)	(490.96)
C. Comparison of Households' Characteristics in Participant and Eligible Nonparticipant Samples*							
Family structure:							
Only father resides in house	1.74	0.27	0.77	0.19	0.37	1.59	2.15
Only mother resides in house	20.02	24.16	19.20	15.30	11.38	19.84	20.52
Neither parent resides in house	1.21	0.81	1.54	0.57	0.96	1.03	1.72
Both parents reside in house	77.04	74.75	78.49	86.13	87.29	77.55	75.61
Household income:							
Mean	902.65	1019.65	1070.33	1189.61	1131.65	900.91	907.87
(s.d.)	(994.37)	(1028.33)	(1025.64)	(1429.10)	(1239.67)	(1083.12)	(685.25)
Per person income:							
Mean	185.18	200.77	201.22	242.00	217.75	183.49	190.25
(s.d.)	(233.10)	(299.72)	(206.34)	(346.42)	(251.58)	(256.71)	(147.67)
Number of persons in household:							
Mean	5.23	5.68	5.75	5.39	5.53	5.23	5.24
(s.d.)	(1.92)	(2.26)	(2.06)	(2.05)	(2.03)	(1.91)	(1.93)
Number of rooms in household:							
Mean	1.61	1.68	1.66	1.95	1.94	1.62	1.56
(s.d.)	(1.08)	(0.99)	(0.93)	(1.27)	(1.27)	(1.07)	(1.09)
Household ownership:							
Own house	35.24	30.23	44.24	30.78	42.35	35.29	35.07
Rented house	33.67	30.68	24.58	27.66	24.39	34.10	32.46
Paid in kind or relative's house	27.69	34.57	28.88	35.47	29.93	27.28	28.84
Other	3.41	4.52	2.31	6.09	3.33	3.33	3.6
Household has a TV							
Mean	0.31	0.37	0.31	0.50	0.42	0.31	0.32
(s.d.)	(0.46)	(0.48)	(0.46)	(0.50)	(0.49)	(0.47)	(0.45)
Participation in outside org.†	58.41	25.15	36.80	27.71	31.89	61.98	48.06

* Includes both rounds of data, but excludes observations from second round who were also included in first round.

† Someone in household participates in neighborhood organizations.

FIGURE 1.—COMPARISON OF GROUP MEAN OUTCOMES

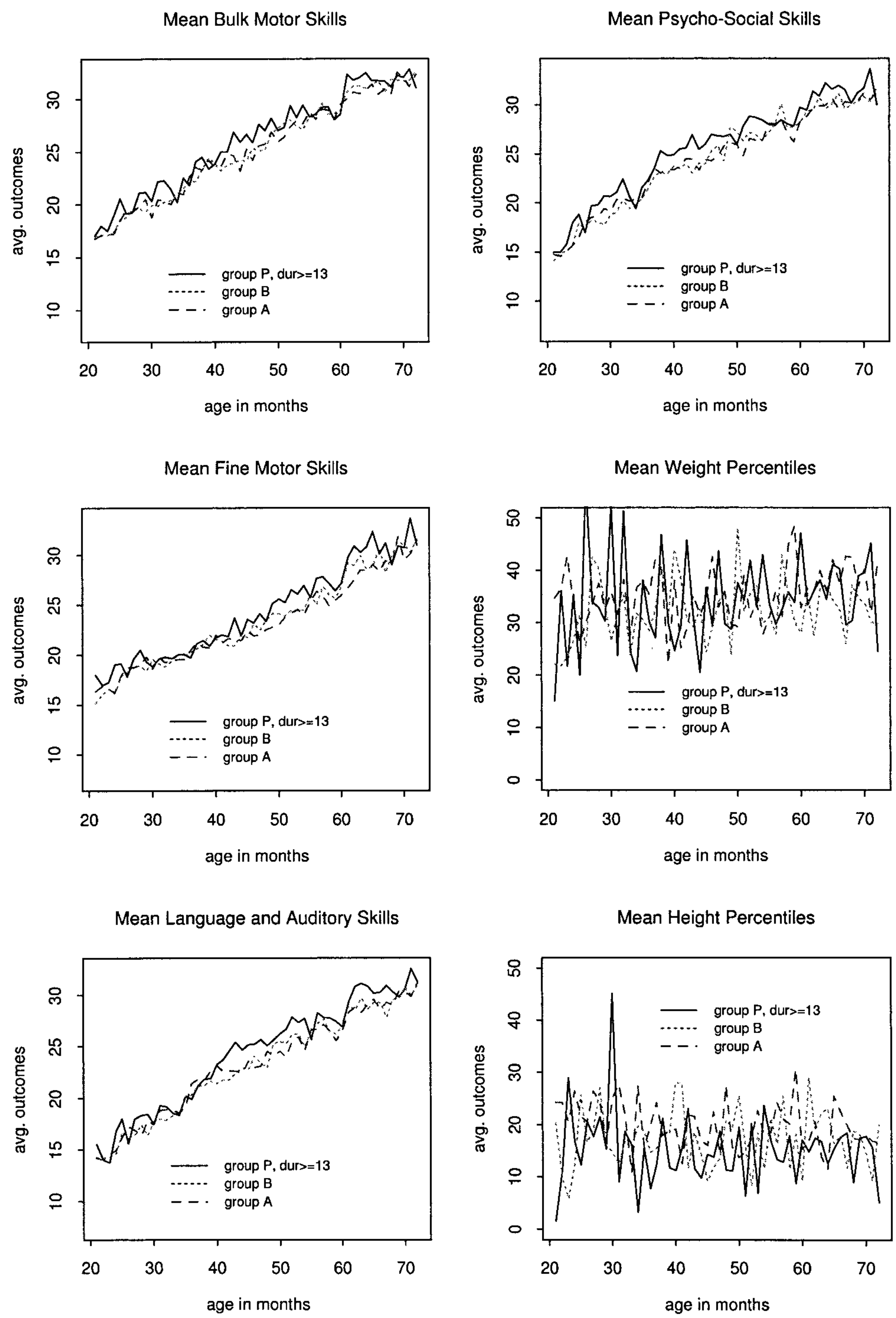


differences between program participants and nonparticipants.

E. *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 IIIA. The mean comparisons in table 2A–C indicate that groups A, B, and P differ along several dimensions that could be relevant to the program participation decision. 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. Our selection of variables is guided by the theoretical model presented in section II that indicated that father's income, child's age and child's pre-

FIGURE 2.—COMPARISON OF GROUP MEAN OUTCOMES, DURATION AT LEAST 13 MONTHS

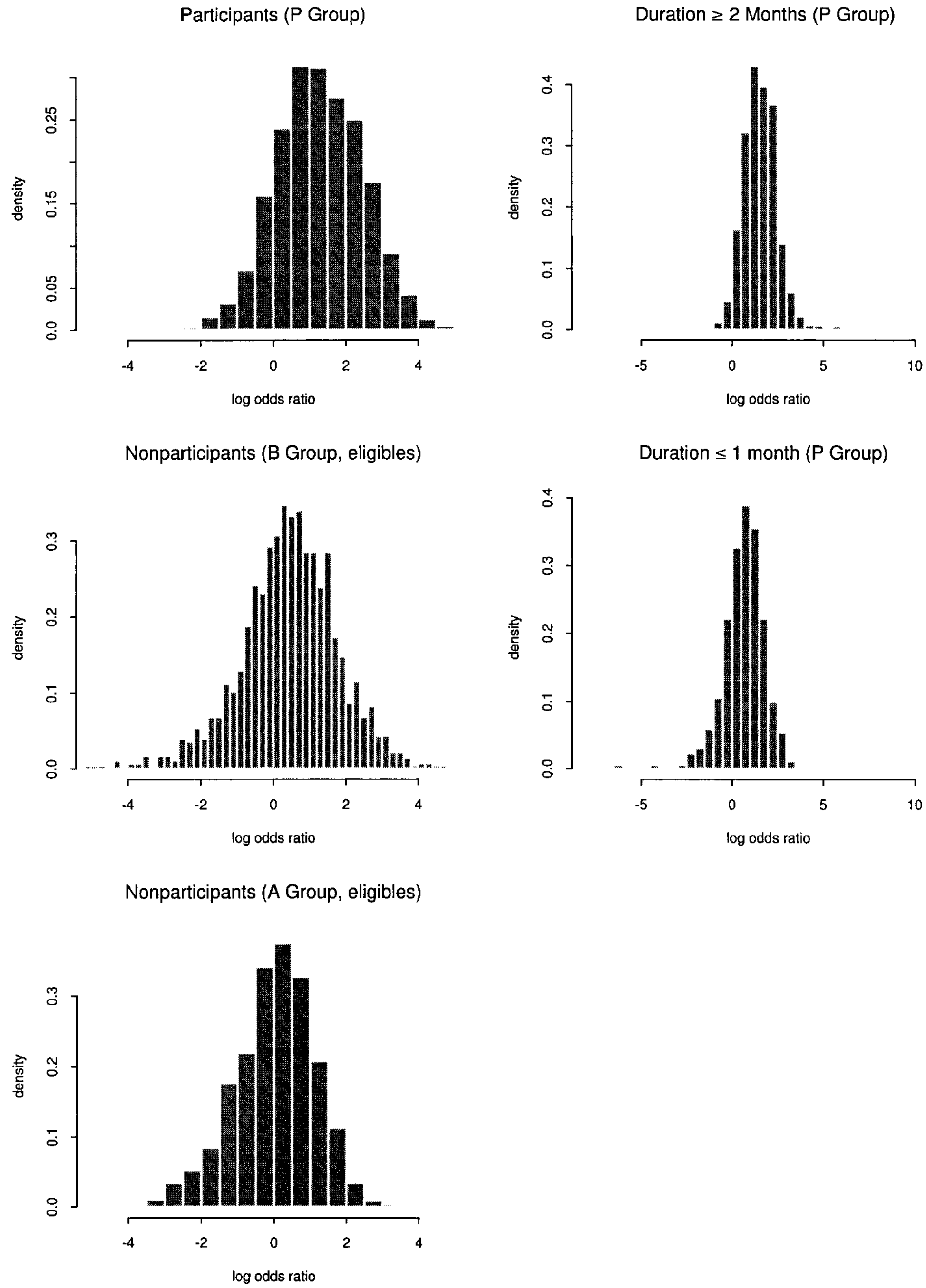


program quality status are potential determinants of participation.[31] Information is available on all these variables, except preprogram quality (see discussion below). We select the particular set of included regressors for the logistic model from those shown in table 2 to maximize the percentage correctly classified by the hit-or-miss criterion. Under the resulting model, 79% of the observations are correctly classified.[32] The included regressors are listed in appendix C. The most useful predictors of participation are

[31] Mother's labor force participation was not included in the participation model out of concern that it might change in response to treatment. The model of section II assumes that the mother's labor force participation is jointly determined with the participation decision.

[32] 77% of the participants and 84% of the controls are correctly classified.

FIGURE 3.—DISTRIBUTION OF ESTIMATED LOG ODDS RATIO



(i) presence of a mother in the household, (ii) education level of the mother, (iii) number of children, (iv) education level of the father, and (v) 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 decisions for group A as for group B, we can impute probabilities of participation for group A families using the coefficients from the participation model that was estimated on groups P and B.[33]

The first column of figure 3 plots the log odds ratio for participating children and for eligible children in the B and A groups. For groups P and B the supports of the log odds ratios overlap, but if group A is used as a comparison group,

[33] This requires assuming that there are no significant unobserved locality characteristics affecting outcomes, so that a similar model for participation for groups P and B can also be applied to group A.

some high values of the log odds ratio are observed for program participants for which no matching values can be found for children in the *A* 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 2 months or more.[34] Appendix C lists of set of regressors we used for this model (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 3.[35]

F. Impacts Estimated by Traditional Regression Methods

Before presenting impact estimates based on the matching estimators, we first report for comparison estimates that are obtained by simple regression estimators. First, we estimate a simple cross-sectional regression model for the three cognitive development tests, the psychosocial ability test, and the two anthropometric indicators, based on the combined *P* and eligible *B* samples. Our specification 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 a set of conditioning variables that is the same as used in estimating the probability of program participation, as described in the previous section.[36] Figure 4 plots the estimated program impact as a function of duration in the program.

The figure shows that estimated program impacts on test scores are mostly positive and on the order of one additional answer correct (out of a possible 32). For the anthropometric outcomes, we find the counterintuitive result of a negative impact of the program on weight and on height. 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, which suggests that the regression models may be misspecified. One potential source of misspecification is that program impacts may depend in a nonadditive way on age and program duration. The matching estimators described below nonparametrically estimate the nature of the dependence.

We also consider estimates based on a DID estimator 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 E2, for the coefficient estimates). The estimates are imprecise due to the substantially reduced sample sizes. In this case, the estimates suggest that the effects of the program are negative for all outcomes except the height-for-age percentile (see figure 5).

G. Cumulative Impacts Estimated by the Method of Matching

We next describe estimated cumulative program impacts based on the matching estimators developed in section III, first conditional on age only and then conditional on both age and duration in the program. We also 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.[37]

Table 3 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 equation (7). Each age interval represents a quintile of the participant age distribution. The "Mean Diff." column shows the difference in raw mean outcomes, the "Mag" column shows the estimated program impact obtained by the matching method, and the "%" column shows 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.

The test score impacts are almost all positive for children aged 37–58 months. For this age group, the program is estimated to increase test scores by roughly one additional correct item, which is 3%–4% 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, widely acclaimed in the United States as a successful program, found an increase in test score outcomes of only 6% (Krueger, 1998). With regard to the anthropometric measures, the estimated program impacts are imprecisely estimated.

Table 4 reports estimated impacts conditional on specific age and duration ranges. The estimates are obtained by first estimating mean impacts at each age and duration value

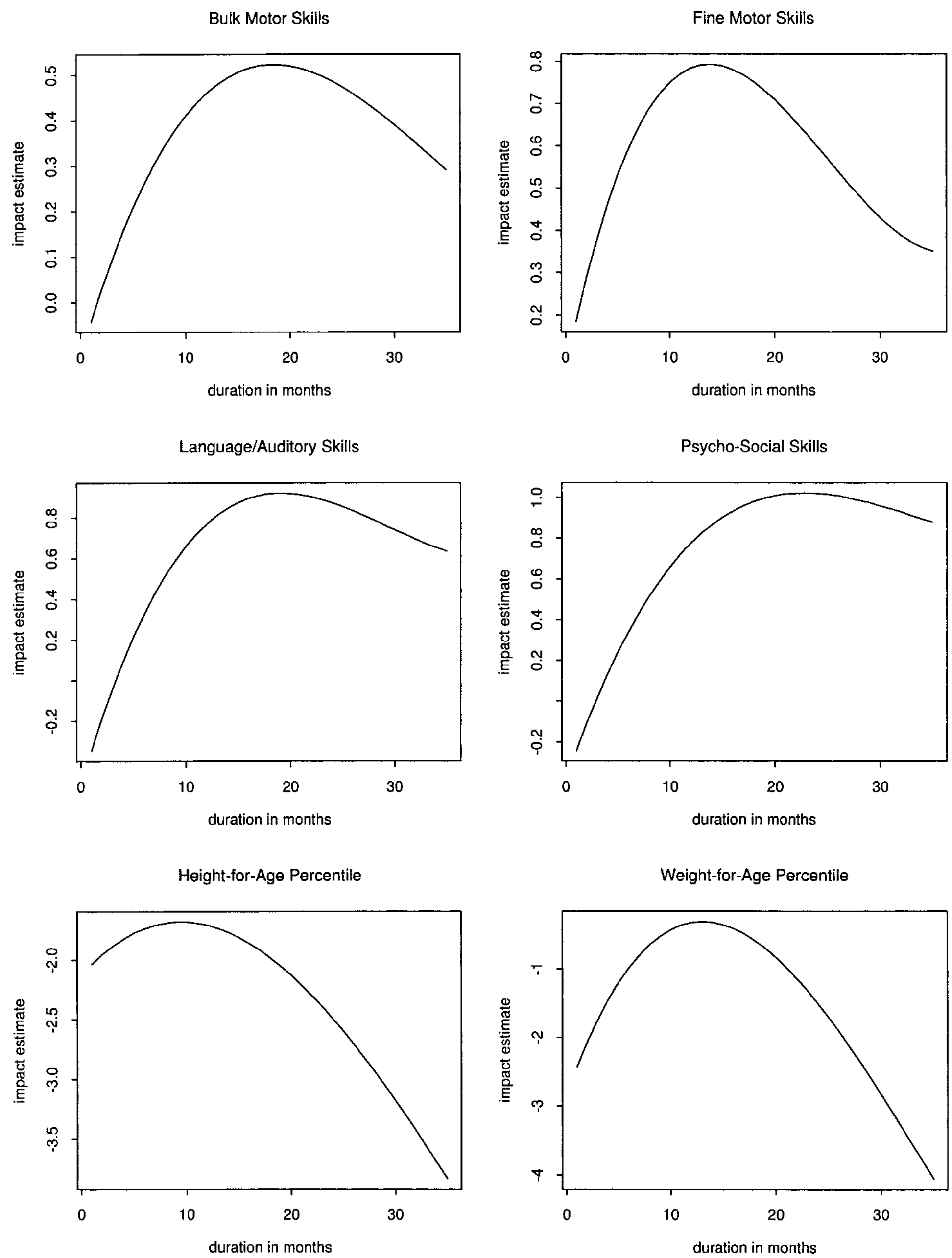
[34] This is a version of the estimator described in section IIIA under "Estimating Marginal Program Impacts" that integrates over the observed program durations greater than or equal to 2 months.

[35] When we use the survival probability calculated only using the data on program participants, there is no 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.

[36] For brevity, the regression estimates are shown in Appendix E (which is available upon request from the authors) in table E1. The model explains considerable shares of the variance in the four test scores (84% or more) but much less of the sample variance in the anthropometric indicators (approximately 4%). Family background characteristics are found to be 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).

[37] The grid is three-dimensional for estimating the expectation conditional on $D_P = 1$, and two-dimensional for estimating the expectation conditional on $D_P = 0$. The values over which we searched are 1.0 through 16.0 for the log odds ratio with a step size of 1, 1.0 through 28.0 for age with a step size of 1 month, and 1.0 through 28.0 for duration with a step size of 1 month.

FIGURE 4.—ESTIMATED PROGRAM IMPACTS FROM CROSS-SECTIONAL, CUBIC-IN-DURATION MODEL, SAMPLES *P* AND *B*

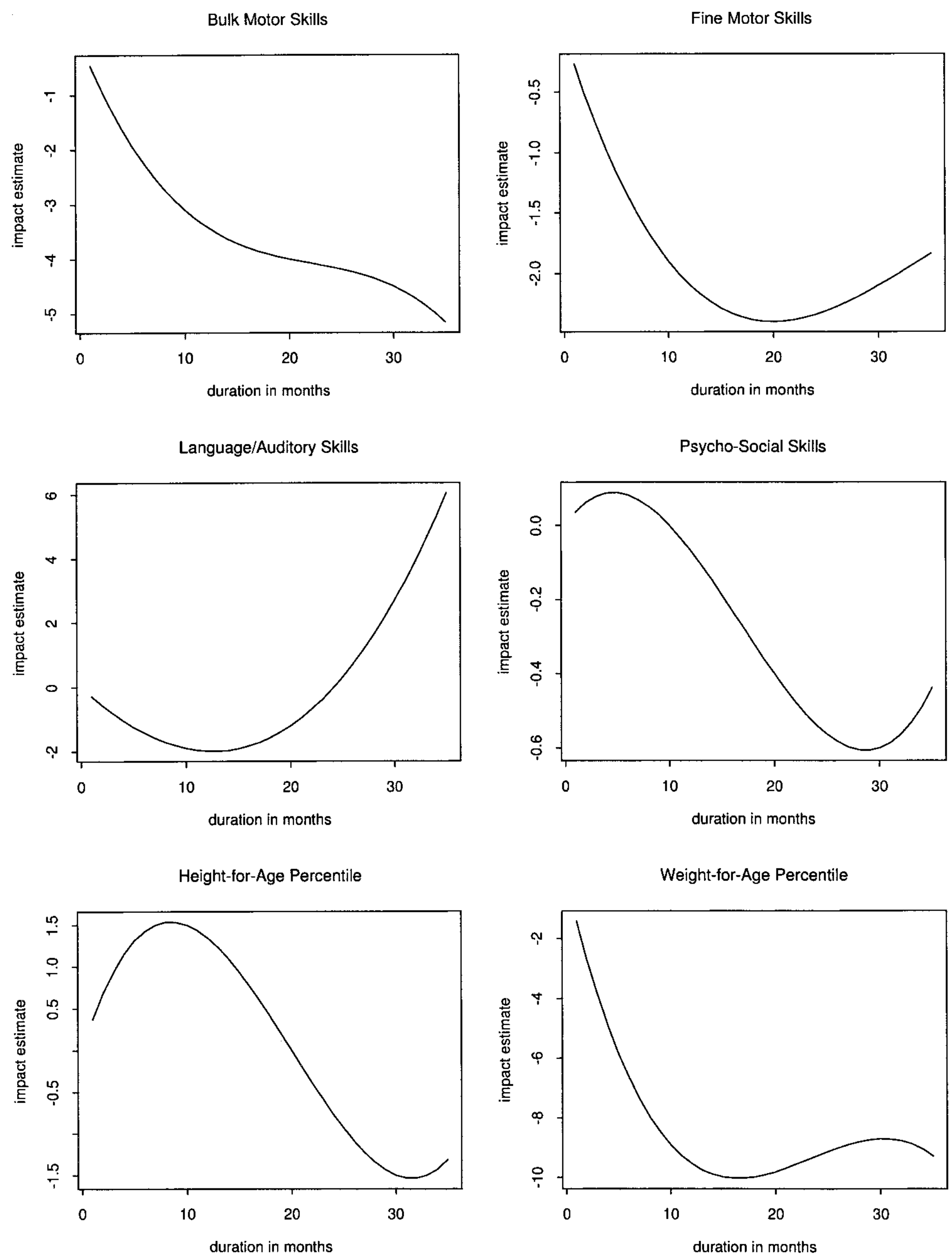


observed in the data and then taking averages over the individual impacts within each age-duration class.[38] The results indicate that average impacts increase as length of exposure to treatment increases. Impacts are almost always positive for children who have participated in the program for at least 13 months (with only two exceptions, both for children under 36 months old who have participated 25+ months) and roughly twice the order of magnitude of the overall average impacts reported in table 3. They tend to be larger than those found under the cubic specification of section IV C. The bottom two panels of table 4 show results for the anthropometric measures, which are insignificantly different from 0.

We 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 II). Tables 5 and 6 report the estimated program impacts in a format identical to the previous two tables. The impact estimates on test scores are generally positive over all age ranges for durations of exposure of at least 7 months. The estimates are more widely statistically significant for the *A* comparison group than for the *B* group;

[38] Averages are therefore self-weighting by the joint duration and age density.

FIGURE 5.—ESTIMATED PROGRAM IMPACTS BASED ON DIFFERENCE-IN-DIFFERENCES MODEL, SAMPLES P AND B



they are significant over half the time for children aged 36 months or less, and almost always for children aged 59+ months. With regard to the anthropometric measures, we find statistically significant negative impacts on weight for short durations of exposure. We think it unlikely that the program could decrease children's weight over a short time interval by as much as estimated. Rather, the estimated negative impacts on weight are probably evidence of bias arising from selection into the program on unobservables that is not taken into account by matching. An important unobservable is preprogram nutritional status, on which, as noted in section IV B, the initial program eligibility criteria placed primary emphasis.

H. *Marginal Program Impacts Estimated by the Method of Matching*

Because preprogram nutritional status represents an important unobservable, our 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 IIIA, the marginal estimator allows selectivity into

TABLE 3.—COMPARISON OF DIFFERENCE IN RAW MEANS AND CUMULATIVE MEAN PROGRAM IMPACTS AFTER ADJUSTING FOR SELECTIVITY INTO THE PROGRAM USING MATCHING METHOD—SAMPLES P AND B

	Age: 6–24			25–36			37–41			42–58			59+ mo.		
	Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact	
Outcome	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%
Bulk motor skills	2.17	−0.03	−0	0.33	−0.04	−0	0.60	0.80	3	0.56	0.85	3	0.09	0.00	0
		(0.21)			(0.29)			(0.24)			(0.22)			(0.29)	
Fine motor skills	2.43	0.34	3	0.23	0.12	1	0.12	0.33	2	0.76	0.52	2	0.23	−0.05	−0
		(0.17)			(0.27)			(0.11)			(0.22)			(0.40)	
Language	1.56	−0.02	−0	0.16	−0.19	−1	1.00	0.73	3	0.78	0.76	3	0.28	0.16	1
		(0.17)			(0.33)			(0.24)			(0.30)			(0.29)	
Psychosocial	1.85	−0.40	−4	0.68	0.03	0	1.10	1.35	6	0.74	0.53	2	0.24	−0.30	−1
		(0.28)			(0.41)			(0.34)			(0.23)			(0.35)	
Weight perc.	−3.8	−3.2	−10	−2.9	−0.36	−1	−5.3	0.40	1	0.41	0.74	2	2.20	−1.6	−5
		(3.99)			(2.03)			(1.64)			(1.40)			(1.97)	
Height perc.	−2.7	2.55	13	−2.5	1.22	7	−4.6	−0.59	−3	−0.36	−0.85	−5	−3.6	0.38	2
		(3.12)			(1.72)			(1.76)			(1.39)			(2.28)	

Bootstrapped standard errors in parentheses.

TABLE 4.—ESTIMATED CUMULATIVE IMPACTS BY DURATION AND AGE CLASSES—SAMPLES P AND B

Age in Months	Duration 1–6	7–12	13–18	19–24	25+ mo.
			Bulk Motor Skills		
6–24	−0.05 (−0%)	0.04 (0%)	0.09 (1%)	·	·
25–36	−0.43 (−2%)	−0.06 (−0%)	0.72 (4%)*	0.99 (5%)*	0.36 (2%)
37–41	0.40 (2%)	0.82 (3%)*	1.32 (6%)*	1.41 (6%)*	1.20 (5%)*
42–58	0.31 (1%)	1.02 (4%)*	1.39 (5%)*	1.12 (4%)*	1.54 (6%)*
59+	−0.49 (−2%)	0.38 (1%)	0.32 (1%)	0.35 (1%)	0.43 (1%)
			Fine Motor Skills		
6–24	0.27 (2%)	0.45 (4%)	0.87 (7%)*	·	·
25–36	0.02 (0%)	0.10 (1%)	0.40 (2%)	0.35 (2%)	−0.24 (−1%)
37–41	0.29 (1%)	0.34 (2%)*	0.53 (2%)*	0.40 (2%)*	0.09 (0%)
42–58	0.03 (0%)	0.81 (3%)*	1.05 (4%)*	0.73 (3%)*	0.98 (4%)*
59+	−0.80 (−3%)	0.38 (1%)	0.59 (2%)	0.53 (2%)	0.57 (2%)
			Language and Auditory Skills		
6–24	−0.11 (−1%)	0.27 (2%)	0.22 (2%)	·	·
25–36	−0.48 (−3%)	0.01 (0%)	0.25 (1%)	0.34 (2%)	−0.71 (−4%)
37–41	0.36 (2%)	1.19 (6%)*	1.42 (7%)*	0.95 (4%)*	0.42 (2%)
42–58	−0.09 (−0%)	1.12 (5%)*	1.65 (7%)*	1.38 (6%)*	1.54 (6%)*
59+	−0.76 (−3%)*	0.77 (3%)*	0.69 (2%)*	1.26 (4%)*	0.68 (2%)
			Psychosocial Skills		
6–24	−0.49 (−4%)*	−0.18 (−2%)	0.01 (0%)	·	·
25–36	−0.57 (−3%)	0.11 (1%)	0.83 (4%)*	1.72 (9%)*	1.51 (8%)*
37–41	0.59 (3%)	1.43 (6%)*	2.09 (9%)*	2.51 (11%)*	2.35 (10%)*
42–58	0.00 (0%)	0.75 (3%)*	1.05 (4%)*	0.81 (3%)*	1.15 (4%)*
59+	−1.11 (−4%)*	0.10 (0%)	0.15 (1%)	0.56 (2%)	0.53 (2%)
			Weight Percentile		
6–12	−4.10 (−13%)	−1.58 (−5%)	0.13 (0%)	·	·
13–18	−1.72 (−5%)	0.20 (1%)	1.02 (3%)	2.38 (7%)	3.61 (11%)
19–24	−0.84 (−2%)	0.65 (2%)	1.38 (4%)	2.00 (5%)	2.87 (8%)
25–30	−0.84 (−3%)	1.13 (3%)	1.57 (5%)	1.73 (5%)	3.43 (10%)
31–36	−3.56 (−11%)	−0.66 (−2%)	−0.18 (−1%)	0.28 (1%)	−0.17 (−1%)
			Height Percentile		
6–12	2.63 (13%)	2.68 (14%)	1.59 (8%)	·	·
13–18	1.80 (10%)	1.36 (8%)	0.76 (4%)	0.24 (1%)	−0.93 (−5%)
19–24	−0.02 (−0%)	−0.33 (−2%)	−1.04 (−5%)	−1.15 (−5%)	−1.52 (−7%)
25–30	−0.37 (−2%)	−0.63 (−4%)	−1.08 (−7%)	−1.46 (−9%)	−1.35 (−9%)
31–36	−0.31 (−2%)	0.92 (5%)	1.09 (6%)	0.86 (5%)	−0.20 (−1%)

* Significant at the 10% level.
Percentage impact shown in parentheses.

TABLE 5.—COMPARISON OF DIFFERENCE IN RAW MEANS AND CUMULATIVE MEAN PROGRAM IMPACTS AFTER ADJUSTING FOR SELECTIVITY INTO THE PROGRAM USING MATCHING METHOD—SAMPLES P AND A

	Age: 6–24			25–36			37–41			42–58			59+ mo.		
	Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact	
Outcome	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%
Bulk motor skills	1.74	0.22	2	0.41	0.28	2	0.65	0.51	3	0.61	0.79	4	0.58	0.79	4
		(0.17)			(0.18)			(0.18)			(0.23)			(0.15)	
Fine motor skills	1.88	0.41	5	0.22	0.20	2	0.38	0.17	1	1.20	1.30	7	1.22	1.61	8
		(0.09)			(0.10)			(0.14)			(0.15)			(0.22)	
Language	1.11	0.02	0	0.14	−0.30	−2	0.40	−0.05	−0	0.97	1.05	6	0.69	0.68	4
		(0.05)			(0.16)			(0.14)			(0.13)			(0.24)	
Psychosocial	1.49	0.16	2	0.37	0.10	1	1.15	0.34	2	0.95	1.10	6	0.68	0.66	3
		(0.09)			(0.21)			(0.18)			(0.29)			(0.29)	
Weight perc.	−5.0	−6.0	−16	−5.1	−7.5	−28	−1.4	−4.9	−15	−0.36	1.07	3	−3.9	−5.0	−14
		(2.10)			(2.69)			(2.39)			(1.37)			(2.48)	
Height perc.	−4.3	3.26	12	−6.7	1.69	9	−0.60	−1.0	−5	−4.2	−.61	−3	−4.7	1.42	7
		(4.27)			(4.72)			(2.87)			(1.88)			(2.68)	

Bootstrapped standard errors in parentheses.

TABLE 6.—ESTIMATED CUMULATIVE IMPACTS BY DURATION AND AGE CLASSES—SAMPLES P AND A

Age in Months	Duration: 1–6	7–12	13–18	19–24	25+ mo.
			Bulk Motor Skills		
6–24	0.13 (1%)	0.37 (4%)*	0.73 (8%)*	·	·
25–36	−0.09 (−1%)	0.41 (3%)	0.96 (7%)*	0.97 (7%)*	0.07 (1%)
37–41	0.15 (1%)	0.58 (3%)*	0.88 (5%)*	0.97 (6%)*	1.07 (6%)*
42–58	0.35 (2%)*	0.88 (4%)*	1.23 (6%)	1.10 (6%)*	1.42 (7%)*
59+	0.43 (2%)	1.09 (5%)*	0.92 (4%)	1.08 (5%)*	1.19 (6%)*
			Fine Motor Skills		
6–24	0.30 (3%)*	0.57 (6%)*	1.02 (11%)*	·	·
25–36	0.10 (1%)	0.20 (2%)	0.46 (4%)*	0.42 (3%)*	−0.21 (−2%)
37–41	0.10 (1%)	0.19 (1%)	0.40 (2%)*	0.24 (1%)	0.01 (0%)
42–58	0.78 (4%)*	1.71 (9%)*	1.82 (10%)*	1.49 (8%)*	1.73 (9%)*
59+	0.92 (5%)*	2.02 (10%)*	2.09 (11%)*	2.22 (11%)*	2.23 (11%)*
			Language and Auditory Skills		
6–24	−0.14 (−2%)	0.37 (4%)	0.56 (6%)	·	·
25–36	−0.57 (−5%)*	−0.03 (−0%)	0.24 (2%)	−0.21 (−2%)	−0.98 (−8%)*
37–41	−0.49 (−3%)	0.36 (3%)	0.77 (5%)*	0.14 (1%)	−0.12 (−1%)
42–58	0.23 (1%)	1.44 (8%)*	1.88 (11%)*	1.72 (10%)*	1.74 (10%)*
59+	−0.25 (−1%)	1.18 (6%)*	1.17 (6%)*	1.83 (10%)*	1.35 (7%)*
			Psychosocial Skills		
6–24	−0.12 (−1%)	0.80 (9%)*	1.08 (12%)*	·	·
25–36	−0.42 (−3%)*	0.45 (4%)	0.84 (7%)*	0.83 (7%)*	0.09 (1%)
37–41	−0.26 (−2%)	0.53 (4%)*	0.95 (7%)*	1.07 (7%)*	1.10 (7%)*
42–58	0.52 (3%)*	1.34 (7%)*	1.63 (9%)*	1.57 (8%)*	1.72 (9%)*
59+	−0.14 (−1%)	1.09 (5%)*	1.23 (6%)*	1.43 (7%)*	1.29 (6%)*
			Weight Percentile		
6–24	−6.84 (−19%)*	−4.37 (−12%)	−2.40 (−7%)	·	·
25–36	−8.86 (−33%)*	−7.63 (−29%)*	−5.50 (−21%)	−4.86 (−18%)	−0.08 (−0%)
37–41	−6.77 (−21%)*	−3.89 (−12%)	−4.23 (−13%)	−3.46 (−11%)	1.06 (3%)
42–58	−0.68 (−2%)	3.63 (11%)	2.52 (8%)	0.48 (1%)	2.62 (8%)
59+	−7.65 (−22%)*	−2.10 (−6%)	−0.79 (−2%)	−4.36 (−12%)	−5.67 (−16%)
			Height Percentile		
6–24	3.14 (12%)	3.73 (14%)	2.66 (10%)	·	·
25–36	2.28 (12%)	1.87 (10%)	1.22 (7%)	0.66 (4%)	−0.57 (−3%)
37–41	−0.48 (−2%)	−0.73 (−4%)	−1.41 (−7%)	−1.62 (−8%)	−1.86 (−9%)
42–58	−0.34 (−2%)	−0.30 (−1%)	−0.74 (−3%)	−1.17 (−5%)	−0.92 (−4%)
59+	0.51 (2%)	2.09 (10%)	2.21 (10%)	1.85 (9%)	1.07 (5%)

* Significant at the 10% level.
Percentage impact shown in parentheses.

TABLE 7.—COMPARISON OF DIFFERENCE IN RAW MEANS AND CUMULATIVE MEAN PROGRAM IMPACTS AFTER ADJUSTING FOR SELECTIVITY INTO THE PROGRAM USING MATCHING METHOD—GROUP P, DUR. ≥ 2 AND DUR. ≤ 1

	Age: 6–24			25–36			37–41			42–58			59+ mo.		
	Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact		Mean	Impact	
Outcome	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%	Diff.	Mag.	%
Bulk motor skills	0.57	−0.08	−1	−0.00	0.19	1	0.16	0.15	1	0.42	0.45	2	0.68	0.53	3
		(0.11)			(0.11)			(0.20)			(0.22)			(0.19)	
Fine motor skills	0.88	0.21	2	−0.16	0.02	0	0.32	−0.03	−0	0.09	0.89	5	0.10	0.69	3
		(0.20)			(0.14)			(0.28)			(0.15)			(0.29)	
Language	1.12	0.30	3	0.23	0.27	2	0.56	0.11	1	0.61	1.41	9	0.81	0.76	4
		(0.31)			(0.19)			(0.43)			(0.29)			(0.17)	
Psychosocial	1.18	0.41	5	0.21	0.66	5	0.88	0.87	6	0.80	0.98	5	1.45	0.74	4
		(0.29)			(0.19)			(0.32)			(0.10)			(0.28)	
Weight perc.	−6.7	−0.10	−0	9.55	1.39	8	1.69	−1.9	−7	−0.43	0.25	1	6.57	0.83	4
		(3.09)			(3.45)			(4.21)			(2.56)			(3.91)	
Height perc.	−5.1	2.89	11	10.2	1.79	17	1.96	−1.1	−9	3.33	−1.5	−11	4.16	0.28	3
		(2.16)			(2.75)			(3.38)			(1.84)			(2.42)	

Bootstrapped standard errors in parentheses.

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 3. The test score impacts are mostly positive for children of all age ranges. The marginal estimates indicate generally somewhat larger effects than do the average estimates and also are suggestive 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 over half the age ranges, although none of the estimates are statistically significant.

Table 8 shows estimated impacts conditional on age and duration in the program. The test score results suggest increasing marginal impacts with greater program exposure. The estimates are mostly positive and tend to be larger than the overall average marginal impacts in table 6 for children who have participated in the program for at least 6 months. For children aged 6–36 months, the estimated impacts on height and weight percentiles are also generally positive for different durations, except for children older than 36 months, for whom the height estimates are negative. For younger children, the height estimates for short durations are surprisingly large and positive.[39] For weight percentiles the marginal effect estimates are more credible than the cumulative estimates (table 4). These comparisons suggest that the first criterion for selecting children into the program (malnourishment) focused on low weight and not low stature.

V. Cost-Benefit Analysis

So far we have considered only the problem of estimating the benefits of the program. Next we consider whether the benefits outweigh the costs, which have been estimated to be approximately $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), which directly affects earnings, (2) by increasing physical stature as an adult, which directly affects earnings, (3) by increasing the number of grades completed, which directly affects earnings and the age *a* of school completion, 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 facilitate earlier entry into school, lessen repetition rates, and lead to more grades completed. Appendix D summarizes empirical evidence on the importance of these four channels from the experience of developing countries.

As our data do not provide information on how higher cognitive skills and better nutrition affect 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 Latin 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.[40] To our knowledge, there has been no research on the cognitive-skills–earnings relationship specifically for Latin American workers, so we base our cost-benefit analysis on a study by Alderman et al. (1996) of the cognitive-skills–earnings relationship for male workers in Pakistan,

[39] A 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.

[40] Their study uses a normal bias correction to control for selectivity into employment.

TABLE 8.—ESTIMATED MARGINAL IMPACTS BY DURATION AND AGE CLASSES—SAMPLE P, DUR. ≥ 2 AND DUR. ≤ 1

Age in Months	Duration 2–6	7–12	13–18	19–24	25+ mo.
			Bulk Motor Skills		
6–24	−0.20 (−1%)	−0.12 (−1%)	0.80 (6%)*	·	·
25–36	−0.14 (−1%)	0.32 (2%)	0.65 (3%)*	0.56 (3%)	−0.86 (−4%)
37–41	−0.04 (−0%)	0.15 (1%)	0.20 (1%)	0.34 (1%)	0.37 (2%)
42–58	−0.24 (−1%)	0.44 (2%)	0.93 (3%)*	0.77 (3%)*	1.09 (4%)*
59+	0.23 (1%)	0.59 (2%)	0.69 (2%)*	0.46 (2%)	0.94 (3%)*
			Fine Motor Skills		
6–24	0.05 (0%)	0.32 (2%)	0.93 (7%)*	·	·
25–36	−0.09 (−0%)	0.08 (0%)	0.28 (1%)	0.10 (1%)	−0.83 (−4%)*
37–41	−0.06 (−0%)	−0.08 (−0%)	0.19 (1%)	−0.01 (−0%)	−0.26 (−1%)
42–58	0.34 (1%)*	0.98 (4%)	1.33 (6%)*	1.08 (5%)*	1.24 (5%)*
59+	0.11 (0%)	0.76 (3%)	0.90 (3%)	1.04 (4%)*	1.20 (4%)*
			Language and Auditory Skills		
6–24	0.10 (1%)	0.51 (4%)	0.95 (8%)*	·	·
25–36	0.07 (0%)	0.42 (2%)*	0.54 (3%)*	0.46 (3%)	−0.62 (−4%)
37–41	−0.18 (−1%)	0.28 (1%)	0.51 (2%)	0.27 (1%)	−0.13 (−1%)
42–58	0.31 (1%)	1.59 (6%)*	2.16 (9%)*	1.95 (8%)*	2.18 (9%)*
59+	−0.06 (−0%)	0.94 (3%)*	1.01 (4%)*	1.37 (5%)*	1.27 (5%)*
			Psychosocial Skills		
6–24	0.15 (1%)	0.69 (6%)	1.21 (10%)*	·	·
25–36	0.13 (1%)	0.75 (4%)*	1.07 (6%)*	1.62 (9%)*	0.88 (5%)
37–41	0.64 (3%)*	0.70 (3%)*	1.01 (4%)*	1.19 (5%)*	1.25 (5%)*
42–58	0.35 (1%)	0.95 (4%)*	1.44 (6%)*	1.31 (5%)*	1.54 (6%)*
59+	−0.00 (−0%)	0.82 (3%)*	1.05 (4%)*	1.26 (4%)*	1.26 (4%)*
			Weight Percentile		
6–24	−0.71 (−3%)	0.22 (1%)	3.36 (12%)	·	·
25–36	0.74 (3%)	0.75 (3%)	1.89 (7%)	3.30 (13%)	5.22 (20%)
37–41	−1.60 (−5%)	−3.03 (−9%)	−3.54 (−11%)	−0.92 (−3%)	1.06 (3%)
42–58	−1.08 (−3%)	1.05 (3%)	1.14 (4%)	−0.21 (−1%)	1.51 (5%)
59+	−2.35 (−7%)	2.93 (9%)	4.27 (13%)	1.27 (4%)	0.34 (1%)
			Height Percentile		
6–24	3.10 (19%)	2.78 (17%)	1.63 (10%)	·	·
25–36	2.28 (19%)	1.94 (16%)	1.44 (12%)	0.98 (8%)	−0.50 (−4%)
37–41	−0.55 (−3%)	−0.54 (−3%)	−1.52 (−9%)	−1.92 (−11%)	−1.94 (−11%)
42–58	−1.22 (−8%)	−1.16 (−7%)	−1.48 (−9%)	−2.05 (−13%)	−1.86 (−12%)
59+	−1.04 (−9%)	1.25 (10%)	1.57 (13%)	0.68 (6%)	−0.18 (−1%)

* Significant at the 10% level.

which finds that a 1% increase in cognitive skills increases earnings by 0.23%.[41] 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.[42,43]

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 s grades completed, cognitive ability c, and height h, and 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* (Integrated Household Survey), which covers the ten most populous urban areas in Bolivia, indicate that for 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.

[41] Their 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.

[42] Measures 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 (Currie & Thomas, 1999).

[43] We 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 small percentage increase in height in comparison with those obtained for some of the estimators in section IV, because we expect that parents may have selected taller children for consideration for the program.

TABLE 9.—COSTS AND ESTIMATED BENEFITS OF THE PIDI PROGRAM IN U.S. DOLLARS UNDER DIFFERENT HYPOTHETICAL IMPACTS*

Educational Level	Mean Annual Earnings‡ ($)	Discount Rate 3%			5%		
		Cost†	Benefit	Benefit/Cost Ratio	Cost	Benefit	Benefit/Cost Ratio
Intermed. (8)	1224 → 1352	1394	5107	3.66	1301	3230	2.48
Secondary (11)	1422 → 1550	1394	3969	2.85	1301	2232	1.72
Intermed. (8)	1224 → 1352	1743	5107	2.93	1626	3230	1.99
Secondary (11)	1422 → 1550	1743	3969	2.28	1626	2232	1.37

* Assuming that children take part in program 3 years, from age 2 to age 5. Impact: Shortens length of time to complete education by 1 year, increases average educational attainment level by 1 year, increases cognitive skills by 5%, and increases height by 2%. Our simulation is based on a point estimate reported in Strauss and Thomas (1997) of a 2.4% increase in earnings for each 1% increase in height, and on a point estimate reported in Alderman et al. (1996), which finds a 0.233% increase in earnings for each 1% increase in cognitive skills.

† The first two lines of estimates are based on a cost of $516/year as estimated by Ruiz (1996). The second set of estimates include a 25% upward adjustment to costs to allow for possible distortionary costs to the government of raising the revenues to pay for the program.

‡ Conversion factor: 7.8 Bolivianos/1 U.S. dollar.

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 $V(s, a, h, c) = \int_a^{60} y(s, h, c)e^{-rt}dt$.[44] This yields a present discounted value of earnings equal to $V(s, a, h, c) = r^{-1}y(s, c, h)(e^{-ra} - e^{-r60})$.

The expected impact of a 2% increase in height is $\bar{y}(s) \times 2 \times 0.024 \times r^{-1}(e^{-ra} - e^{-r60})$, where $\bar{y}(s)$ is the average earnings for men with s grades completed and we use the results, as noted, from Strauss and Thomas's (1997) study and from Alderman et al. (1996).

The earnings gain that would result from a decrease in the school completion age from a_1 to a_2 without changing the level of school attainment is given by $\bar{y}(s)r^{-1}[e^{-ra_2} - e^{-ra_1}]$. 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

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

On the cost side, the cost of participating in the program for 4 years between ages 2 and 5 is given by $\$516 \int_2^5 e^{-rt}\, dt$.

Table 9 reports the cost-benefit estimates under hypothetical program impacts that are in the range of some of the impacts observed in the impact analysis of section IV E (table 8) and for average male earnings levels associated with three different education levels: 8, 11, and 14 years of education.[45] Specifically, we obtain an earnings gain resulting from an impact of 2% on height, a 5% increase in cognitive skills, and a 1-year increase in grades completed and a corresponding 1-year increase in the age of school completion. The tables shows estimates for two values of the discount rate, $r = 3\%$ or $r = 5\%$.

The single impact that has the largest effect among the ones considered is increasing the number of grades completed (under the assumption that there is a corresponding 1-year increase in the age of completion), which alone would generate a benefit-cost ratio greater than 1 for both discount rates and both education levels. When multiple types of program impacts are considered together (as shown in the table), the benefit/cost ratios range from 1.7 to 3.7. We also estimated the cost-benefit ratios adjusting the costs by an additional 25% to allow for distortionary costs to the government of raising the revenues to pay for the program.[46]

VI. Conclusions

This paper analyzes the impact of a preschool program in a developing country using a relatively large, nonexperimental data set. To do so, we generalize matching methods to allow the program impact to vary with a continuous treatment dose (the duration of time spent in the program) and to depend in a flexible way on the age of the child. We also develop a marginal effect estimator that assumes that program participants with differing durations of participation can be made comparable by conditioning on observed child and family characteristics. Advantages of the marginal effect estimator are that it does not require assumptions on the process governing selection into the program, can accommodate the case where selection into the program is on unobservable characteristics, and can be implemented using data only on program participants.

We applied several different estimators to evaluate the effectiveness of the PIDI program in Bolivia, which is aimed at improving early cognitive skills and nutrition. We developed a dynamic model for the decision to enroll children that provided an interpretation for the treatment impact estimates and guided our selection of matching variables. Impact estimates based on cross-sectional regres-

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

[45] Mean 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 modal number of years of education for these males is 8.

[46] The 25% figure is based on studies such as Devarajan, Squire, and Suthiwart-Narueput (1997), Feldstein (1995), and Harberger (1997). As seen in the table, even with this adjustment the cost-benefit ratios are substantially greater than 1.

sion estimators indicate a positive, statistically significant effect on test scores. Our matching estimators show that test score gains depend strongly on duration of exposure to the program, with positive effects observed for children who participated at least 7 months (for the marginal estimator) and increasing effects observed with longer durations. However, the impacts on the anthropometric outcomes are not precisely estimated.

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, under the assumptions of our simulations, the expected benefit of the program outweighs the costs by a fair amount.

REFERENCES

Alderman, Harold, Jere R. Behrman, Victor Levy, and Rekha Menon, "Child Health and School Enrollment: A Longitudinal Analysis," *Journal of Human Resources* 36:1 (2001), 185–205.

Alderman, Harold, Jere R. Behrman, David Ross, and Richard Sabot, "The Returns to Endogenous Human Capital in Pakistan's Rural Wage Labour Market," *Oxford Bulletin of Economics and Statistics* 58:1 (1996), 29–56.

Barnett, Stephen, "The Benefits of Compulsory Preschool Education," *Journal of Human Resources* 27:2 (1992), 279–312.

Behrman, Jere R., "The Economic Rationale for Investing in Nutrition in Developing Countries," *World Development* 21:11 (1993), 1749–1771.

Behrman, Jere R., and Anil B. Deolalikar, "Wages and Labor Supply in Rural India: The Role of Health, Nutrition and Seasonality," in David E. Sahn (Ed.), *Causes and Implications of Seasonal Variability in Household Food Security* (Baltimore, MD: The Johns Hopkins University Press, 1989).

Bloom, H., L. Orr, G. Dave, S. H. Bell, and F. Doolittle, *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months* (Bethesda, MD: Abt Associates, 1993).

Boissiere, Maurice, John B. Knight, and Richard H. Sabot, "Earnings, Schooling, Ability and Cognitive Skills," *American Economic Review* 75 (1985), 1016–1030.

Cleveland, William, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association* 74 (1979), 829–836.

Currie, Janet, and Duncan Thomas, "Early Test Scores, Socioeconomic Status and Future Outcomes," National Bureau of Economic Research working paper no. 6943 (1999).

Dehejia, Rajeev, and Sadek Wahba, "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association* 94:448 (1999), 1053–1062.

Deolalikar, Anil B., "Nutrition and Labor Market Productivity in Agriculture: Estimates for Rural South India," *Review of Economics and Statistics* 70:3 (1988), 406–413.

Devarajan, S., L. Squire, and S. Suthiwart-Narueput, "Beyond Rate of Return: Reorienting Project Appraisal," *The World Bank Research Observer* 12:1 (1997), 35–36.

Fan, J., "Design Adaptive Nonparametric Regression," *Journal of the American Statistical Association* 87 (1992), 998–1004.

Feldstein, Martin, "Tax Avoidance and the Deadweight Loss of the Income Tax," National Bureau of Economic Research working paper no. 5055 (1995).

Glewwe, Paul, "The Relevance of Standard Estimates of Rates of Return to Schooling for Education Policy: A Critical Assessment," *Journal of Development Economics* 51:2 (1996), 267–290.

Glewwe, Paul, and Hanan G. Jacoby, "An Economic Analysis of Delayed Primary School Enrollment in a Low-Income Country: The Role of Early Childhood Nutrition," this REVIEW 77:1 (1995), 156–159.

Glewwe, Paul, Hanan G. Jacoby, and Elizabeth King, "Early Childhood Nutrition and Academic Achievement: Analysis Using Longitudinal Data," *Journal of Public Economics* 81:3 (2001), 345–368.

Grantham-McGregor, S., C. Walker, S. Chang, and C. Powell, "Effects of Early Childhood Supplementation With and Without Stimulation on Later Development in Stunted Jamaican Children," *American Journal of Clinical Nutrition* 66 (1997), 247–453.

Haas, J., S. Murdoch, J. Rivera, and R. Martorell, "Early Nutrition and Later Physical Work Capacity," *Nutrition Reviews* 54 (1996), S41–S48.

Haddad, Lawrence, and Howarth Bouis, "The Impact of Nutritional Status on Agricultural Productivity: Wage Evidence from the Philippines," *Oxford Bulletin of Economics and Statistics* 53:1 (1991), 45–68.

Hahn, Jinyong, "On the Role of the Propensity Score in Efficient Estimation of Average Treatment Effects," *Econometrica* 66:2 (1998), 315–331.

Harberger, Arnold, "New Frontiers in Project Evaluation? A Comment on Devarajan, Squire and Suthiwart-Narueput," *The World Bank Research Observer* 12:1 (1997), 73–39.

Heckman, J. J., H. Ichimura, J. Smith, and P. Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica* 66 (1998), 1017–1098.

Heckman, J. J., H. Ichimura, and P. Todd, "Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies* 64:4 (1997), 605–654.

——— "Matching As an Econometric Evaluation Estimator," *Review of Economic Studies* 65 (1998), 261–294.

Heckman, J. J., and P. Todd, "Adapting Propensity Score Matching and Selection Models to Choice-based Samples," University of Chicago, unpublished manuscript (1995).

Jamison, Dean T., "Child Nutrition and School Performance in China," *Journal of Development Economics* 20:2 (1986), 299–310.

Krueger, Alan, "Reassessing the View that American Schools Are Broken," *Federal Reserve Bank of New York Economic Policy Review* 4:1 (1998), 29–43.

Lavy, Victor, Jennifer Spratt, and Nathalie Leboucher, "Patterns of Incidence and Change in Moroccan Literacy," *Comparative Education Review* 41:2 (1997), 120–141.

Martorell, R., "Results and Implications of the INCAP Follow-up Study," *Journal of Nutrition* 125 (1995), 1127S–1138S.

——— "The Nature of Child Malnutrition and its Long-term Implications," *Food and Nutrition Bulletin* 20 (1999), 288–292.

Martorell, R., K. L. Kahn, and D. G. Schroeder, "Reversibility of Stunting: Epidemiological Findings in Children from Developing Countries," *European Journal of Clinical Nutrition* 48:Suppl. (1994), S45–S57.

Martorell, Reynaldo, Juan Rivera, and Haley Kaplowitz, "Consequences of Stunting in Early Childhood for Adult Body Size in Rural Guatemala" (Stanford, CA: Food Policy Research), Stanford University mimeograph (1989).

Mincer, Jacob, "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy* 66:4 (1958), 281–302.

Moock, Peter R., and Joanne Leslie, "Childhood Malnutrition and Schooling in the Terai Region of Nepal," *Journal of Development Economics* 20:1 (1986), 33–52.

Myers, Robert, *The Twelve Who Survive: Strengthening Programmes of Early Childhood Development in the Third World* (Ypsilanti, MI: High Scope Press, 1995).

Neal, Derek, and William Johnson, "The Role of Pre-market Factors in Black-White Wage Differences," *Journal of Political Economy* 104:5 (1996), 869–895.

Pitt, Mark M., Mark R. Rosenzweig, and Donna M. Gibbons, "The Determinants and Consequences of the Placement of Government Programs in Indonesia," *The World Bank Economic Review* 7:3 (1993), 319–348.

Pollitt, Ernesto, *Malnutrition and Infection in the Classroom* (Paris: UNESCO, 1990).

Psacharopoulos, George, "Returns to Investment in Education: A Global Update," *World Development* 22:9 (1994), 1325–1344.
Ramey, Craig T., Frances A. Campbell, and Clancy Blair, "Enhancing the Life Course for High-Risk Children," in Jonathon Crane (Ed.), *Social Programs that Work* (New York: Russell Sage Foundation, 1998).
Rosenbaum, P., and D. Rubin, "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika* 70 (1983), 41–55.
Rosenzweig, Mark R., "Why Are There Returns to Schooling?" *American Economic Review* 85:2 (1995), 153–158.
Rosenzweig, Mark R., and Kenneth J. Wolpin, "Evaluating the Effects of Optimally Distributed Public Programs," *American Economic Review* 76:3 (1986), 470–487.
Ruiz, Fernando, "Estudio de Costos del Proyecto Integral de Desarrollo Infantil (PIDI)," La Paz, Bolivia: Ruiz Guissani Consultores mimeograph (1996).
Smith, Jeffrey, and Petra Todd, "Reconciling Conflicting Evidence on the Performance of Propensity Score Matching Methods," *American Economic Review* 91:2 (2001), 112–118.
——— "Does Matching Address LaLonde's Criticism of Nonexperimental Estimators," *Journal of Econometrics,* forthcoming (2004).
Strauss, John, "Does Better Nutrition Raise Farm Productivity?" *Journal of Political Economy* 94 (April 1986), 297–320.
Strauss, John, and Duncan Thomas, "Human Resources: Empirical Modeling of Household and Family Decisions," in J. Behrman and T. N. Srinivasan (Eds.), *Handbook of Development Economics,* vol. 3A (Amsterdam: North Holland Publishing Company, 1995).
——— "Health, Nutrition and Economic Development," *Journal of Economic Literature* 36:2 (1998), 766–817.
Sweinhart, Lawrence J., and David P. Weikart, "High/Scope Perry Preschool Program Effects at Age Twenty-Seven," in Jonathon Crane (Ed.), *Social Programs that Work* (New York: Russell Sage Foundation, 1998).
Thomas, Duncan, and Strauss, John, "Health and Wages: Evidence on Men and Women in Urban Brazil," *Journal of Econometrics* 77:1 (1997), 159–185.
Todd, Petra E., and Kenneth I. Wolpin, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *The Economic Journal* 113:485 (2003), F3–F33.
van de Walle, Dominique, and Kimberly Nead, *Public Spending and the Poor* (Baltimore: Johns Hopkins University Press, 1995).
World Bank, *World Development Report: The State in a Changing World* (New York and Oxford: Oxford University Press for the World Bank, 1997).

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, one 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 psychosocial tests, information on vaccination records and 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 (LLR) methods.[47] 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 that converges to 0 as $n \to \infty$. The estimator of the conditional mean is $\hat{a}$. If b_1 were constrained to equal 0, then $\hat{a}$ would give the standard kernel regression estimator. Thus, kernel regression can be viewed as a special case of 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.[48] The smaller bias associated with the LLR estimator implies that it is more rate-efficient than the kernel estimator. Another advantage 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.

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 1 month (used in comparing groups with durations >1 and durations ≤1 month). 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 data set 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 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 between number of rooms in house and age of child, interaction between employment status of father and age of child, interaction between number of siblings and age of child, interaction between monthly income of father and number of siblings, interaction between education level of mother and age of child.

Variables in the model for the probability of experiencing a duration that exceeds 1 month: age of child, sex of child, indicator for whether family participates in outside organizations, indicators for whether mother and father reside in household, education level of mother, job type of mother, age of father, education of father, monthly income of father, number of siblings, number of rooms in house, indicator for whether family owns house, indicator for whether house has running water, indicator for whether there is a bathroom in the house, indicator for whether household has a television set, interaction between number of

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

[48] The 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).

rooms and age of child, interaction between employment status of father and age of child, interaction between employment status of mother and the age of child, interaction between number of siblings and age of child, interaction between age of father and number of siblings, interaction between monthly income of father and number of siblings, interaction between education level of mother and age of child.

APPENDIX D

Empirical Evidence on Impact of Preschool Child Nutrition and Cognitive Development on Postschooling Earnings

To simulate benefits of improved preschool child nutrition and cognitive development on adult earnings, a number of channels must be considered as noted in section V. 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 Philippines; Strauss (1986) for Côte 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. (2001) for rural Pakistan; Glewwe and Jacoby (1995) for Ghana; and Glewwe, Jacoby, and King (2001) for the Philippines.

For our illustrative simulations, we use estimates from Alderman et al. (1996) for (1) and Thomas and Strauss (1997) 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.