

# Children's Health Opportunities and Project Evaluation

## Mexico's Oportunidades Program

*Dirk Van de gaer*  
*Joost Vandenbossche*  
*José Luis Figueroa*

The World Bank  
Development Economics Vice Presidency  
Partnerships, Capacity Building Unit  
January 2013



## Abstract

This paper proposes a methodology to evaluate social projects from the perspective of children's opportunities on the basis of the effects of these projects on the distribution of outcomes. The evaluation is conditioned on characteristics for which individuals are not responsible; in this case, parental education level and indigenous background. The methodology is applied to evaluate the effects on children's health opportunities

of Mexico's Oportunidades program, one of the largest conditional cash transfer programs for poor households in the world. The evidence from this program shows that gains in health opportunities for children from indigenous backgrounds are substantial and are situated in crucial parts of the distribution, whereas gains for children from nonindigenous backgrounds are more limited.

---

This paper is a product of the Partnerships, Capacity Building Unit, Development Economics Vice Presidency. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [Dirk.Vandegaer@ugent.be](mailto:Dirk.Vandegaer@ugent.be), [Joost.Vandenbossche@UGent.be](mailto:Joost.Vandenbossche@UGent.be), and [joseluis.figueroaoropeza@ugent.be](mailto:joseluis.figueroaoropeza@ugent.be).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

# Children's Health Opportunities and Project Evaluation: Mexico's Oportunidades Program

Dirk Van de gaer, Joost Vandenbossche, and José Luis Figueroa

**Keywords:** project evaluation, opportunities, Oportunidades program.

**JEL classification codes:** I18, I38, D63

This paper evaluates the change in health opportunities for children aged two to six years who participate in the Mexican Oportunidades program. Oportunidades is a large-scale, conditional cash transfer program initiated in 1998 through which poor rural households receive cash in exchange for their compliance with preventive health care requirements, nutrition supplementation, education, and monitoring. In 2010, approximately 5.8 million families participated in the program, and cash transfers to the participants totaled \$4.8 billion. The average treatment effects of the program on the health of young children have been shown to be positive (see the literature surveyed in Parker et al. 2008). We propose a methodology that focuses on the conditional cumulative distribution functions of health outcomes to identify whether and where in the distribution the program is effective for children whose parents have certain characteristics. Our methodology evaluates the program from the perspective of children's opportunities rather than average treatment effects.

Fiszbein et al. (2009) report that in 1997, only three developing countries (Mexico, Brazil, and Bangladesh) had conditional cash transfer programs in place; by 2008, this number had increased to 29, with many more countries planning to implement such programs. It is important to develop techniques to evaluate the effects of these programs on children's opportunities because these programs are increasingly popular in developing countries, they are sometimes conducted on a large scale, and their focus is on breaking the intergenerational poverty cycle. Despite the recent emergence of substantial empirical literature measuring inequality of opportunity (e.g., Paes et al. 2009 and the references below), no such techniques currently exist.

In the recent literature on equality of opportunity (e.g., Bossert 1995; Fleurbaey 1995, 2008; Roemer 1993), a distinction is generally drawn between two types of factors that influence

the outcome under consideration. On the one hand, there are circumstances and characteristics for which an individual is not responsible, such as race, sex, and parental background; these are the characteristics upon which we condition the cumulative distribution function. On the other hand, there are other characteristics for which individuals are considered responsible, such as having a good work ethic. The idea is that public policies, including conditional cash transfer programs, should compensate for the former while respecting the influence of the latter.<sup>1</sup>

We apply the framework to health outcomes of children aged two to six years. We consider the following circumstances for which parents are not responsible: race, in particular, whether either parent is indigenous; educational level, determined by whether either parent had primary education; and participation in the program. Each possible combination of circumstances corresponds to a “type,” in Roemer’s terminology (Roemer 1993). Therefore, we have eight types. To evaluate the program, we take the health outcomes of children who belong to families enrolled in the program for each of the four types, which are defined on the basis of the parents’ race and education level, and we compare those outcomes with the health outcomes of children whose parents belong to the corresponding type that was not enrolled in the program. Within each type, outcomes can (and will) differ because of factors that are unobserved and ascribed to parental responsibility, such as parental health investments in children. In section II, we argue that an opportunity perspective implies that the comparison of treatment and control types must be based on first- or second-order stochastic dominance.

The idea of using first- or second-order stochastic dominance to investigate equality of opportunity for a particular outcome is not novel. However, until now, this method has been applied only to study whether opportunities are equal within a particular population (see O’Neill et al. 2000 and Lefranc et al. 2009 for studies in which the outcome is income; see Rosa Dias

2009 and Trannoy et al. 2010 for adults' self-assessed health studies; for comparisons between different countries, see Lefranc et al. 2008 for income-based outcomes; for comparisons between regions, see Peragine and Serlenga 2008 for education-based outcomes). Our paper makes three primary contributions to this literature. First, and most important, we conduct our evaluation by establishing the effect of Oportunidades on children's health opportunities. Second, we consider opportunity in the health of young children because their health is crucial for their adult outcomes (see, e.g., Black et al. 2007 and Alderman et al. 2006) and because it is important in its own right. Third, in contrast to previous literature that tested for stochastic dominance in the context of equality of opportunity, our test procedure is based on Davidson and Duclos (2009) and Davidson (2009). Thus, we test the null of nondominance against the alternative of dominance so that rejection of the null logically entails dominance.

Most of the literature on program evaluation focuses on estimating average treatment effects. However, we are interested in establishing or rejecting stochastic dominance between the distributions of health outcomes of children when their parents are either in or out of the program. This exercise is not trivial because we cannot observe the same child both in and out of the program; in other words, we cannot simply resort to a comparison of the cumulative distributions of treatment and control types without making additional assumptions (Heckman 1992). One such assumption is perfect positive quantile dependence (see Heckman et al. 1997), which stipulates that those who are at the  $q$ th quantile in the distribution with treatment would have been at the  $q$ th quantile in the distribution without treatment. Roemer's identification axiom (Roemer 1993) is usually invoked in empirical applications of equality of opportunity when responsibility characteristics are unobserved. This axiom posits that the parents of children who are at the same percentile of their type distribution have exercised comparable responsibility. We

argue below that this axiom provides a normatively inspired alternative to perfect positive quantile dependence by reducing the problem to a comparison of the cumulative distribution functions of the corresponding treatment and control types. The literature on average treatment effects stresses that treatment and control samples must be comparable in terms of preprogram characteristics. We show that this is also imperative when testing for stochastic dominance. Following the literature on average treatment effects, we propose a propensity score matching technique on the basis of preprogram characteristics to better compare treatment and control types. Finally, it is noteworthy that two authors recently suggested incorporating stochastic dominance into project evaluation: Verme (2010) proposed a stochastic dominance approach to determine the effect of a perfectly randomized experiment based on the measures establishing poverty line dominance (i.e., dominance for a range of poverty lines) developed by Foster et al. (1984). Our approach, based on equality of opportunity, stresses that we should focus on the distributions that are conditional on circumstances instead of comparing the distributions of all treatment and control samples. Therefore, we compare the distributions of corresponding treatment and control types. Moreover, our propensity score matching technique makes this approach effective for imperfectly randomized experiments. Naschold and Barrett (2010) allow for nonrandomized treatment by focusing on stochastic dominance between treatment and control samples of the distribution of the difference in outcome, both before and after treatment. They do not focus on types, and the results are difficult to interpret because dominance in terms of differences does not imply that treatment leads to a dominating distribution, which fundamentally depends on who gains and who loses.

Our main findings are that the treatment has substantial positive effects on the health opportunities of children from indigenous families. The effects on children growing up in

nonindigenous families are weaker, although we still find significant positive treatment effects for that group.

The paper is structured as follows. Section I provides definitions and explains the methodology. The data are described in section II. Section III presents the empirical results, including a discussion of the relationship with previous studies. Section IV concludes.

## ***I. DEFINITIONS AND METHODOLOGY***

Let a child's health outcome be represented by the variable  $h \in H = [\underline{h}, \bar{h}] \subseteq \mathbb{R}$ , and let higher values for  $h$  mean better health. A child's health is the result of two types of variables. The first variable,  $c \in C$ , represents circumstances and characteristics for which the child's parents are not responsible, such as race, educational background, and whether the family participates in the program.<sup>2</sup> The second variable,  $r \in R$ , represents characteristics for which parents are responsible, such as health investments in children. Each combination of circumstances corresponds to a type. Social programs should improve children's opportunities, and from the perspective of the equality of opportunity literature, they should compensate for health differences that are caused by circumstances. Moreover, they should respect the influence of parental responsibility, at least to some extent (see, e.g., Swift 2005 for a defense of this position).

In many empirical applications, responsibility is unobserved, as it is here. In such cases, the equality of opportunity framework is usually operationalized using the identification axiom proposed by Roemer (1993), which states that the parents of two children who are at the same percentile of their type distribution of health have exercised identical responsibility.<sup>3</sup> Thus, if the cumulative distribution function of health for a type whose family participated in the program



lies below the cumulative distribution function of health for the corresponding type who did not participate in the program, the type in the program needs less parental effort to obtain a particular level of child health than the type not in the program. If this holds for all levels of health, program participation unambiguously improves the opportunities for this type. Consequently, if the distribution of a type with treatment first-order stochastically dominates the distribution of the corresponding type that did not receive treatment, the program improves this type's opportunities. Similar reasoning applies to second-order stochastic dominance, with the caveat that second-order stochastic dominance can also be obtained by within-type, inequality-reducing transfers of health that do not fully respect the influence of parental responsibility.<sup>4</sup> Roemer's identification axiom does not necessarily imply that we would find children with and without treatment at exactly the same  $q$ th quantile (which is the perfect positive quantile dependence found in Heckman et al. 1997); instead, it merely states that the comparison of the quantiles of the treated and corresponding untreated type is normatively relevant because it compares the health outcomes of children of parents who behaved equally responsibly.

Let  $F^C(h|c)$  denote the conditional distribution of children's health for parents with circumstances  $c$  in the control sample, and let  $F^T(h|c)$  denote the same distribution in the treatment sample. We say that the project improves the opportunities for the health of children with parental circumstances  $c$  if the conditional distribution  $F^T(h|c)$  first-order stochastically dominates the conditional distribution  $F^C(h|c)$ , and we test whether first-order stochastic dominance occurs. Thus, the issue of statistical inference arises. We follow Davidson and Duclos (2009), starting from nondominance as the null hypothesis. To illustrate the procedure for testing first-order dominance and to describe the test more formally, let  $U \subseteq H$  be the union of the

supports of  $F^C(h|c)$  and  $F^T(h|c)$ . We test the null hypothesis of nondominance of  $F^C(h|c)$  by  $F^T(h|c)$ ,

$$\max_{z \in U} (F^T(z|c) - F^C(z|c)) \geq 0,$$

against the alternative hypothesis that  $F^T(h|c)$  first-order stochastically dominates  $F^C(h|c)$ ,

$$\max_{z \in U} (F^T(z|c) - F^C(z|c)) < 0.$$

This approach has the advantage of allowing us to draw the conclusion of dominance if we succeed in rejecting the null hypothesis; in other words, when the null is rejected, the only other possibility is dominance. By contrast, if dominance is the null hypothesis, as is the case in most empirical work to date, failure to reject dominance does not allow us to accept dominance. As Davidson and Duclos (2009) point out, taking nondominance as the null with continuous distributions comes at the cost that it is not possible to reject nondominance in favor of dominance over the entire support of the distribution.<sup>5</sup> Rejecting nondominance is normally possible only over *restricted* ranges of the observed variable. Thus, another merit of this approach is that it allows us to identify the maximal range over the supports of the distribution for which we are able to reject the null of nondominance and, therefore, to accept dominance in favor of the project. In this way, we can check whether we have dominance over ranges of the observed variable that are of special importance, such as the range below minus two standard deviations from the reference height for standardized height, which indicates stunting.

Of course, we must use the identical procedure to test the null of nondominance of  $F^T(h|c)$  by  $F^C(h|c)$  against the alternative hypothesis that  $F^C(h|c)$  dominates  $F^T(h|c)$ . If rejection occurs, we identify the maximal range over the support of the distribution for which we

are able to reject the null of nondominance and to accept dominance against the project.<sup>6</sup> These elements are incorporated in the following weak version of improvements in opportunities, which encompasses most of the work in this paper.

### ***First-Order Improvements***

The project leads to a first-order improvement of the opportunities of children with parental circumstances  $c$  if (i) there exists  $U^0 \subseteq U$  such that we can reject the null of nondominance of  $F^C(h|c)$  by  $F^T(h|c)$  against the alternative that  $F^T(h|c)$  dominates  $F^C(h|c)$  over  $U^0$  and (ii) there exists no  $U^1 \subseteq U$  such that we can reject the null of nondominance of  $F^T(h|c)$  by  $F^C(h|c)$  against the alternative that  $F^C(h|c)$  dominates  $F^T(h|c)$  over  $U^1$ .

Assuming that the influence of parental responsibility on children's health need not be fully respected and that health is cardinally measurable, equalizing health outcomes within type becomes desirable, and it becomes meaningful to ask whether the conditional distribution  $F^T(h|c)$  second-order stochastically dominates the conditional distribution  $F^C(h|c)$ , if the project does not lead to a first-order improvement. Similar statistical issues arise here as for first-order stochastic dominance (see Davidson 2009), leading to the following definition.

### ***Second-Order Improvements***

The project leads to a second-order improvement of the opportunities of children with parental circumstances  $c$  if (i) the project does not lead to a first-order improvement, (ii) there exists  $U^0 \subseteq U$  such that we can reject the null of absence of second-order dominance of  $F^C(h|c)$  by  $F^T(h|c)$  against the alternative that  $F^T(h|c)$  second-order stochastically dominates  $F^C(h|c)$  over  $U^0$ , and (iii) there exists no  $U^1 \subseteq U$  such that we can reject the null of absence of second-order stochastic dominance of  $F^T(h|c)$  by  $F^C(h|c)$  against the alternative that  $F^C(h|c)$  second-order stochastically dominates  $F^T(h|c)$  over  $U^1$ .

Finally, when comparing conditional distribution functions to evaluate a program, it is important to note that inaccurate conclusions may be drawn when preprogram characteristics are not accounted for and when they differ for the treatment types in comparison with the control types (including compensation characteristics). Suppose that we have two sets of characteristics, preprogram characteristics  $x \in X$ , which are not accounted for, and observable circumstances  $c$ . For the type with observed circumstances  $c_1$ , we then have

$$\begin{aligned} F(h|c_1) &= \frac{\int_{\underline{h}}^h f(\tilde{h}, c_1) d\tilde{h}}{f(c_1)} = \frac{\int_X \int_{\underline{h}}^h f(\tilde{h}, c_1, x) d\tilde{h} dx}{f(c_1)} \\ &= \int_X \int_{\underline{h}}^h f(\tilde{h}|c_1, x) \frac{f(c_1, x)}{f(c_1)} d\tilde{h} dx = \int_X F(h|c_1, x) f(x|c_1) dx. \end{aligned}$$

This equation clearly shows that the composition of the  $c_1$  type in terms of  $x$  matters. Indeed, suppose that the treatment has no effect ( $F^C(h|c_1, x) = F^T(h|c_1, x)$ ), but the composition of those with circumstances  $c_1$  differs between the control and treatment types. Suppose that  $f^C(x|c_1)$  is higher than  $f^T(x|c_1)$  for favorable preprogram characteristics  $x$ , or characteristics for which  $F^C(h|c_1, x)$  is lower, and that  $f^C(x|c_1)$  is lower than  $f^T(x|c_1)$  for unfavorable preprogram characteristics. As a result,  $F^C(h|c_1)$  is smaller than  $F^T(h|c_1)$ , and we might erroneously infer that the treatment had an adverse effect on the opportunities of those with circumstances  $c_1$ .

## II. DATA DESCRIPTION

In this section, we describe the Oportunidades program and the construction of treatment and control samples. We describe the selection of circumstances and outcomes and examine the data used to evaluate the program.

### *The Oportunidades program*

The Oportunidades program is a conditional cash transfer program in which bimonthly cash transfers are provided to households in extreme poverty. The cash transfers are conditioned on the attendance of children in school, health care visits for all members of the household, and attendance at information sessions on primary health care and nutrition. Money for schooling constitutes the largest part of the conditional cash transfer. The total amount that a household receives depends on the number, age, and sex of its children. On average, households receive approximately 20 percent of their household consumption from such cash transfers.

Interventions for young children and their mothers are particularly emphasized. Prenatal and postpartum care visits, growth monitoring, immunization, and management of diarrhea and antiparasitic treatments are provided to mothers and young children. Children between the ages of four months and 23 months must have nine periodic medical check ups. From the age of 23 months until the child turns 19 years old, household members must have at least two check ups per year. Children between the ages of six and 23 months, lactating women and low-weight children between the ages of two and four years receive milk-based and micronutrient fortified foods containing the daily recommended intake of zinc, iron, and essential vitamins.<sup>7</sup>

### *Sample design*

The selection of immediate and delayed treatment samples was undertaken in several steps (see, e.g., INSP 2005). Highly deprived localities were identified by using a deprivation index computed on the basis of relevant sociodemographic data available from national censuses. Localities with at least 500 and not more than 2,500 inhabitants, that were categorized as having high or very high deprivation and that had access to an elementary school, a middle school and a health clinic were eligible for treatment. Localities were identified, and a random sample was constructed that was stratified by locality size. Within each state, localities were randomly

assigned into treatment and control groups. A sample of 506 localities was finally selected for the study. A random procedure assigned 320 of these localities to receive immediate treatment; the remaining 186 began receiving treatment approximately 18 months later. In the selected localities, the poverty conditions of all households were evaluated, and households categorized as experiencing extreme poverty were included in the program. This categorization was based on household income, characteristics of the head of household, and variables related to dwelling conditions. Comments by a community assembly on the inclusion and exclusion of households were considered if they met certain criteria to identify beneficiary families. The randomized design enabled us to use the immediate treatment sample as the treatment group and the delayed treatment sample as the control group.<sup>8</sup> However, when we consider the effect of the program on the health outcomes of children between the ages of two and six years in 2003, most of these children grew up in families that were in the program for their entire lives. For children born before the delayed treatment began, this comparison can only show the effect of the difference in exposure when the children were young.<sup>9</sup> Therefore, and because we want to limit our study to an analysis of households that actually received cash transfers (this information is not available for the initial treatment sample), our treatment sample is a subset of the delayed treatment sample.<sup>10</sup> Once the delayed treatment sample began receiving treatment, we had to construct a new control sample, with the intention of making it as similar as possible to the treatment samples (see, e.g., Todd 2004 and Behrman et al. 2006). First, localities that did not meet the criteria for access to an elementary school, a middle school, and a health clinic were excluded. Next, a propensity score method was used that was based on data at the local level as a function of observed characteristics from the 2000 Census that permitted comparison with the localities of the original sample. This procedure led to a selection of 151 localities in which households that

met the criteria for program eligibility were included in the control sample. We compare this control sample to the subset of the delayed treatment sample, as described above.

As we explained at the end of section I, the households in the treatment and control samples must be comparable in terms of preprogram characteristics. There are important problems with the way the control sample was selected.<sup>11</sup> Matching at the local level was performed on the basis of a comparison with observable characteristics in 2000. By this time, the treatment sample had already received treatment. However, matching should have been performed on the basis of characteristics before treatment began. In addition, matching at the local level does not imply matching at the household level (see also Behrman and Todd 1999). Moreover, we do not have data on all children of the households that were in the delayed treatment sample for three reasons (see table A.1 in appendix 1). First, some households dropped out of the sample because of sample attrition. Second, health data were only collected for a subsample of children. Third, because of problems with household identifiers, it was impossible to match all of the children for whom health data were available with only one household each. We only included unique matches in our samples (accounting for more than 80 percent of the children, fortunately). The second and third problems were also present in the control sample. As a result, the treatment and control samples may have differences in terms of preprogram characteristics.

For our empirical strategy in section III, we first use a logistic regression approach to test whether there are statistically significant differences in composition between the treatment and control samples in 1997 for the households with children that were observed in 2003.<sup>12</sup> We use a propensity score matching technique to match the four treatment types with the corresponding control types to correct for possible under- and overrepresentation of households with certain

preprogram characteristics. This technique entails weighted sampling (see appendix 3). We compare the resulting weighted distributions at crucial points (such as standardized height below minus two standard deviations from the reference height, indicating stunting) to establish whether the treatment led to first- or second-order improvements of opportunities for each type by performing stochastic dominance tests on the weighted distribution functions.

### *Circumstances and outcomes*

Ideally, normative theory requires us to obtain a full description of parental circumstances. In reality, an exhaustive description is not available from surveys, and the inclusion of an extensive set of circumstances is statistically unworkable for nonparametric procedures such as ours because of the limited number of observations. For these reasons, we limit ourselves to program participation and two additional circumstances.

The first circumstance refers to parental educational background. In the literature on equality of opportunity, this variable is used most frequently, is always statistically significant, and has been shown to be the most important circumstance in Latin American countries (see, e.g., Bourguignon et al. 2007 and Ferreira and Gignoux 2011). We measure educational background with a dichotomous variable indicating whether at least one parent completed primary education.<sup>13</sup> The second circumstance variable refers to parents' indigenous background. There is substantial literature indicating that indigenous people remain disadvantaged in Mexico (Olaiz et al. 2006; Psacharopoulos and Patrinos 1994; Rivera et al. 2003; SEDESOL 2008). We consider parents to have an indigenous background if at least one of them can speak or understand an indigenous language.

Combining these two binary characteristics with a binary characteristic indicating program participation yields eight types in Roemer's terminology. We partition the samples on the basis of parental indigenous origin (indigenous or nonindigenous) and parental level of



education (primary or less than primary) to form the following types: indigenous, less than primary education (IL); indigenous, primary education (IP); nonindigenous, less than primary education (NL); nonindigenous, primary education (NP). Table 1 shows that there are remarkable differences in the composition of the control sample and the treatment sample among these groups. Clearly, the control sample contains fewer indigenous children and more nonindigenous children with at least one parent who completed primary education than the treatment sample. Because we are comparing cumulative distribution functions of types in the control sample with the corresponding types in the treatment sample, this creates no problem for our analysis. However, as shown in section I, problems arise when there are important differences in terms of preprogram characteristics between the treatment and control types that are compared.

<< insert table 1 about here.>>

We focus on several health outcomes. Two important measures of malnutrition for children are anemia, which is defined as hemoglobin levels lower than 11 grams per deciliter, and stunting, which covers a wider range of nutritional deficiencies and is defined as height for age below minus two standard deviations from the WHO International Growth Reference. The latter implies that in a reference population, approximately 2.3 percent of the population is stunted. As reviewed by Grantham-McGregor and Ani (2001), anemia (iron deficiency) in infancy has been associated with poorer cognition, school achievement, and behavioral problems into middle childhood. Branca and Ferrari (2002) point out that stunting is associated with developmental delay, delayed achievement of developmental milestones (such as walking), later deficiencies in cognitive ability, reduced school performance, increased child morbidity and mortality, higher risk of developing chronic diseases, impaired fat oxidation (stimulating the

development of obesity), small stature later in life, and reduced productivity and chronic poverty in adulthood. In addition to actual stunting, height has a positive effect on completed years of schooling, earnings (see, e.g., Alderman et al. 2006), and cognitive and noncognitive abilities (see, e.g., Case and Paxson 2008 and Schick and Steckel 2010) throughout the distribution. Therefore, we treat our two measures of malnutrition as dichotomous and continuous variables, focusing on the fraction of anemic (stunted) children and on the entire distribution of hemoglobin levels (standardized height). Another health outcome is based on the standardized Body Mass Index (BMI); children are at risk of being overweight if their standardized BMI is larger than 1.15.<sup>14</sup> In a reference population, this cutoff value indicates that 15 percent of children are at risk of being overweight. Overweight children have delayed skill acquisition at young ages (Cawley and Spiess 2008), are more likely to have psychological or psychiatric problems, have increased cardiovascular risk factors, have increased incidence of asthma and diabetes (Reilly et al. 2003), are more likely to be obese as adults (Serdula et al. 1993), and may earn lower wages (Cawley 2004). A final health outcome is based on the number of days parents reported that the child was sick during the previous four-week period. We consider the percentage of children reporting zero days and more than three days. Table 2 provides information on the outcome variables of the control and treatment samples.

<< insert table 2 about here.>>

Considering all households, it is striking that the different entries are similar for all health outcomes in the control and treatment samples, with the exception of the number of days sick; fewer sick days were reported for children in the treatment sample than in the control sample. Approximately one child in four is anemic, and one in three is stunted. Compared with the

reference population, our sample contains far too many stunted children and too many children at risk of being overweight.

Interesting but predictable patterns emerge when considering the distribution of health outcomes over the types.<sup>15</sup> Comparing the IL type with the NL type and the IP type with the NP type, indigenous children have worse health outcomes than nonindigenous children, except for the risk of being overweight in the treatment sample. The differences are substantial, particularly for hemoglobin concentration and standardized height in the control sample. Comparing the IL type with the IP type and the NL type with the NP type, the differences between children who had at least one parent who completed primary education and children whose parents had less than primary education are less obvious. The largest differences occur for standardized height; here having a parent who completed primary education is a clear advantage. Overall, these results are in line with the previous literature (see, e.g., Backstrand et al. 1997; Fernald and Neufeld 2006; González de Cossío et al. 2009; Rivera and Sepúlveda 2003; Rivera et al. 2003).

### ***III. EMPIRICAL RESULTS***

We now use the data described in the previous section to evaluate the Oportunidades program. We show that the treatment and control samples are not comparable in terms of preprogram characteristics, and we apply a propensity score matching technique to make them comparable. We apply the methodology presented in section I on the resulting samples to evaluate the program. We then compare the results to previous studies.

#### ***Comparison of weighted treatment and control types***

As stated at the end of section I, a crucial assumption in the identification of treatment effects on the basis of a simple comparison of the outcomes of treatment and control samples is

that  $f^C(x|c_1) = f^T(x|c_1)$ , implying that the two samples must be similar in terms of preprogram characteristics. If that is the case, after conditioning on  $c_1$ , observing  $x$  does not provide any information about whether an observation belongs to the treatment or control sample. We test this hypothesis as described below.

We construct a sample containing members of both the control and treatment samples. Next, we perform a logistic regression in which the dependent variable takes the value one if the observation belongs to the control sample and the value zero if it belongs to the treatment sample.

Explanatory variables are characteristics of the family, characteristics of the family's dwelling, family assets, and state of residence (see appendix 2 for more details). These characteristics were measured in 1997, before the program started.<sup>16</sup> The results are reported in table A.2 in appendix 2. We find that many of the characteristics significantly affect the probability that the observation comes from the control sample, indicating that the hypothesis that treatment and control samples are comparable in terms of the composition of their preprogram characteristics must be rejected.

In the identification of average treatment effects, a standard way to address differences in the composition of the treatment and control samples is to use propensity score matching techniques. The goal is to make the treatment and control samples more comparable by weighting different observations based on the estimated probability that the observation belongs to the control sample, as determined by the logistic regression discussed in the previous paragraph. Appendix 3 explains this procedure and how the weighting is used to obtain estimates of the relevant distribution functions. The weighting procedure has a substantial effect on the Roemer motivation for considering cumulative distribution functions (Roemer's identification

axiom), as we discuss in appendix S2.<sup>17</sup> Appendix S3 provides the equivalent of table 2 for the weighted (matched) samples. Supplemental appendices S2 and S3 are available at <http://wber.oxfordjournals.org/>.

In table 3, we use the weighted samples to consider the effect of the treatments on the fraction of children who are anemic, stunted, or at risk of being overweight. We use the same samples to examine the fraction of children for whom zero sick days or more than three sick days during the previous four weeks were reported. Effects that are statistically significantly different from zero at the 5 percent level of significance are indicated by “\*\*,” and effects that are statistically significantly different from zero at the 10 percent level of significance are indicated by one “\*.” Each entry provides the effect of the treatment. From an opportunity perspective, a desirable effect on these fractions indicates that less responsibility allows parents to prevent their children from being anemic, stunted, at risk of being overweight, or sick for more than three days in the previous four-week period.

<< insert table 3 about here.>>

We see that the treatment effects reported in table 3 are substantial, and all significant effects of the program are in a desirable direction. For each health indicator, we find at least one significant desirable treatment effect for one of the types. The table suggests that the program works well, particularly for children of indigenous origin without a parent who completed primary education. This type is likely to be the most disadvantaged, as table 2 suggests.

Children of indigenous origin with a parent who completed primary education have an improvement in all indicators, although the effects are only significant for the fraction of anemic and stunted children. For nonindigenous children, the results are less obvious. The fraction of

nonindigenous children who are anemic decreases because of the program, but the results presented in table 3 identify no other significant treatment effects for nonindigenous children.

Figure 1 presents the results of the stochastic dominance tests, using the procedure explained in section I.<sup>18</sup> The horizontal axis denotes the numerical value of the variable of interest (hemoglobin concentration, standardized height, standardized BMI, and reported days sick).

The black (grey) boxes depict the maximal range over the support of the distributions for which the null of nondominance is rejected at the 5 percent level of significance in favor of a desirable (undesirable) effect of the treatment. Hatched (white) boxes indicate the same at a significance level of 10 percent. When hatched (white) boxes are adjacent to a black (grey) box, they show how far the rejection range of the null can be extended for the 10 percent level of significance. Each row contains an acronym “XYi,” of which the first two characters, “XY”, indicate the name of the types that are compared (XY = IL, IP, NL, or NP), and the character “i” indicates whether the test refers to first- (i = 1) or second- (i = 2) order stochastic dominance. The numbers in parentheses behind the boxes show the percentage of observations of the treated type within the black or grey (hatched or white) box.

<<insert figure 1 about here.>>

For example, in the top left panel of figure 1, the hatched box labeled “IL1” shows that, using a 10 percent level of significance, the null hypothesis that the cumulative distribution of the treatment type does not first-order stochastically dominate the distribution of the control type must be rejected against the alternative, that the distribution of the treatment type first-order stochastically dominates the distribution of the control type over the range [7.5, 11.2], which contains 35.5 percent of the treated type. The hypothesis of nondominance can only be rejected

at the 10 percent level of significance. Thus, we tested the null hypothesis of the absence of second-order stochastic dominance in favor of the treatment against the alternative, that the distribution of the treatment type second-order stochastically dominates the distribution of the control type at the 5 percent level of significance. We failed to reject the null, such that no box “IL2” is drawn. For IP types, the black box labeled “IP1” indicates that the null hypothesis of nondominance can be rejected at the 5 percent level of significance over the range [8.1, 14.5], which contains 97 percent of the treated IP type. When we increase the level of significance to 10 percent, the hatched box shows that the rejection interval enlarges only marginally, to [8.0, 14.5]. For NL types, when testing for first-order stochastic dominance, we find a white box over the small range of [9.7, 9.9] with very few observations of the treatment type and a solid black box further up in the distribution. When testing NL types for second-order stochastic dominance, we find a small white box. On balance, the evidence for this type against treatment is not strong. Finally, for NP types, we have first a solid black and then a white box. The latter is only significant at the 10 percent level of significance and occurs at a less important part of the distribution (above 11, when children are no longer anemic). When testing for second-order stochastic dominance, we see a solid black box labeled “NP2,” indicating that the project leads to second-order improvement,<sup>19</sup> and this type is also positively affected by the program.

The other panels in figure 1 can be similarly interpreted. In the top right panel, we see that the treatment leads to first-order improvements in the standardized height for IL and IP types over large and crucial parts of the support (standardized height below minus two standard deviations from the reference height). For NL types, we find a first-order stochastic dominance effect in favor of the treatment in an important part of the distribution (standardized height below minus two standard deviations from the reference height) and an adverse effect higher up in the

distribution. There is evidence of a marginal perverse first-order treatment effect at a significance level of 10 percent on standardized height for NP types over a small range of  $[-2.11, -2.00]$ , which contains only 3 percent of the observations of the treated type, and a positive effect higher up in the distribution. No second-order stochastic dominance effects can be established for the nonindigenous types. In the bottom left panel, we concentrate on what occurs at the right of the dotted vertical line, which represents children at risk of being overweight. We see positive, first-order stochastic dominance effects at the 5 percent level of significance for IL types and some evidence of marginally significant perverse treatment effects for IP and NP types. The bottom right panel shows first-order improvements for IL, NL, and NP types. The intervals reported here, except for IL, contain few observations, because of the high frequency of zero reported sick days (see table 2).

The results reported in table 3 and figure 1 are consistent. The stochastic dominance results provide more detail and identify effects in important parts of the distribution that would otherwise go unnoticed, such as the positive first-order stochastic dominance effect on standardized height for NL children. If first-order improvements cannot be found and the influence of parental responsibility is not to be fully respected, then second-order stochastic dominance provides a way to determine whether the program has positive effects. Second-order improvements occur only once in our application, for the hemoglobin concentration of NP types. In summary, we find strong evidence of positive treatment effects for children of indigenous origin, particularly for those without a parent who completed primary education. The evidence for children from nonindigenous origin is not as strong, but enrollment in the program also seems to have positive effects on health opportunities for these children, on balance.



### *Comparison to previous studies*

Diaz and Handa (2006) use propensity score matching techniques to construct alternative control samples from the Mexican national household survey. They compute average treatment effects by comparing the immediate treatment sample after eight months of receiving program benefits with the delayed treatment sample (who had not yet received benefits), on the one hand, and their newly constructed control samples, on the other. They conclude, “The PSM [propensity score matching] technique requires an extremely rich set of covariates, detailed knowledge of the beneficiary selection process, and the outcomes of interest need to be measured as comparably as possible in order to produce viable estimates of impact” (p.341). In our case, the outcomes are measured in identical ways in the delayed treatment and control samples, and the control sample is constructed following the beneficiary selection process as closely as possible. Our selection of covariates for the propensity score matching closely follows Behrman et al. (2009b), who use almost identical covariates in comparing the effects on schooling outcomes of the short-run differential exposure (between the immediate and delayed treatment samples) with the long-run differential exposure (between the immediate treatment and control samples). They find that longer exposure produces larger effects, and the differences between the order of magnitude of the short- and long-run effects are reasonable. This finding suggests that the propensity score matching technique we use can produce reliable estimates of average treatment effects.

The interpretation of the difference between the distributions of the weighted treatment and control samples as a treatment effect depends on the extent to which the weighting procedure manages to correct for possibly unobserved heterogeneity caused by the imperfect randomness of the assignment to treatment and control groups. Of course, it is not possible to test this directly, but we can compare our results to the findings in the literature that consider differences in

children's health outcomes between immediate and delayed treatment samples. Rivera et al. (2004) compare the health outcomes of children younger than 12 months old in 1997. They find that in 1999 after 12 months of treatment, children in the immediate treatment sample had higher mean hemoglobin values than the children from the delayed treatment sample, who were untreated up to that point. After the immediate treatment sample had received 24 months of treatment and the delayed treatment sample had received approximately six months of treatment, children from the immediate treatment sample had grown more than children in the delayed treatment sample, and the differences in height were significantly larger for households with low socioeconomic status (a score based on dwelling characteristics, possession of durable goods, and access to water and sanitation). Gertler (2004) finds similar results for children aged 0 to 35 months in 1997, stating that "treatment children were 25.3 percent less likely to be anemic and grew about 1 centimeter more during the first year of the program" (p. 340). Both of these differences are statistically significant at the 1 percent level. Unfortunately, Gertler does not report whether the effect differs for different subgroups, such as our types. Hemoglobin levels, unlike height, were not observed before the program started. Therefore, the results for hemoglobin levels do not control for child fixed effects as opposed to growth effects, as noted by Behrman and Hoddinott (2005). They investigate the effect on the height of children who were between 4 and 48 months of age when treatment began in August 1998. They find that when child fixed effects are not included, treatment has a significant negative effect on child height for children between 4 and 36 months of age. However, if child fixed effects are controlled (by considering the difference between 1999 and 1998), the treatment effect becomes significantly positive at approximately one centimeter, as in Gertler (2004).<sup>20</sup> Notably, program effects are

larger for children in households in which the head of the household speaks an indigenous language and the mother is more educated.<sup>21</sup>

Finally, Fernald et al. (2008) use a different approach. They combine the data of both the immediate and delayed treatment samples to estimate the effect of the size of the conditional cash transfer received on children between 24 and 68 months of age in 2003, when the children's height was measured. Increasing the size of the transfer leads to higher height-for-age scores, a lower prevalence of stunting and a lower prevalence of obesity. Parental level of education and whether the head of the household spoke an indigenous language were not significant controls in their model.

Overall, these findings are in line with ours. The program has significant positive effects on children's height and hemoglobin concentration levels. Larger effects tend to be found for households in which an indigenous language is spoken. This finding is compatible with Fernald et al. (2008) because, in general, indigenous families receive larger cash transfers than nonindigenous families based on the finding that they tend to have more children. Our results indicate where in the distribution the program is most effective for the different types, and we can see that the program is most powerful for the most disadvantaged types, children of indigenous origin.

#### ***IV. CONCLUSION***

There is a growing body of literature on the measurement of inequality of opportunity (for an overview, see, e.g., Ramos and Van de gaer 2012). Thus far, the ideas in the literature have not been applied to evaluate social programs. We propose a methodology to do so.

We bring together insights from the literature on equality of opportunity, the literature on program evaluation, and the literature on testing for stochastic dominance. Roemer's (1993)

normative approach to equality of opportunity indicates that we should focus on types and that, if responsibility characteristics are unobserved, individuals at the same percentile of the distribution of the outcome within their type have exercised a comparable degree of responsibility. This approach provides a normative foundation for the comparison of cumulative distribution functions of corresponding treatment and control types. The literature on program evaluation stresses that care should be taken to ensure that the treatment and control samples are comparable in terms of preprogram characteristics. If they are not, propensity score matching techniques can be used to make the samples more comparable. Hence, we test whether the treatment and control samples are comparable in terms of preprogram characteristics and since the test fails, we propose a weighted sampling method based on standard propensity score matching techniques to make the treatment and control types comparable. Finally, Davidson and Duclos (2009) and Davidson (2009) propose a new technique to test for stochastic dominance, taking nondominance as the null so that rejection of the null implies dominance. Their test procedure is particularly suited to our study because it allows us to see where dominance can be established along the distribution.

We applied our procedure to study the effect of the Mexican Oportunidades program on children's health opportunities. We can draw two conclusions about the proposed methodology. First, in our application (as in the applications by Lefranc et al. 2008, Lefranc et al. 2009, Peragine and Serlenga 2008, and Rosa Dias 2009), looking for second-order stochastic dominance does not significantly add to the conclusions drawn from first-order stochastic dominance. Thus, whether the influence of parental responsibility is to be fully respected does not substantially affect the conclusions. Second, the treatment and control samples differed substantially in terms of preprogram characteristics. Therefore, it is important to use weighted

sampling based on techniques such as propensity score matching to make the samples (more) comparable. Concerning the actual effects of the program, our results indicate that the Oportunidades program has a substantially favorable effect on the health opportunities of the most disadvantaged children, that is, those with parents of indigenous origin and without a parent who completed primary education. Additionally, the effects on children of indigenous origin with a parent who completed primary education are sizable and important. The effects on nonindigenous children are less obvious, but the overall evidence in this paper indicates that the program also results in better health opportunities for these children

## *APPENDICES*

### *APPENDIX 1. Sampling Procedure*

<< insert table A.1 here.>>

When we compare the sample sizes in the column “1997 data available” with the sizes in table 1 in the main text, we see that 12 (three) observations dropped out in the final control (treatment) sample because of missing observations on circumstances.

### *APPENDIX 2. Results of the logistic regression*

Our specification for the logistic regression is close to the specification used for propensity score matching by Behrman et al. (2009b) and Behrman and Parker (2010). The dependent variable equals one if the observation comes from the control sample and zero otherwise. Explanatory variables are based on preprogram characteristics of the treatment sample and the 1997 recall characteristics of the control sample. We have five types of explanatory variables:

- (1) Household characteristics, which include the ages of the head of the household and spouse (in years); the sex of the head of the household; whether the head of the household and spouse speak an indigenous language; whether the parents completed primary education; whether the parents work; and the composition of the household (number of children and women and men of different ages)
- (2) Dwelling conditions of the household, which include the number of rooms in the house and a list of dummy variables indicating the presence of electric light, running

water on the property, running water in the house (which implies the presence of running water on the property), a dirt floor, and whether the roof and walls are of poor quality

(3) Asset information, which includes dummy variables indicating whether the family owns animals or land and whether the family possesses a blender, refrigerator, fan, gas stove, gas heater, radio, stereo, TV, video, washing machine, car, or truck

(4) State of residence, which includes a list of dummy variables indicating the state in which the family lives, with the reference state (all state of residence dummies equal to zero) of Veracruz

(5) Dummy variables for missing characteristics whose effects could be meaningfully estimated, following Behrman et al. (2009b) and Behrman and Parker (2010); the variable “Miss Asset” takes the value of one if any of the assets listed in the table between “Animals” and “Truck” is missing

Table A.2 gives the estimated coefficients.

<< insert table A.2 about here.>>

### ***APPENDIX 3. Matching estimator and construction of the corresponding distribution function.***

<< insert table A.3 about here.>>

#### **Step 1: Propensity score matching**

The estimated logistic regressions allow us to compute, for each observation, the propensity score  $P_i$ , the probability that the observation is in the control sample given its preprogram characteristics  $x_i$ . Figure A.1 depicts the estimated propensity scores because we

matched the treatment into the control sample for each of the four combinations of race and parental level of education, and we determined the common support for each of these four comparisons as the overlap of the support of the control and treatment samples. Table A.3 above gives the common support and the number of observations in the common support for each of the types.

We tested the balancing property score using Stata. The optimal number of blocks was 11, and we had 54 explanatory variables, resulting in 594 tests. In 14 cases, the balancing property was rejected. As an additional test, we reran the logistic equation from table A.2 using the weighted sample. Only four coefficients out of 54 were significant. These results are encouraging.

#### Step 2: Construction of the cumulative distribution function

Let  $I_1$  denote the set of individuals in the treatment sample,  $I_0$  denote the set of individuals in the control sample, and  $S_p$  denote the region of common support. The number  $n_0$  gives the number of individuals in the set  $I_0 \cap S_p$ . The outcome of individual  $j$  in the control sample is  $Y_{0j}$ , and the outcome of individual  $i$  in the treatment sample is  $Y_{1i}$ . Let  $D = 1$  for program participants and  $D = 0$  for those who do not participate in the program.

The purpose is to match each individual in the control sample with a weighted average of individuals in the treatment sample. The usual estimator of the average treatment effect thus becomes

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_p} [E(Y_{1j} | D = 1, P_j) - Y_{0j}],$$

with  $E(Y_{1j} | D = 1, P_j) = \sum_{i \in I_1} W(i, j) Y_{1i}$ .



The construct  $E(Y_{1j}|D = 1, P_j)$  is the outcome of the hypothetical individual matched to individual  $j$ . The average treatment effect can be written as

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} \sum_{i \in I_1} W(i, j) Y_{1i} - \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} Y_{0j}.$$

The first term is the average of the matched observations, which attaches to each of the original observations  $Y_{1i}$  a weight

$$\omega_i = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} W(i, j).$$

It is therefore natural (and consistent with the standard model of the estimation of average treatment effects) to use for each observation  $Y_{1i}$  the weight  $\omega_i$  to construct the cumulative distribution function.

Many possible ways exist to determine the weights  $W(i, j)$ . We use a Kernel estimator, such that

$$W(i, j) = \frac{G\left(\frac{P_i - P_j}{\alpha}\right)}{\sum_{k \in I_1} G\left(\frac{P_k - P_j}{\alpha}\right)},$$

where  $G(\cdot)$  is the Epanechnikov kernel function and  $\alpha$  is a bandwidth parameter. The bandwidth parameter was chosen in an optimal way using the formula in Silverman (1986,45–47):

$$\alpha = 1.06 \min\left(\sigma, \frac{\rho}{1.34}\right),$$

where  $\sigma$  is the standard deviation and  $\rho$  is the interquartile range of the distribution of propensity scores. The resulting bandwidths for each of the types are given in the last column of table A.3.

<< insert figure A.1 about here.>>

## REFERENCES

- Alderman, Harold, John Hoddinott, and Bill Kinsey. 2006. "Long-term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58 (3): 450–74.
- Backstrand, Jeffrey R., Lindsay H. Allen, Gretel H. Pelto, and Adolfo Chávez. 1997. "Examining the Gender Gap in Nutrition: An Example from Rural Mexico." *Social Science & Medicine* 44 (11): 1751–9.
- Behrman, Jere R., and John Hoddinott. 2005. "Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition." *Oxford Bulletin of Economics and Statistics* 67 (4): 547–69.
- Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five Year Follow-up of PROGRESA/Oportunidades." *Journal of Human Resources* 46: 93-122.
- Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2009a. "Medium-Term Impact of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico." In *Poverty, Inequality and Policy in Latin America*, ed. S. Klasen and F. Nowak-Lehmann, 219–70. Cambridge: MIT Press.
- Behrman, Jere R., Susan W. Parker, and Petra E. Todd. 2009b. "Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico." *Economic Development and Cultural Change* 57 (3): 439–77.
- Behrman, Jere R., Piyali Sengupta, and Petra E. Todd. 2005. "Progressing through PROGRESA: an Impact Assessment of a School Subsidy Experiment in Rural Mexico." *Economic Development and Cultural Change* 54 (1): 237–75.

- Behrman, Jere R., and Petra E. Todd. 1999. *Randomness in the experimental samples of PROGRESA –Education, Health, and Nutrition Program*. International Food Policy Research Institute.
- Behrman, Jere R., Petra E. Todd, Bernardo Hernández, José Urquieta, Orazio Attanasio, Manuela Angelucci, and Mauricio Hernández. 2006. *Evaluación externa de impacto del programa Oportunidades 2006*. Instituto Nacional de Salud Pública.
- Black, Sandra E., Paul Devereux, and Kjell Salvanes. 2007. “From the cradle to the labor market? The Effect of Birth Weight on Adult Outcomes.” *The Quarterly Journal of Economics* 122 (1): 409–39.
- Bossert, Walter. 1995. “Redistribution Mechanisms Based on Individual Characteristics.” *Mathematical Social Sciences* 29 (1): 1–17.
- Bourguignon, François, Francisco H.G. Ferreira, and Marta Menéndez. 2007. “Inequality of Opportunity in Brazil.” *Review of Income and Wealth* 53 (4): 585–618.
- Branca, Francesco, and Marika Ferrari. 2002. “Impact of Micronutrient Deficiencies on Growth: The Stunting Syndrome.” *Annals of Nutrition and Metabolism* 46 (Suppl. 1): 8–17.
- Case, Anne, and Christina Paxson. 2008. “Stature and Status: Height, Ability and Labor Market Outcomes.” *Journal of Political Economy* 116 (3): 499–532.
- Cawley, John. 2004. “The Impact of Obesity on Wages.” *Journal of Human Resources* 39 (2): 451–74.
- Cawley, John, and C. Katharina Spiess. 2008. “Obesity and Skill Attainment in Early Childhood.” *Economics and Human Biology* 6: 388–97.

- Chen, Wen-Hao, and Jean-Yves Duclos. 2008. *Testing for Poverty Dominance: An Application to Canada*. IZA Discussion Paper N 2829.
- Davidson, Russell. 2009. "Testing for Restricted Stochastic Dominance: Some Further Results." *Review of Economic Analysis* 1 (1): 34–59.
- Davidson, Russell, and Jean-Yves Duclos. 2009. *Testing for Restricted Stochastic Dominance*. GREQAM Document de Travail 2009-38 (06-09).
- Diaz, Juan José, and Sudhanshu Handa. 2006. "An assessment of Propensity Score Matching as a Nonexperimental Impact Estimator: Evidence from Mexico's PROGRESA Program." *Journal of Human Resources* 41 (2): 319–45.
- Fernald, Lia C.H., Paul J. Gertler, and Lynnette M. Neufeld. 2008. "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's Oportunidades." *The Lancet* 371 (9615): 828–37.
- Fernald, Lia C.H., and Lynnette M. Neufeld. 2006. "Overweight with Concurrent Stunting in Very Young Children from Rural Mexico: Prevalence and Associated Factors." *European Journal of Clinical Nutrition* 61 (5): 623–32.
- Ferreira, Francisco. H. G., and Jérémie Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income and Wealth* 57(4): 622-54.
- Fiszbein, Ariel, Norbert Schady, Francisco H.G. Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty, a World Bank policy research report*. The World Bank, Washington.

- Fleurbaey, Marc. 1995. "The Requisites of Equal Opportunity." In *Social Choice, Welfare and Ethics*, ed. M. Salles and N. Schofield, 37–53. Cambridge University Press.
- Fleurbaey, Marc. 1998. "Equality among responsible individuals." In *Freedom in Economics: New Perspectives in Normative Economics*, ed. J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy, 206–234. London: Routledge.
- Fleurbaey, Marc. 2008. *Fainness, Responsibility and Welfare*. Oxford: Oxford University Press.
- Foster, James, Joel Greer, and Erik Thorbeke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–66.
- Gertler, Paul J. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *American Economic Review* 94 (2): 336–41.
- González de Cossío, Teresa, Juan A. Rivera, Dinorah González Castell, Mishel Unar Munguía, and Eric A. Monterrubio. 2009. "Child Malnutrition in Mexico in the Last Two Decades: Prevalence using the New WHO 2006 Growth Standards." *Salud Pública de México* 51 (Supp 4): S494-S506.
- Grantham-McGregor, Sally, and Cornelius Ani. 2001. "A Review of Studies on the Effect of Iron Deficiency on Cognitive Development in Children." *The Journal of Nutrition* 131 (2): 649S –68S.
- Heckman, James J. 1992. "Randomization and social policy evaluation." In *Evaluating Welfare and Training Programs*, ed. C. Manski and I. Garfinkel, 201–230. Cambridge: Harvard University Press.

- Heckman, James J., Jeffrey Smith, and Nancy Clements. 1997. "Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts." *Review of Economic Studies* 64 (4): 487–535.
- INSP. 2005. *General Rural Methodology Note*. Instituto Nacional de Salud Pública. Cuernavaca, Mexico. INSP2005.
- Lefranc, Arnaud, Nicolas Pistolessi, and Alain Trannoy. 2008. "Inequality of Opportunities vs. Inequality of Outcomes: Are Western Societies All Alike?" *Review of Income and Wealth* 54 (4): 513–46.
- Lefranc, Arnaud, Nicolas Pistolessi, and Alain Trannoy. 2009. "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France." *Journal of Public Economics* 93 (11-12): 1189–1207.
- Naschold, Felix, and Christopher B. Barrett. 2010. *A Stochastic Dominance Approach to Program Evaluation with an Application to Child Nutritional Status in Kenya*. Working Paper.
- Olaiz, Gustavo, Juan A. Rivera, Teresa Shamah, Rosalba Rojas, Salvador Villalpando, Mauricio Hernández, and Jaime Sepúlveda. 2006. *Encuesta Nacional de Salud y Nutrición 2006* [National Health and Nutrition Survey 2006]. Instituto Nacional de Salud Pública.
- O'Neill, Donal, Olive Sweetman, and Dirk Van de gaer. 2000. "Equality of Opportunity and Kernel Density Estimation: An Application to Intergenerational Mobility." In *Advances in Econometrics*, Volume 14, ed. T. Fomby and R. C. Hill, 259–274. Stanford: JAI Press.

- Parker, Susan W., Luis Rubalcava, and Graciela Teruel. 2008. "Evaluating Conditional Schooling and Health Programs." In *Handbook of Development Economics*, Volume 4, ed. T. Schultz and J. Strauss, 3963–4035. Elsevier.
- Paes de Barros, Ricardo, Francisco H.G. Ferreira, José R. Molinas Vega, and Jaime Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunities in Latin America and the Caribbean*. The World Bank.
- Peragine, Vito, and Laura Serlenga. 2008. "Higher education and equality of opportunity in Italy." In *Inequality of opportunity: papers from the Second ECINEQ Society Meeting, Research on Economic Inequality*, Volume 16, ed. J. Bishop and B. Zheng, 67–97. Bingley: Emerald Group Publishing.
- Psacharopoulos, George, and Harry A. Patrinos. 1994. *Indigenous People and Poverty in Latin America: An Empirical Analysis*. Washington DC: The World Bank.
- Ramos, Xavi, and Dirk Van de gaer. 2012. *Empirical Approaches to Inequality of Opportunity: Principles, Measures and Evidence*. FEB Working Paper 12/792. Ghent: Faculty of Economics and Business Administration, Ghent University.
- Reilly, John J., E. Methven, Zoe C. McDowell, Belinda Hacking, D. Alexander, Laura Stewart, and Christopher J.H. Kelnar. 2003. "Health Consequences of Obesity." *Archives of Disease in Childhood* 88 (9): 748–52.
- Rivera, Juan A., Eric Monterrubio, Teresa González-Cossío, Raquel García-Feregrino, Armando García-Guerra, and Jaime Sepúlveda. 2003. "Nutritional Status of Indigenous Children Younger than Five Years of Age in Mexico: Results of a National Probabilistic Survey." *Salud Pública de México* 45: S466–76.

- Rivera, Juan A., and Jaime Sepúlveda. 2003. "Conclusions from the Mexican National Nutrition Survey 1999: Translating Results into Nutrition Policy." *Salud Pública de México* 45: S565–75.
- Rivera, Juan A., Daniela Sotres-Alvarez, Jean-Pierre Habicht, Teresa Shamah, and Salvador Villalpando. 2004. "Impact of the Mexican Program for Education, Health, and Nutrition (PROGRESA) on Rates of Growth and Anemia in Infants and Young Children." *The Journal of the American Medical Association* 291 (21): 2563–70.
- Roemer, John. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs* 22 (2): 146–66.
- Roemer, John. 1998. *Equality of Opportunity*. Cambridge MA: Harvard University Press.
- Rosa Dias, Pedro. 2009. "Inequality of Opportunity in Health: Evidence from a UK Cohort Study." *Health Economics* 18 (9): 1057–74.
- Schick, Andreas, and Richard H. Steckel. 2010. *Height as a Proxy for Cognitive and Non-Cognitive Ability*. NBER Working Paper N 16570 .
- Schultz, T. Paul. 2004. "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program." *Journal of Development Economics* 74 (1): 199–250.
- SEDESOL. 2008. *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007)*. Ministry of Social Development of Mexico (SEDESOL).
- Serdula, Mary K., Donna Ivery, Ralph J. Coates, David S. Freedman, David F. Williamson, and Tim Byers. 1993. "Do Obese Children become obese Adults? A Review of the Literature." *Preventive Medicine* 22: 167–77.



- Silverman, Bernard. W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRH.
- Swift, Adam. 2005. “Justice, Luck, and the Family: The Intergenerational Transmission of Economic Advantage from a Normative Perspective.” In *Unequal chances: family background and economic success*, ed. S. Bowles, H. Gintis, and M. Osborne Groves, 256–76. Princeton University Press.
- Todd, Petra E. 2004. *Design of the Evaluation and Method used to Select Comparison Group Localities for the Six Year Follow-Up Evaluation of Oportunidades in Rural Areas*. Technical report, International Food Policy Research Institute.
- Trannoy, Alain, Sandy Tubeuf, Florence Jusot, and Marion Devaux. 2010. “Inequality of Opportunities in Health in France: A First Pass.” *Health Economics* 19 (8): 921–38.
- Verme, Paolo. 2010. “Stochastic Dominance, Poverty and the Treatment Effect Curve.” *Economics Bulletin* 30 (1): 365–73.

Dirk Van de gaer (corresponding author) is Professor in Economics, Vakgroep Sociale Economie and SHERPPA, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Associate Fellow at Université Catholique de Louvain, CORE, B-1348, Louvain-la- Neuve, Belgium. The research was completed while he was visiting IAE - CSIC, Campus UAB, 08193 - Bellaterra, Barcelona, Spain. Tel: +32-(0)9-2643490. Fax: +32-(0)9-2648996. E-mail: [Dirk.Vandegaer@ugent.be](mailto:Dirk.Vandegaer@ugent.be).

Joost Vandebossche is a PhD student in Economics, SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Aspirant FWO - Flanders. E-mail: [Joost.Vandebossche@UGent.be](mailto:Joost.Vandebossche@UGent.be).

José Luis Figueroa is a PhD student in Economics, SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and CES, Katholieke Universiteit Leuven. E-mail: [joseluis.figueroaoropeza@ugent.be](mailto:joseluis.figueroaoropeza@ugent.be).

This work was supported by the Belgian Program on Inter University Poles of Attraction, initiated by the Belgian State, Prime Minister's Office, Science Policy Programming [Contract No. P6/07] and by the FWO Flanders, project number 3G079112. We thank the editor, two referees, Bart Cockx, Aitor Calo Blanco, Gaston Yalonetzky, Alain Trannoy, Stefan Dercon, Francisco Ferreira, Vito Peragine, and Nicolas Van de Sijpe for many useful comments and suggestions and Jean-Yves Duclos for showing us how to incorporate the survey design into the bootstrap procedure. We gratefully acknowledge comments received on preliminary versions presented at the GREQAM-IDEF workshop "The Multiple Dimensions of Equality and Fairness" (Marseilles, France, November 17, 2010), the OPHI workshop "Inequalities of

Opportunities” (Oxford, UK, November 22–23, 2010), the UAB workshop “Equality of Opportunity and Intergenerational Mobility” (Barcelona, Spain, December 17, 2010), the winter school on “Inequality and Social Welfare Theory” (Canazei, Italy, January 10–13, 2011), the faculty seminar in Caen (France, March 28, 2011), the workshop “Equity in Health” (Louvain la Neuve, Belgium, May 11–13, 2011), the ABCDE conference (Paris, France, May 30–June 01, 2011), the conference “Mind the Gap: from Evidence to Policy” (Cuernavaca, Mexico, June 15–17, 2011), the conference “Micro Evidence on Innovation in Developing Countries” (San Jose, Costa Rica, June 27–28, 2011), the ECINEQ conference (Catania, Italy, July 18–20, 2011) and the EEA conference (Oslo, Norway, August 25–29, 2011). A supplemental appendix to this paper is available at <http://wber.oxfordjournals.org/>.

1 Recently, Lefranc et al. (2009) extended this framework with a third factor, random factors that are legitimate sources of inequality “as long as they affect individual outcomes and circumstances in a neutral way” (p. 1192).

2 Race and educational background are circumstances because they should not influence the health opportunities parents can obtain for their children. Whether the family participates in the program is largely determined by the locality in which they lived at the time the program began; therefore, this is outside of parental control.

3 See Roemer (1993) and Roemer (1998) for a defense of this principle, and see Fleurbaey (1998) for a discussion of the assumptions involved.

4 Fully respecting the influence of responsibility means that the health differences caused by responsibility are fully preserved by the program. Alternative notions of responsibility are weaker and require, for instance, that the program does not change the rank order of children’s health. This weaker requirement is compatible with second-order stochastic dominance.

5 Let  $\underline{h}$  be the lower bound of  $U$ . Evidently,  $F^T(\underline{h}|c) - F^C(\underline{h}|c) = 0$ ; therefore, the maximum over  $U$  is never less than zero. Moreover, close to the boundaries of the support, there may be too little information to reject nondominance.

6 Supplemental appendix S1 contains more details about stochastic dominance tests. The appendix is available at <http://wber.oxfordjournals.org/>.

7 These supplements may also be given to children in households that are not receiving treatment (including children in the control sample) if signs of malnutrition are detected. This may lead to a downward bias of the estimated effect of Oportunidades (see also Behrman et al. 2009b, footnote 8).

8 Most studies focus on a comparison of the immediate and delayed treatment samples and therefore evaluate the effect of differences in duration of program participation; see, e.g., Schultz (2004), Behrman et al. (2005), or Behrman et al. (2009a).

9 In the working paper version, we repeat the analysis for children born after April 1998 (when the original treatment started) and before October 1999 (when delayed treatment started), taking the original treatment sample as the treatment sample and the delayed treatment sample as the control. The program effects are less clearly shown, but some positive treatment effects remain; see also note 21.

10 Sensitivity analysis (reported in the working paper version, available at [http://www.feb.ugent.be/nl/Ondz/WP/Papers/wp\\_11\\_749.pdf](http://www.feb.ugent.be/nl/Ondz/WP/Papers/wp_11_749.pdf)) shows that the results are similar when we compare the entire delayed treatment sample (including those for which no positive transfers were reported) and the control sample.

11 This may explain why the control sample has rarely been used in academic papers. Recently, however, matched sampling was used to compare schooling (Behrman et al. 2009b and

Behrman et al. 2010) and work outcomes (Behrman et al. 2010) in immediate treatment, delayed treatment, and control samples.

12 In 2003, in addition to the regular household data, an additional questionnaire with recall data was collected. The purpose of these retrospective questions was to compare the preprogram characteristics for the treatment samples with the new control sample.

13 In the working paper version, we report the results when parental background is measured on the basis of mother's education only. The results are similar to the ones we present here.

14 The incidence of underweightness is lower than in a reference population.

15 The types may differ in terms of characteristics that do not enter the definition of type and in terms of preprogram characteristics.

16 For the control sample, this is based on recall data (see also note 12).

17 Because health is also influenced by preprogram characteristics, we can no longer infer from the percentile in the distribution of health for each type the corresponding responsibility; the same percentile will be obtained by people with different combinations of responsibility and preprogram characteristics. In the supplemental appendix S2, we show that, under certain assumptions, the weighting procedure guarantees that individuals at the same percentile in the weighted treatment and the control sample have identical expected responsibility.

18 Because of the many zero observations, this test procedure cannot be used for the number of days sick. Here, the stochastic dominance test is based on a standard test for the difference between the cumulative distribution functions at the natural numbers between 0 and

30. The intervals shown for this health outcome connect the points in the support where the difference between the cumulative distribution functions is statistically significant.

19 Observe that the “NP1” interval is not a subset of the “NP2” interval. This is because the test procedure for first-order (second-order) stochastic dominance identifies the point in the support where the difference between the cumulative (cumulated) distribution functions is most significant and then constructs the interval around this point. There is no reason why the point (and, hence, the intervals) identified should be the same or why the intervals should be related by set inclusion. Moreover, first-order stochastic dominance over a particular interval does not imply second-order stochastic dominance over that same interval because, for second-order dominance, the values of the cumulative distribution functions to the left of the first interval are also relevant. Hence, it may occur that we find an interval over which we reject non-first-order stochastic dominance, but we cannot find an interval over which we reject non-second-order stochastic dominance.

20 Behrman and Hoddinott (2005) obtain the same pattern when considering standardized height-for-age scores.

21 We compare the health outcomes of immediate and delayed treatment in the working paper version of the paper for children born between the beginning of the initial treatment and the beginning of the delayed treatment. This substantially limits the size of the sample. Moreover, because all of these children received at least three years of treatment by the time their health outcomes were measured, few significant effects can be found, particularly for hemoglobin concentration and reported days sick. This indicates that these variables are more sensitive to nutritional status in the immediate past than in the more distant past. We find a significant positive effect on standardized height for indigenous children without parental

primary education over a large range of the support of the distribution and for nonindigenous children with parental primary education over a limited support of the distribution. Again, the evidence is in favor of the program.

Figure 1. Stochastic dominance intervals for health outcomes among IL, IP, NL, and NP groups.

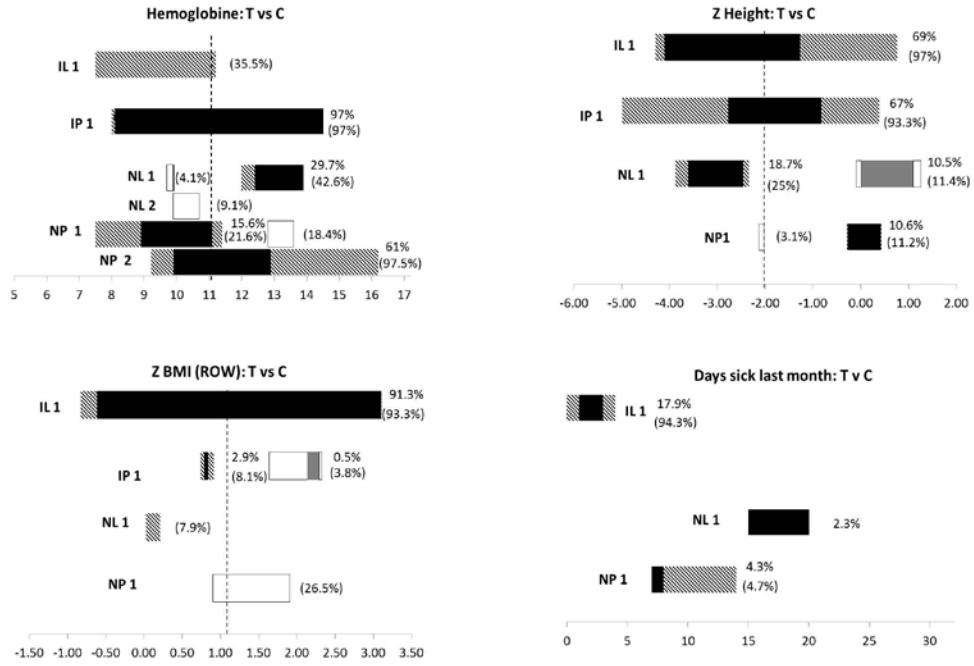




Figure A.1. Estimated propensity scores

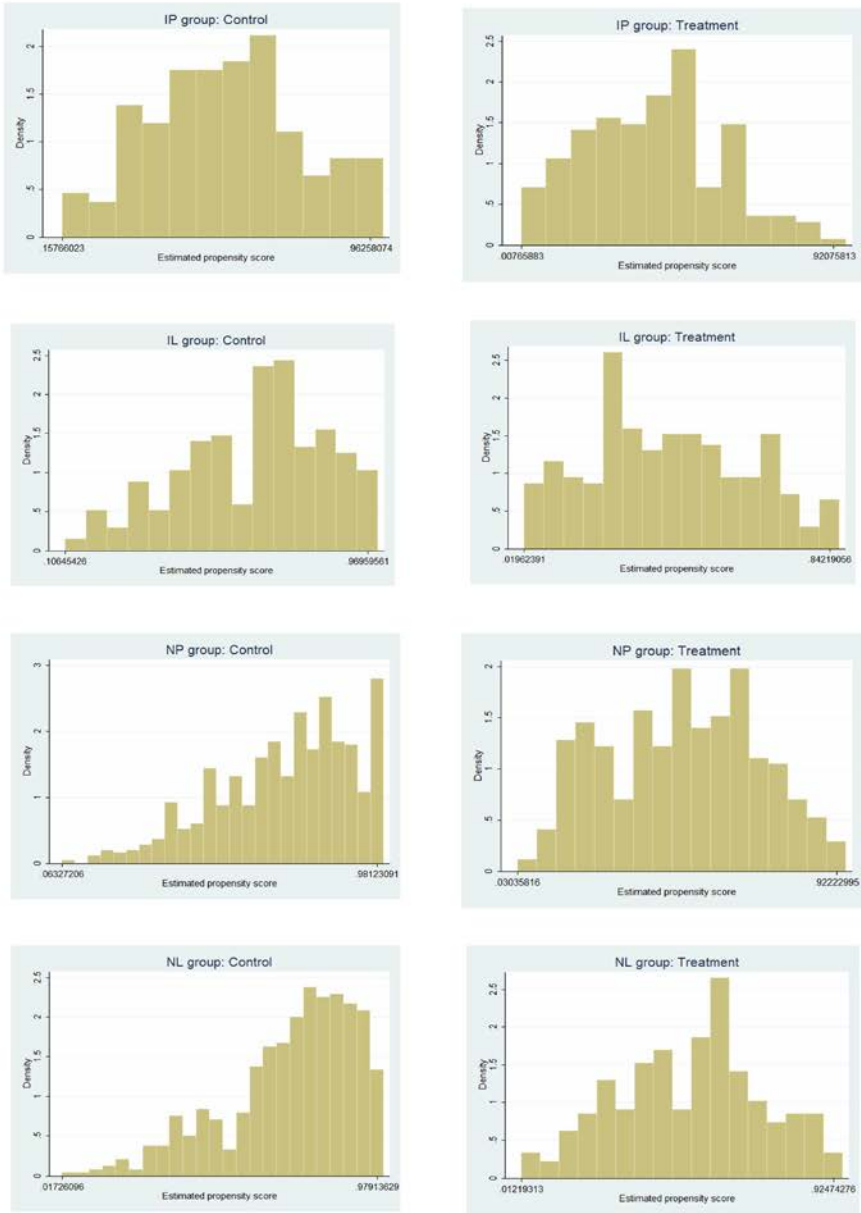


TABLE 1. Composition of the Samples

	Control sample		Treatment sample	
	#	%	#	%
All	1859	100	1125	100
IL	241	13.0	274	24.4
IP	173	9.3	209	18.6
NL	621	33.4	321	28.5
NP	824	44.3	321	28.5

*Source:* Authors' analysis based on data sources discussed in the text.

*Note:* The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

TABLE 2. Health Outcomes of Two- to Six-Year-Old Children in 2003

A. Control sample							
	Hemoglobin		zheight		zBMI	Days sick	
	Anemic	Median	Stunted	Median	ROW	0	>3
All	0.24	12.00	0.32	-1.46	0.24	0.58	0.17
IL	0.30	11.90	0.64	-2.40	0.30	0.64	0.13
IP	0.36	11.60	0.50	-1.99	0.23	0.57	0.19
NL	0.25	12.00	0.32	-1.47	0.25	0.58	0.18
NP	0.18	12.20	0.20	-1.13	0.22	0.56	0.18
B. Treatment sample							
	Hemoglobin		zheight		zBMI	Days sick	
	Anemic	Median	Stunted	Median	ROW	0	>3
All	0.23	12.10	0.34	-1.58	0.20	0.67	0.12
IL	0.29	11.70	0.43	-1.82	0.16	0.72	0.11
IP	0.27	12.00	0.35	-1.63	0.14	0.64	0.14
NL	0.24	12.20	0.33	-1.58	0.22	0.63	0.16
NP	0.13	12.50	0.26	-1.32	0.24	0.68	0.10

Source: Authors' analysis based on data sources discussed in the text.

Note: The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education. ROW, risk of being overweight.

TABLE 3. Difference between Control and Treatment Groups in the Fraction of Anemic, Stunted, at Risk of Overweight Children and Days Sick. Weighted Samples

	Anemic	Stunted	Risk overweight	0 days sick	>3 days sick
All	-0.03	0.01	-0.04	0.09**	-0.06**
IL	-0.05	-0.18*	-0.11**	0.10*	-0.05*
IP	-0.17**	-0.17**	-0.08	0.09	-0.06
NL	0.00	-0.01	-0.04	0.06	-0.02
NP	-0.08**	0.05	0.03	0.07	-0.09**

*Source:* Authors' analysis based on data sources discussed in the text.

*Note:* The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

TABLE A.1. Sampling Process

	Original number of children (a)	Matched children		1997 data available	
		number (b)	% of (a)	number	% of (b)
Control	2,247	1,871	83	1,871	100
Treatment	2,615	2,200	84	1,128	51
Total	4,862	4,071	84	2,999	73

*Source:* Authors' analysis based on data sources discussed in the text.

TABLE A.2. Logistic Regression Results.

Variable	Coef.	SE	z	Variable	Coef.	SE	z
Age Hh. head	-0.013	0.007	-1.96	Blender	-0.169	0.132	-1.27
Age spouse	-0.012	0.007	-0.61	Fridge	0.054	0.200	0.27
Sex Hh. head	-2.197	0.351	-6.25	Fan	0.142	0.120	0.71
Indig. Hh. head	-0.718	0.272	-2.64	Gas stove	0.377	0.145	2.60
Indig. Spouse	0.249	0.278	0.90	Gas heater	0.709	0.360	1.97
Educ. Hh. Head	-0.229	0.114	-2.01	Radio	-0.600	0.100	-5.96
Educ. spouse	-0.386	0.116	-3.32	Hifi	-0.361	0.251	-1.44
Work Hh. head	1.124	0.262	4.29	TV	-0.635	0.188	-5.53
Work spouse	0.623	0.161	3.86	Video	0.498	0.345	1.44
# Children 0-5	-0.090	0.048	-1.89	Washing machine	-0.35	0.330	-0.11
# Children 6-12	-0.211	0.042	-5.06	Car	1.229	0.465	2.64
# Children 13-15	-0.160	0.084	-1.91	Truck	0.243	0.282	0.86
# Children 16-20	-0.016	0.073	-0.22	Guerrero	-0.548	0.190	-2.88
# Women 20-39	-0.014	0.119	-0.12	Hidalgo	-0.937	0.209	-4.48
# Women 40-59	0.040	0.155	0.26	Michoacán	-0.582	0.176	-3.30
# Women 60+	0.040	0.185	0.22	Puebla	-1.097	0.150	-7.33
# Men 20-39	-0.162	0.106	-1.54	Querétaro	0.119	0.219	0.54
# Men 40-59	0.366	0.161	2.28	San Luis	-0.462	0.153	-3.02
# Men 60+	0.698	0.234	2.99	Miss Age Sp.	-4.297	0.713	-6.03
# Rooms	-0.006	0.010	-0.58	Miss Indg. Hh.	0.799	1.959	0.41
Electrical light	0.036	0.115	0.32	Miss Indg. Sp.	-2.102	1.894	-1.11
Running water land	0.879	0.115	7.67	Miss Work Hh.	3.461	1.871	1.85
Running water house	-0.435	0.208	-2.10	Miss Work Sp.	3.817	1.844	2.07
Dirt floor	0.096	0.118	0.81	Miss water land	0.871	1.640	0.53
Poor quality roof	-0.026	0.108	-0.24	Miss water house	0.699	0.827	0.84
Poor quality wall	-0.483	0.126	-3.82	Miss Assets	-4.121	2.398	-1.72
Animals	-0.168	0.113	-1.48	Constant	3.860	0.422	9.13
Land	-0.545	0.105	-5.17				
Number of Obs.	2,741						
LR $\chi^2$ (54)	730.0			Pseudo $R^2$	0.198		
Prob. > $\chi^2$	0.000			Log Likelihood	-1478.75		

Source: Authors' analysis based on data sources discussed in the text.

Note: Dependent variable equals one if the observation is in control and zero if the observation is in treatment group.

TABLE A.3. Propensity Score Matching: Common Support and Number of Observations in the Common Support

	Common support	Control #	Treatment #	Bandwidth
IL	[0.106, 0.868]	228	260	0.074
IP	[0.158, 0.957]	155	193	0.074
NL	[0.017, 0.952]	586	318	0.071
NP	[0.063, 0.949]	668	318	0.071
Total		1,637	1,089	

*Source:* Authors' analysis based on data sources discussed in the text.

*Note:* The acronyms refer to the following types: IL, indigenous, less than primary education; IP, indigenous, primary education; NL, nonindigenous, less than primary education; NP, nonindigenous, primary education.

## Supplemental Appendix 1: testing stochastic dominance

We explain the approach by focussing on tests for first order stochastic dominance of  $F^T$  over  $F^C$ . Davidson(2009) shows how the approach must be generalized to test for stochastic dominance of arbitrary order.

It is assumed that samples of the control and treatment types that are compared are independent, and their weighted empirical distribution functions  $\hat{F}^C$  and  $\hat{F}^T$  are defined in the usual way. If for the empirical distribution functions  $\hat{F}^C$  and  $\hat{F}^T$ , there exists a  $y \in R$  such that  $\hat{F}^T(y) \geq \hat{F}^C(y)$ , there is non-dominance in the sample and we do not wish to reject the null.

Davidson and Duclos (2009) restrict the test to a test of the frontier of the null hypothesis against the alternative hypothesis of dominance of  $T$  over  $C$ . The frontier of the null hypothesis is the case where  $\hat{F}^C(y) > \hat{F}^T(y)$  for all  $y \in R$  except for one point  $y^*$  where  $\hat{F}^C(y^*) = \hat{F}^T(y^*)$ . They show that, for configurations of non-dominance that are not on the frontier, the rejection probabilities of their test are no greater than they are for configurations on the frontier.

For each point in  $R$ , we calculate an unconstrained empirical likelihood ratio statistic and a constrained empirical likelihood ratio statistic, the statistic under the frontier of the null (i.e. imposing the null of non-dominance). The square root of the double difference between these two statistic is the test statistic.<sup>1</sup> Denote this value by  $LR$ . Next, determine the value for which  $LR$  is minimal, as this is the most likely point at which non-dominance cannot be rejected and compute the probabilities  $p_t^X$  associated with each point in sample  $X$  ( $x = C, T$ ) that maximizes the empirical likelihood function subject to  $\hat{F}^C(y^*) = \hat{F}^T(y^*)$ . These probabilities are estimates of the population probabilities under the assumption of non-dominance and are used to set up the following bootstrap data-generating process on the frontier of the null of non-dominance.

We compute 3000 bootstrap samples from the two distributions  $p_t^C$  and  $p_t^T$ , following the original sample design, as suggested by ?. Our samples contain  $C_1^X, \dots, C_c^X, \dots, C_{n^X}^X$  clusters (villages),  $X = C, T$ . Each cluster in the sample contains  $n_c^X$  children ( $c = 1, \dots, n^X$ ). We mimic this sample design as follows. First, define for each cluster

$$\pi_c^X = \frac{\sum_{t \in C_c^X} p_t^X}{\sum_{t \in \cup_{c=1 \dots n^X} C_c^X} p_t^X},$$

which gives the probability that an observation is drawn from cluster  $c$ . Now, draw the identity of the first cluster from the  $n^X$  clusters, such that each cluster has a probability  $\pi_c^X$  of being drawn. This gives, say cluster  $k$ . Next, draw  $n_1^X$  observations from cluster  $k$  with replacement, where each observation has a probability  $p_t^k / \sum_{t \in C_k^X} p_t^X$  of being drawn. Do the same for all the other  $n^X - 1$  clusters. This gives the first bootstrap sample. Repeat the procedure 3000 times. For each bootstrap sample, we calculate the minimal  $LR$  statistic to get an idea of the distribution of the minimal  $LR$  under the frontier of the null hypothesis.

<sup>1</sup>For first order stochastic dominance, this statistic can be analytically obtained. For second order dominance the statistic has to be numerically determined using the Newton method to solve a set of non-linear equations -see Davidson (2009).



The  $p$ -value of the sample statistic is then the fraction of bootstrap-statistics greater than the sample statistic.

When there is dominance in the sample, we report the results by giving the longest interval  $[\hat{r}^-, \hat{r}^+]$  for which the hypothesis

$$\max_{z \in [\hat{r}^-, \hat{r}^+]} (F^T(z) - F^C(z)) \geq 0,$$

can be rejected. For a given level of significance  $\alpha$ ,  $\hat{r}^-$  ( $\hat{r}^+$ ) is the smallest (greatest) value of  $r^-$  ( $r^+$ ) for which the hypothesis

$$\max_{z \in [r^-, r^+]} (F^T(z) - F^C(z)) \geq 0$$

can be rejected at level  $\alpha$ . The larger is this interval, given  $\alpha$ , the more powerful our rejection of non-dominance. We ignore the stochastic nature of the sampling weights.

## Supplemental Appendix 2: Roemer's identification axiom and matching estimator (weighted treatment distribution)

(1) The standard Roemer model and its assumptions

In the standard model health,  $h$ , is determined only by parental circumstances,  $c$ , and a scalar representing parental responsibility,  $p$ :

$$h(c, p).$$

Define, for each type  $\hat{h}^i$  as the level of health such that a fraction  $R$  of type  $i$  has a health not better than  $R$ :

$$\int_P I(h(c_i, p) \leq \hat{h}^i) f_p^i(p) dp = R, \quad (1)$$

where  $I(\cdot)$  is the indicator function. The first assumption typically made to derive Roemer's identification axiom is

**A1:**  $h(c, p)$  is strictly increasing in  $p$ .

As a result of this assumption, there exists for each type a value  $p^i$  such that

$$I(h(c_i, p) \leq \hat{h}^i) = 1 \Leftrightarrow p \leq p^i,$$

and we get from (1),

$$R = \int_{\underline{p}}^{p^i} f_p^i(p) dp.$$

Imposing the second assumption,

**A2:** For all  $i$ ,  $f_p^i(p) = f_p(p)$ ,

which says that responsibility is distributed independently from circumstances, we get

**RIA:** For all  $i$ ,  $p^i = p^*$ ,

which is Roemer's identification axiom: those that are at the same percentile in the distribution of health within their type, have the same responsibility.

(2) Weighted treatment observations and a variant of RIA

Suppose children's health is influenced by parental circumstances,  $c$ , pre-program characteristics,  $x$ , and a scalar representing parental responsibility,  $p$ :

$$h(c, x, p).$$

Define for the treatment sample after weighting the observations the value  $\hat{h}^T$  and for the control sample the value  $\hat{h}^C$  such that the same fraction in both samples has a health smaller than or equal to these critical values.

$$\int_P \int_X I(h(c^T, x, p) \leq \hat{h}^T) \hat{f}_{x,p}^T(x, p) dx dp =$$

$$\int_P \int_X I(h(c^C, x, p) \leq \hat{h}^C) f_{x,p}^C(x, p) dx dp, \quad (2)$$

where

$$\hat{f}_{x,p}^T(x, p) = f_{p|x}^T(p|x) f_x^C(x),$$

the joint distribution of  $x$  and  $p$  after weighting the observations in the treatment sample, which ensures that the marginal distribution of  $x$  is the same in the control and treatment sample. A first assumption that can be made is

**A3:**  $f_{p|x}^T(p|x) = f_{p|x}^C(p|x)$ .

This says that the distribution of responsibility conditional on  $x$  is the same in the treatment and control group. It implies that

$$\hat{f}_{x,p}^T(x, p) = f_{x,p}^C(x, p). \quad (3)$$

As a result, (2) reduces to

$$\int_P \int_X \left[ I(h(c^T, x, p) \leq \hat{h}^T) - I(h(c^C, x, p) \leq \hat{h}^C) \right] f_{x,p}^C(x, p) dx dp = 0. \quad (4)$$

A second assumption that can be made is that the function  $h(c, x, p)$  is additively separable between  $c$  and  $(x, p)$ .

**A4:** There exist functions  $v(x)$  and  $w(c, p)$  such that  $h(c, x, p) = v(x) + w(c, p)$ .

This allows us to write (4) as

$$\int_P \int_X \left[ I(w(x, p) \leq \hat{h}^T - v(c^T)) - I(w(x, p) \leq \hat{h}^C - v(c^C)) \right] f_{x,p}^C(x, p) dx dp = 0.$$

As this equation must hold for arbitrary distribution functions  $f_{x,p}^C(x, p)$ , it follows that

$$\hat{h}^T - v(c^T) = \hat{h}^C - v(c^C).$$

As a result,

$$\begin{aligned} h(c^T, x, p) = \hat{h}^T &\Leftrightarrow v(c^T) + w(x, p) = \hat{h}^T \Leftrightarrow w(x, p) = \hat{h}^T - v(c^T) \\ &\Leftrightarrow w(x, p) = \hat{h}^C - v(c^C) \Leftrightarrow h(c^C, x, p) = \hat{h}^C. \end{aligned}$$

Now consider the expected value of  $p$  in the weighted treated and control sample, given that health is at the same percentile.

$$E(p|h = \hat{h}^T) = \frac{1}{\hat{f}_h^T(h)} \int_P p \int_X I(h(c^T, x, p) = \hat{h}^T) \hat{f}_{p,x}^T(p, x) dx dp, \quad (5)$$

$$E(p|h = \hat{h}^C) = \frac{1}{f_h^C(h)} \int_P p \int_X I(h(c^C, x, p) = \hat{h}^C) f_{p,x}^C(p, x) dx dp. \quad (6)$$

We have shown that weighting the treatment sample and A3 implies (3) and that A3 together with A4 imply  $h(c^T, x, p) = \hat{h}^T \Leftrightarrow h(c^C, x, p) = \hat{h}^C$ , such that the expressions behind the

first integral sign in (5) and (6) are equal. What needs to be shown is that the marginal distributions  $\widehat{f}_h^T(h)$  and  $f_h^C(h)$  are equal. This follows directly from the previous reasoning, upon observing that

$$\begin{aligned}\widehat{f}_h^T(h) &= \int_P \int_X I(h(c^T, x, p) = \widehat{h}^T) \widehat{f}_{p,x}^T(p, x) dx dp \quad \text{and} \\ f_h^C(h) &= \int_P \int_X I(h(c^C, x, p) = \widehat{h}^C) f_{p,x}^C(p, x) dx dp.\end{aligned}$$

Conclusion: if both assumptions A3 and A4 hold true, then the weighting procedure guarantees that those that are at the same percentile in the distribution of health in the weighted treatment and control sample have the same expected value for responsibility.

## Supplemental Appendix 3: treatment and control effects in matched samples

**Table S.1:** Health outcomes of 2-6 year old children in 2003.

(a) Control sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.24	12.0	0.32	-1.47	0.24	0.58	0.17
IL	0.30	11.9	0.63	-2.36	0.30	0.63	0.13
IP	0.36	11.5	0.46	-1.91	0.23	0.54	0.19
NL	0.24	12.0	0.32	-1.47	0.26	0.58	0.17
NP	0.18	12.2	0.19	-1.12	0.21	0.57	0.18

(a) Treatment sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.20	12.1	0.32	-1.47	0.19	0.67	0.11
IL	0.25	11.7	0.45	-1.86	0.18	0.71	0.07
IP	0.19	12.0	0.30	-1.52	0.14	0.66	0.12
NL	0.25	12.3	0.30	-1.41	0.21	0.64	0.15
NP	0.10	12.4	0.24	-1.10	0.25	0.68	0.09

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Less than primary;  
 NP = Non-indigenous, Primary education; NL = Non-indigenous, Less than primary.

Source: Authors' analysis based on data sources discussed in the text

As expected since we match the treatment sample to the control samples, the characteristics of the matched control sample are very similar to those of the original control sample in table 2. The differences between the matched and original treatment sample are larger.