UNIVERSITEIT GENT

FACULTEIT ECONOMIE EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Children's health opportunities and project evaluation: Mexico's *Oportunidades* program

Dirk Van de gaer

Joost Vandenbossche

José Luis Figueroa

October 2011

2011/749

D/2011/7012/54

Children's health opportunities and project evaluation: Mexico's *Oportunidades* program*

Dirk Van de gaer† Joost Vandenbossche‡ José Luis Figueroa§

October 2011

Abstract

We propose a methodology to evaluate social projects from an (equality of) opportunity perspective by looking at their effect on (parts of) the distribution of outcomes conditional on morally irrelevant characteristics, taken here to be parental education level and indigenous background. The methodology is applied to evaluate the effects on children's health outcomes of Mexico's Oportunidades program, one of the world's largest conditional cash transfer programs for poor households. The evidence shows that the gains in health opportunities for children from indigenous background are substantial and situated in crucial parts of the distribution, while the gains for children from non-indigenous backgrounds are more limited.

Keywords: project evaluation, opportunities, oportunidades program.
JEL classification: I18, I38, D63

*This work was supported by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office, Science Policy Programming [Contract No. P6/07]. We thank Bart Cockx, 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 how to incorparate the survey design in the bootstrap procedure. We gratefully acknowledge comments received on preliminary versions presented at the GREQAM-IDEP workshop "The Multiple Dimensions of Equality and Fairness" (Marseille, France, 17 November 2010), the OPHI workshop "Inequalities of Opportunities" (Oxford, UK, 22-23 November 2010), the UAB workshop "Equality of Opportunity and Intergenerational Mobility" (Barcelona, Spain, 17 December 2010), the winterschool on "Inequality and Social Welfare Theory" (Canazei, Italy, 10-13 January 2011), the faculty seminar in Caen (France, 28 March 2011), the workshop "Equity in Health" (Louvain la Neuve, Belgium, 11-13 May 2011), the ABCDE conference (Paris, France, 30 May - 01 June 2011), the conference "Mind the Gap: from Evidence to Policy" (Cuernavaca, Mexico, 15-17 June 2011), the conference "Micro Evidence on Innovation in Developing Countries" (San Jose, Costa Rica, June 27-28, 2011), the ECINEQ conference (Catania, Italy, 18-20 July 2011) and the EEA conference (Oslo, Norway, 25-29 August 2011).

†Corresponding author. SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Université Catholique de Louvain, CORE, B-1348, Louvain la Neuve, Belgium. Tel:+32-(0)9-2643490. Fax: +32-(0)9-2648996. E-mail: Dirk.Vandegaer@ugent.be

‡SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University, Tweekerkenstraat 2, B-9000 Gent, Belgium and Aspirant FWO - Flanders.

§SHERPPA, Vakgroep Sociale Economie, F.E.B., Ghent University,Tweekerkenstraat 2, B-9000 Gent, Belgium and CES, Katholieke Universiteit Leuven.

1 Introduction

The present paper evaluates the change in opportunities for health of two to six years old children brought about by the Mexican *Oportunidades* program. This program is a large scale conditional cash transfer program that started in 1998, where poor rural households receive cash in exchange for complying with preventive health requirements and nutrition supplementation, education and monitoring. In 2010 about 5.8 million families participated in the program and the transfers to the poor totalled $4.8 billion. The impact of the program on the health outcomes of young children has been analyzed before, but we add to this literature by doing program evaluation from the perspective of children's opportunities rather than by identifying average treatment effects. Fiszbein et al. (2009) report that in 1997 only 3 developing countries (Mexico, Brazil and Bangladesh) had conditional cash transfer programs in place, while, by 2008, this number had increased to 29 and many more countries were planning to start such programs. Given the increased popularity of conditional cash transfer programs in developing countries, their sometimes large scale, their focuss on breaking the intergenerational poverty cycle and the recent emergence of a substantial empirical literature measuring inequality of opportunity (see, e.g., de Barros et al. (2009) and the references below), the development of techniques for program evaluation from the equality of opportunity perspective is an important task.

To operationalize the opportunities, we find inspiration in recent theories of (equality of) opportunities. In this literature (see, e.g., Fleurbaey (1995), Bossert (1995) and Roemer (1993) or, for a recent survey, Fleurbaey (2008)) one makes a distinction between two kinds of factors that influence the outcome under consideration. At the one hand, there are circumstances, characteristics for which the individual is not responsible, such as his race, sex and parental background. At the other hand, there are characteristics for which individuals are taken to be responsible, such as hard work. The idea is that public policies and thus also conditional cash transfer programs should compensate for the former, while respecting the influence of the latter.[1]

In our context, where we apply the framework to health outcomes of children aged between two and six, we take as circumstances race (whether the child's family is indigenous or not), whether a parent had primary education or not and whether the child's family participated in the program or not. With each combination of circumstances corresponds a "type" in Roemer (1993)'s terminology. Hence we have 8 types. In order to evaluate the program, we compare the health outcomes of the children belonging to a family enrolled in the program for each of the 4 types defined on the basis of their race and parental education level with the health outcomes of the children belonging to a family of the corresponding type that was not enrolled. Within each type outcomes can (and will) be different, due to factors that have not been accounted for by conditioning on type. In the present context, children's genetic make-up will probably be the most important factor. Different normative theories treat genetic make-up as either a responsibility or a compensation factor. We argue in section 2 that in the former case, comparison of treatment and control types has to be limited to first order stochastic dominance of the cumulative distribution of health outcomes conditional on type, while in the latter case second order stochastic dominance is acceptable.

[1] Recently Lefranc et al. (2009) extend 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).

The idea to use first and/or second order stochastic dominance to investigate equality of opportunity for a particular outcome is not new. So far, it has only been applied to study whether opportunities are equal within a particular population (see, e. g., O'Neill et al. (2000) and Lefranc et al. (2009) where the outcome is income, or Rosa Dias (2009) and Trannoy et al. (2010) for adults' self-assessed health) or between different countries (see, e.g., Lefranc et al. (2008) for income) or regions (see, e.g., Peragine and Serlenga (2008) for education). Compared to this literature, our paper has 3 contributions. First, and most importantly, we perform program evaluation: we establish the effect of the *oportunidades* program on children's health opportunities. Second, we look at opportunity for health of very young children, as their health is not only important for their adult outcomes (see, e.g. Black et al. (2007) and Alderman et al. (2006)), but is crucial in its own right. Third, contrary to the previous literature testing for stochastic dominance in the context of equality of opportunity, our test procedure is based on Davidson and Duclos (2009) and Davidson (2009): we test the null of non-dominance against the alternative of dominance, such that rejection of the null logically entails dominance.

Most of the literature on program evaluation focusses on the estimation of average treatment effects. We are interested in establishing or rejecting stochastic dominance between the distributions of health outcomes of children when their families are in and out of the program. This is not a trivial exercise as we cannot observe the same child in and out of the program; 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 that can be made is perfect positive quantile dependence (see Heckman et al. (1997)), which says that those that are at the $q-$ th quantile in the distribution with treatment would have been at the $q-$ th quantile in the distribution without treatment. We argue below that Roemer (1993)'s identification axiom, which is usually invoked in empirical applications of equality of opportunity when responsibility characteristics are unobserved and says that those that are at the same percentile of their type distribution have a comparable responsibility, provides a normatively inspired alternative to perfect positive quantile dependence that reduces 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 have to be comparable in terms of pre-program characteristics. We show this is also an issue when testing for stochastic dominance, and, following the literature on average treatment effects, we propose a propensity score matching technique on the basis of pre-program characteristics to make treatment and control types (more) comparable. Finally, it is worth observing that recently two authors suggested to incorporate ideas of stochastic dominance into project evaluation. Verme (2010) proposes a stochastic dominance approach to determine the effect of a perfectly randomized experiment on Foster et al. (1984) poverty measures to establish poverty line dominance (i.e. dominance for a range of poverty lines). Our approach, based on equality of opportunity, stresses that we should not compare the distributions of the entire treatment and control samples, but the distributions of corresponding treatment and control types. Moreover, our propensity score matching technique makes the approach workable for imperfectly randomized experiments. Naschold and Barrett (2010) also allow for non-randomized treatment by focussing on stochastic dominance between treatment and control samples of the distribution of the difference in outcome before and after treatment. They do not focus on types and the problem is that the results are difficult to interpret, as dominance in terms of differences does not imply that treatment leads to a dominating

distribution since this crucially 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 poor indigenous families. The effects on children growing up in poor non-indigenous families are less strong, but we still find some significant positive treatment effects.

The paper is structured as follows. Section 2 provides definitions and explains the methodology. The data are described in section 3, section 4 gives the empirical results, including a discussion of the relationship with previous studies. Section 5 concludes.

2 Definitions and methodology

Let a child's health outcome, $h \in H = \left[\underline{h}, \overline{h}\right] \subseteq \mathbb{R}$ and higher values for h mean better health. A child's health is the result of two types of variables: circumstances for which it is not responsible, $c \in C$ (race, parental background, family being in the program) and a variable $l \in L$ representing genetic luck. Each combination of circumstances corresponds to a type. From the perspective of the equality of opportunity literature, social programs should compensate differences that are due to circumstances. There is some discussion whether genetic luck is a circumstance or a responsibility variable. According to the libertarian principle of self-ownership agents are entitled to the full benefit of their natural personal endowment (Nozick (1977)), which implies that differences in health due to genetic luck should be respected, such that l becomes a responsibility variable. Most people, however, will probably find that genetic luck is not a responsibility characteristic, but a compensation characteristic (as advocated by, e.g., Rawls (1971)), which means that full equality of health for children at as good a level as possible should be the ideal.

In many empirical applications, responsibility is unobserved, and so it is here as we lack observations on genetic luck. In such cases the equality of opportunity framework is usually operationalized using Roemer (1993)'s identification axiom, which translates as saying that two persons at the same percentile of their type distribution of health have a comparable degree of genetic luck.[2] Hence, if the cumulative distribution function of health for one of the four types whose family was in the program lies below the cumulative distribution function of the corresponding type that did not participate in the program, for this type, being in the program ensured that a worse genetic endowment suffices to obtain a particular level of health. If this holds for all levels of health, program participation unambiguously improved 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. A similar reasoning applies to second order stochastic dominance, but remember that second order stochastic dominance can also be obtained by within type inequality reducing transfers of health, which do not respect the influence of genetic luck, and are therefor only acceptable if genetic luck is considered to be a compensation variable. Roemer's identification axiom does not necessarily imply that we would find exactly the same individual with and without treatment at the $q-$ th

[2]See Roemer (1993) and Roemer (1998) for a defense of this principle and Fleurbaey (1998) for a discussion of the assumptions involved.

quantile (which is Heckman et al. (1997)'s perfect positive quantile dependence). It just says that the comparison of the quantiles of the treated and corresponding untreated type is the normatively relevant one.

Let the conditional distribution of health for those with circumstances c in the control sample be denoted by $F^C(h \mid c)$ and in the treatment sample by $F^T(h \mid c)$. We say that the project improves the opportunities for health of children with circumstances c if the conditional distribution $F^T(h \mid c)$ first order stochastically dominates the conditional distribution $F^C(h \mid c)$ and we test whether first order stochastic dominance occurs. Hence the issue of statistical inference arises. We follow Davidson and Duclos (2009) and start from non-dominance as the null hypothesis. To illustrate the procedure for testing first order dominance and describe the test more formally, let $U \subseteq H$ be the union of the supports of $F^C(h \mid c)$ and $F^T(h \mid c)$. We test the null hypothesis of non-dominance of $F^C(h \mid c)$ by $F^T(h \mid c)$:

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

against the alternative hypothesis that $F^T(h \mid c)$ first order stochastically dominates $F^C(h \mid c)$:

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

This approach has as a first merit that, if we succeed in rejecting the null, the only other possibility is dominance, enabling us to draw the conclusion of dominance. By contrast, if, as is the case in most empirical work to date, dominance would be the null hypothesis, failure to reject dominance does not enable us to accept dominance. As Davidson and Duclos (2009) point out, taking non-dominance as the null comes at the cost that (with continuous distributions), it is not possible to reject non-dominance in favor of dominance over the entire support of the distribution[3]. Rejecting non-dominance is typically possible only over *restricted* ranges of the observed variable. This leads us to a second merit of the approach, as it allows us to identify the maximal range over the supports of the distribution for which we are able to reject the null of non-dominance and, hence, accept dominance in favor of the project. That way we can check whether we have dominance over ranges of the observed variable that are of special importance, for example the range below -2 for standardized height which indicates stuntedness.

Of course, we also have to test the null of non-dominance of $F^T(h \mid c)$ by $F^C(h \mid c)$ against the alternative hypothesis that $F^C(h \mid c)$ dominates $F^T(h \mid c)$ using the same procedure. Also here, if rejection occurs, we identify the maximal range over the support of the distribution for which we are able to reject the null of non-dominance and, hence, accept dominance against the project.[4] These issues are incorporated in the following weak version of improvements in opportunities which suffices for most of what we do in this paper.

Definition FOI (First Order Improvements): the project leads to a first order improvement of the opportunities of children with circumstances c if (i) there exists $U^0 \subseteq U$ such that we can reject the null of non-dominance of $F^C(h \mid c)$ by $F^T(h \mid c)$ against the alternative that

[3]Let $\underline{h}$ be the lower bound of U. Evidently, $F^T(\underline{h} \mid c) - F^C(\underline{h} \mid c) = 0$, and so the maximum over U is never less than 0. Moreover, close to the boundaries of the support there may be too little information to reject non-dominance.

[4]Appendix 5 contains more details about the stochastic dominance tests.

$F^T(h \mid c)$ dominates $F^C(h \mid c)$ over U^0 and (ii) there does not exist $U^1 \subseteq U$ such that we can reject the null of non-dominance of $F^T(h \mid c)$ by $F^C(h \mid c)$ against the alternative that $F^C(h \mid c)$ dominates $F^T(h \mid c)$ over U^1.

Assuming that genetic luck l is a compensation variable and that h is cardinally measurable, equalizing health outcomes within type becomes desirable such that, in case the project does not lead to a first order improvement, it becomes meaningful to ask whether the conditional distribution $F^T(h \mid c)$ second order stochastically dominates the conditional distribution $F^C(h \mid c)$. Similar statistical issues as for first order stochastic dominance arise (see Davidson (2009)), leading to the following definition.

Definition SOI (Second Order Improvements): the project leads to a second order improvement of the opportunities of children with 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 $F^C(h \mid c)$ by $F^T(h \mid c)$ against the alternative that $F^T(h \mid c)$ second order stochastically dominates $F^C(h \mid c)$ over U^0 and (iii) there does not exist $U^1 \subseteq U$ such that we can reject the null of absence of second order stochastic dominance of $F^T(h \mid c)$ by $F^C(h \mid c)$ against the alternative that $F^C(h \mid c)$ second order stochastically dominates $F^T(h \mid c)$ over U^1.

A final comment on the empirical procedure is important. When comparing conditional distribution functions for program evaluation, one must be aware that the presence of unaccounted for pre-program characteristics (including unaccounted for compensation characteristics) that differ between the compared treatment and control types can lead to wrong conclusions. Suppose we have two sets of characteristics, observable circumstances, c and unaccounted for pre-program characteristics $x \in X$. We then have for the type with observed circumstances c_1

$$F(h \mid c_1) = \frac{\int_{\underline{h}}^{h} f\left(\widetilde{h}, c_1\right) d\widetilde{h}}{f(c_1)} = \frac{\int_X \int_{\underline{h}}^{h} f\left(\widetilde{h}, c_1, x\right) d\widetilde{h} dx}{f(c_1)}$$

$$= \int_X \int_{\underline{h}}^{h} f\left(\widetilde{h} \mid c_1, x\right) \frac{f(c_1, x)}{f(c_1)} d\widetilde{h} d\widetilde{c}_2 = \int_X F(h \mid c_1, x) f(x \mid c_1) dx.$$

This clearly shows that the composition of the c type in terms of x matters. Indeed, suppose the treatment has no effect ($F^C(h \mid c_1, x) = F^T(h \mid c_1, x)$), but the composition of those with circumstances c_1 is different between control and treatment type, say $f^C(x \mid c_1)$ is higher (lower) for (un-)favorable pre-program characteristics x -i.e. characteristics for which $F^C(h \mid c_1, x)$ is lower (higher)- than $f^T(x \mid c_1)$. This makes $F^C(h \mid c_1)$ smaller (higher) than $F^T(h \mid c_1)$, such that we would erroneously infer that the treatment had an adverse (benign) effect on the opportunities of those with circumstances c_1.

3 Data description

3.1 The oportunidades program

The *Oportunidades* program is a conditional cash transfer program: bimonthly cash transfers are provided to households in extreme poverty conditionally on school attendance of children, health care visits for all members of the household and presence at information sessions

about primary health and nutrition. School benefits constitute the largest part of the conditional cash transfer. The total amount of the transfers a households receives depends on the number, age and sex of its children. On average households receive about 20% of household consumption.

Special emphasis is placed on intervention for small children and their mothers. Prenatal and postpartum care visits, growth monitoring, immunization, management of diarrhea and antiparasitic treatments are provided to mothers and small children. Children between 4 and 23 months attend periodical medical check-ups, 9 in total. After 23 months and up to the age of 19 years, two check-ups per year are obligatory for all household members. Children between 6 and 23 months of age, lactating women and low weight children between 2 and 4 years of age receive milk-based micronutrient fortified foods with a daily equivalent recommended intake of zinc, iron and essential vitamins [5].

3.2 Sample Design

The selection of an immediate and delayed treatment sample proceeded in several steps (see, e.g., INSP (2005)). Highly deprived localities were identified on the basis of a deprivation index for each of the localities in the country for which socio-demographic data in national censuses were available. Localities with at least 50 and less than 2500 inhabitants that were categorized as having high or very high deprivation and had access to elementary school, middle school and a health clinic were eligible for treatment. A random procedure stratified by locality size proportional to the number of localities determined which localities receive treatment. In the selected localities poverty conditions of all households were evaluated and households categorized as in extreme poverty were included in the program. This categorization was based on household income, characteristics of the household head and variables related to households' dwelling conditions. Comments by a community assembly on the inclusion and exclusion of households were taken into account, if these comments met certain criteria previously established for the identification of beneficiary families. These steps led to a selection of 506 localities (and their households). A random procedure assigned 320 of them to receive immediate treatment; the remaining 186 started to receive treatment about 18 months later. As we limit our sample to those that actually received conditional cash transfers, and these data were only available for the delayed treatment sample, our treatment sample is a subsample of the delayed treatment sample. Sensitivity analysis (reported in appendix B) shows that the results are very similar when we take the entire delayed treatment sample.

Once the delayed treatment sample started to receive treatment, there was a need to construct a new control sample. The intention was to make 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 on access to elementary school, middle school and access to a health clinic were excluded. Next, a multiple matching propensity score method was used, based on data at the locality level as a function of several observed characteristics at the community level from the Census in 2000 that permit comparison with the localities of the original sample. This

[5] These supplements may also be given to children in households 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 impact of *Oportunidades* (see also Behrman et al. (2009b), footnote 8)

led to a selection of 151 localities, the households of which that met the criteria for program eligibility were included in the control sample.

As stressed at the end of section 2, the households in the treatment and control sample have to be comparable in terms of pre-program characteristics. There are important problems with the way the control sample was selected[6]. Matching at the locality level was done on the basis of a comparison with observable characteristics in 2000, which is at a time that the treatment sample already received treatment, while it should be done on the basis of characteristics before treatment started. Matching at the locality level does not imply matching at the household level (see also Behrman and Todd (1999)). Moreover, we do not have data on all the children of the households that were in the delayed treatment sample for three reasons (see table A.1 in appendix 1). First, some households drop out of the sample due to classical problems of sample attrition. Second, the health data were only collected for a sub-sample of children. Third, due to problems with the household identifiers, it appeared impossible to match all children for which health data are available uniquely to one household. We only included the unique matches in our samples (which fortunately accounts for more than 80% of the children). The second and third problem also occurred in the control sample. As a result of all these issues, the treatment and control sample can be different in terms of pre-program characteristics.

Our empirical strategy in section 4 is the following. First we use a logistic regression approach to test whether there are statistically significant differences in composition between the treatment and control sample in 1997 for the households with children that were observed in 2003[7]. We use a propensity score matching technique to match each of the four treatment types with the corresponding control type to correct for the possible under- (and over-) representation of some household types. This amounts to weighted sampling (see appendix 3). We compare the resulting weighted distributions at crucial points (such as standardized height smaller than -2 indicating stuntedness) and establish whether the treatment leads to first or second order improvements of types' opportunities by performing stochastic dominance tests on the weighted distribution functions.

3.3 Circumstances and outcomes

Ideally, normative theory requires us to obtain a full description of children's circumstances. In reality an exhaustive description is not available in surveys and due to the limited number of observations the inclusion of an extensive set of circumstances is statistically unworkable for non-parametric procedures such as ours. For these reasons, we limit ourselves to program participation and two additional circumstances.

The first circumstance refers to parental background. In the literature on equality of oppor-

[6]This can explain why the control sample has hardly been used in academic papers. Most studies focus on a comparison of the immediate and delayed treatment samples and thus evaluate the effect of differences in duration of program participation, see, e.g., Schultz (2004), Behrman et al. (2005) or Behrman et al. (2009a). Recently, however, matched sampling was used to compare schooling (Behrman et al. (2009b) and Behrman and Parker (2010)) and work outcomes (Behrman and Parker (2010)) of the immediate treatment, delayed treatment and control samples.

[7]In 2003, besides the regular household data, an additional questionnaire with recall data was collected. The purpose of these restrospective questions was to compare households' pre-program characteristics for the treatment samples with the new control sample.

Table 1: Composition of the samples.

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

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

tunity for income this variable is used most frequently, is always statistically significant and has been shown to be the most important circumstance characteristic (see, e.g., Bourguignon et al. (2007) and Ferreira and Gignoux (2011)). We measure parental background by a dichotomous variable indicating whether at least one of the child's parents completed primary education. In appendix C we report the results when parental background is measured on the basis of mother's education only. The results are very similar to the ones we present in the main text. The second circumstance variable refers to the indigenous background of the child. There is a substantial literature indicating that indigenous people remain disadvantaged in Mexico (Rivera et al. (2003), SEDESOL (2008), Olaiz et al. (2006), Psacharopoulos and Patrinos (1994)). We consider the child to have an indigenous background if at least one parent can speak and/or understand an indigenous language. We don't include the sex of the child (boy or girl), as existing evidence (see, e.g., Backstrand et al. (1997)) indicates that sex is an unimportant determinant for young children's health outcomes in Mexico.

Combining these 2 binary characteristics and a binary characteristic indicating program participation gives us 8 types in Roemer's terminology. Table 1 shows that there are remarkable differences in sample composition between the control and the treatment sample when we partition the samples on the basis of indigenous origin (I = indigenous, N=non-indigenous) and parental level of education (P = Primary, L = Less than primary). Clearly, the control sample contains less indigenous children and more non-indigenous children with at least one parent completing primary education than the treatment sample. Since we compare 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 2, problems arise when there are important differences in terms of pre-program characteristics between the treatment and control types that are compared.

We focus on several health outcomes. Two important measures of malnutrition for children are anemia, which is defined as having hemoglobin levels lower than 110 g/l, and stunting, which covers a wider range of nutritional deficiencies and is defined as height for age below -2 standard deviations of the WHO International Growth Reference. The latter implies that in a reference population about 2.3% of the population is stunted. As reviewed by Grantham-McGregor and Ani (2001), anemia (iron deficiency) in infancy has been shown to be associated with poorer cognition, school achievement and behavior problems into middle childhood. Branca and Ferrari (2002) point out that stunting is associated with develop-

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

(a) Control sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.36	11.6	0.50	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.40	0.30	0.64	0.13
NP	0.18	12.2	0.20	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.47	0.25	0.58	0.18

(b) Treatment sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.23	12.1	0.34	-1.58	0.20	0.67	0.12
IP	0.27	12.0	0.35	-1.63	0.14	0.64	0.14
IL	0.29	11.7	0.43	-1.82	0.16	0.72	0.11
NP	0.13	12.5	0.26	-1.32	0.24	0.68	0.10
NL	0.24	12.2	0.33	-1.58	0.22	0.63	0.16

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

mental delay, retarded achievement of development milestones such as walking, later deficits in cognitive ability, reduced school performance, increased child morbidity and mortality, higher risk of developing chronic diseases, impaired fat oxidation which stimulates the development of obesity, small stature later in live and reduced productivity and chronic poverty in adulthood. Not just stuntedness matters: throughout the distribution height has a postive effect on completed years of schooling (see, e.g., Alderman et al. (2006)) and cognitive and non-cognitive abilities (see, e.g., Case and Paxson (2008) and Schick and Steckel (2010)) and thereby on earnings. Hence we treat our two measures of malnutrition as dichotomous and continuous variables; we focus 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 (ROW) if their standardized BMI is larger than 1.15 [8]. This cut-off value is such that in a reference population 15 % is at risk of being overweight. Overweight children have delayed skill acquisition at very young ages (Cawley and Spiess (2008)), are more likely to have psychological or psychiatric problems, increased cardiovascular risk factors, increased incidence of asthma and diabetes (Reilly et al. (2003)), are more likely to be obese as adults (Serdula et al. (1993)) and earn possibly lower wages (Cawley (2004)). A final health outcome is based on the parents reported number of days that the child was sick during the previous 4-week period. Here we consider the percentage of children for which zero days and for which more than 3 days were reported. Table 2 provides information on the outcome variables of the control and treatment samples.

Looking at "All" households it is striking that the different entries are similar for all health outcomes in the contol and treatment sample, except for days sick: fewer sick days were

[8]The incidence of underweightedness is lower than in a reference population.

reported for children in the treatment sample than in the control sample. About one child in four is anemic and one in three is stunted. Compared to a reference population, our sample contains far too many stunted children and too many children at risk of being overweight.

When looking at the distribution of health outcomes over the types, unsurprising but nevertheless interesting patterns emerge[9]. Comparing IP with NP and IL with NL, we see that, except for the risk of being overweight in the treatment sample, indigenous children have worse health outcomes than non-indigenous children. The differences are substantial, especially for hemoglobin concentration and standardized height in the control sample. Comparing IP with IL and NP with NL, we see that the differences between children that had at least one parent that completed primary education and children whose parents have less than primary education are less outspoken. The largest differences occur for standardized height, where having a parent that completed primary education is a clear advantage. Overall, these results are very much in line with the literature, see, e.g., Backstrand et al. (1997), Rivera and Sepúlveda (2003), Rivera et al. (2003), Fernald and Neufeld (2006) and González de Cossío et al. (2009).

4 Empirical Results

4.1 Comparison of weighted treatment and control types

As stated at the end of section 2, a crucial assumption to identify treatment effects on the basis of a simple comparison of the outcomes of treatment and control samples is that $f^C(x \mid c_1) = f^T(x \mid c_1)$, implying that the two samples must be similar in terms of pre-program characteristics. If that is the case, after conditioning on c_1, observing x does not provide any information on whether an observation belongs to the treatment or control sample. We test this hypothesis as follows.

We define a sample containing those that belong to both the control and treatment sample. Next we perform a logistic regression, where the dependent variable takes the value 1 if the observation belongs to the control sample and the value 0 if it belongs to the treatment sample. Explanatory variables are characteristics of the family, the family's dwelling characteristics, possession of assets and the state of residence (see appendix 2 for more details). All these characteristics are measured in 1997, before the program started[10]. 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, such that the hypothesis that treatment and control samples are comparable in terms of composition of pre-program characteristics has to be rejected.

In the identification of average treatment effects, a standard way to deal with differences in composition of treatment and control sample is to use propensity score matching techniques. The idea is to make the treatment and control sample more comparable by weighting different observations with weights that depend on the estimated probability that the observation belongs to the control sample, as estimated by the logistic regression discussed in the previous paragraph. Appendix 3 explains this procedure and how the weights are used to obtain

[9]Keep in mind, however, that the types might differ in terms of characteristics that do not enter the definition of type and in terms of pre-program characteristics.

[10]Remember that for the control sample this is based on recall data -see also footnote 7.

Table 3: Difference between control and treatment in fraction of anemic, stunted, at risk of being overweight 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**
IP	-0.17**	-0.17**	-0.08	0.09	-0.06
IL	-0.05	-0.18*	-0.11**	0.10*	-0.05*
NP	-0.08**	0.05	0.03	0.07	-0.09**
NL	0.00	-0.01	-0.04	0.06	-0.02

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education;NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) "*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

estimates of the relevant distribution functions. Appendix 4 provides the equivalent of table 2 for the weighted (matched) samples.

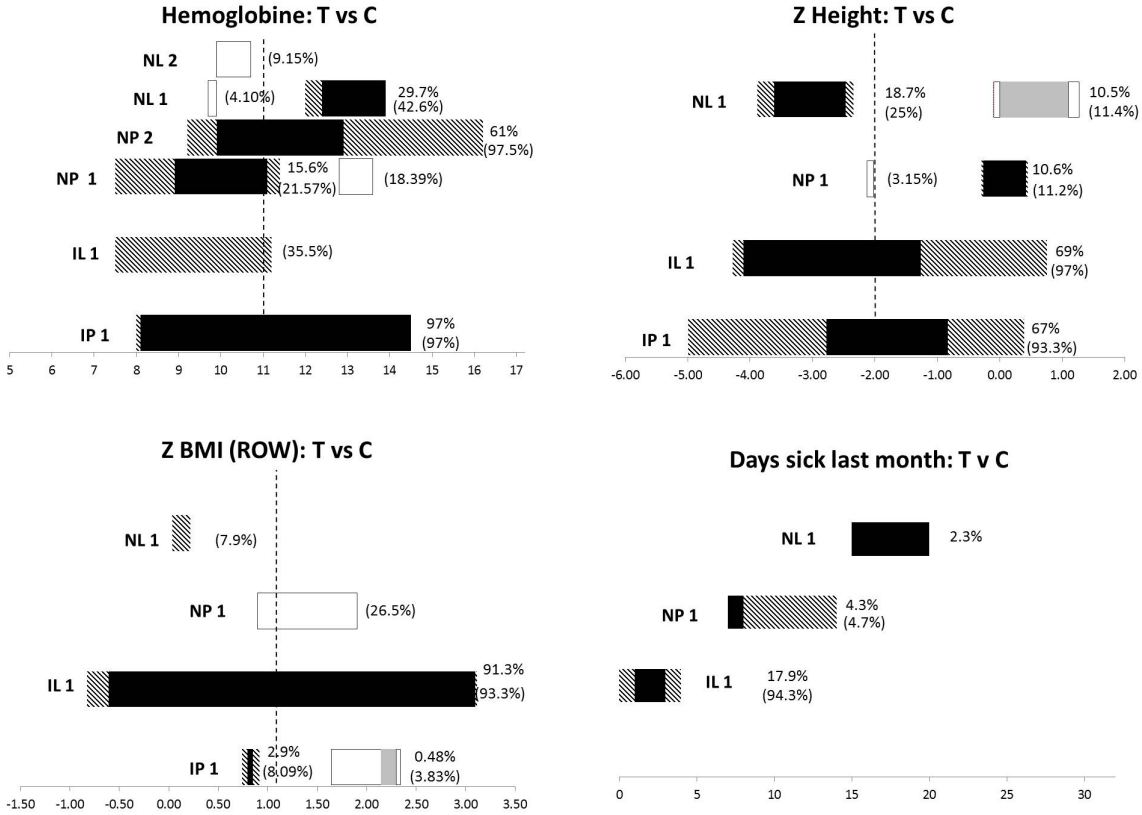
In table 3 we use the weighted samples to look at the effect of the treatments on the fraction of anemic, stunted, at risk of being overweight and the fraction of children for which zero sick days or more than three sick days during the last 4 weeks were reported. Effects that are statistically significantly different from zero at five (ten) percent are indicated by two (one) "*". Each entry gives the effect of the treatment. Remember that there are two interpretations possible. If children are responsible, a desirable effect on these fractions means that less genetic luck allows them to escape the bad condition of being anemic, stunted, at risk of being overweight, being sick or more than three days sick. If children are not responsible, a desirable effect simply means that treatment manages to avoid the bad conditions of being anemic, stunted, at risk of being overweight, being sick or more than three days sick for more children of a particular type.

We see that the treatment effects reported in table 3 are quite 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 quite well, especially for the children form indigenous origin that had no parent who completed primary education, which is, as table 2 suggests, probably the worst-off type. Also children from indigenous origin with a parent that completed primary education see an improvement in all indicators, although the effects are only significant for the fraction of anemic and stunted children. For non-indigenous children the results are less outspoken. The fraction of anemic non-indigenous children decreases thanks to the program, but no other significant treatment effects for non-indigenous children are identified in table 3.

Figure 1 presents the results of the stochastic dominance tests, using the procedure explained in section 2[11]. 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 non-dominance is rejected at 5 percent level of significance in favor of an

[11]Due to the many zero observations at zero, 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.

Figure 1: Stochastic dominance results.



(un)desirable 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=IP, IL, NP or NL), and the last character "i" indicates whether the test refers to first (i=1) or second (i=2) order stochastic dominance. The numbers (in parenthesis) behind the boxes give the percentage of observations of the treated type within the black or grey (hatched or white) box.

Take the top left panel in figure 1. The solid black box labelled "IP1" shows that for the IP types, using a 5% level of significance, the null that the cumulative distribution of the treatment type does not first order stochastically dominates the distribution of the control type has to 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 [8.1,14.5], which contains 97% of the treated IP type. When we increase the level of significance to 10%, the hatched box shows that the rejection interval enlarges only marginally to [8.0,14.5]. For the IL types, the null hypothesis of non-dominance can only be rejected at 10%, so we tested the null hypothesis of 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 5% level of significance. We failed to reject

the null, such that no box "IL2" is drawn. For the NP types we have first a solid black and then a white box. The latter is only significant at 10 % and occurs at a less important part of the distribution (above 11, where children are no longer anemic). Moreover, testing for second order stochastic dominance, we see a solid black box labelled "NP2", showing that the project leads to a second order improvement. Hence this type is also positively affected by the program. Finally, for the NL type, testing for first order stochastic dominance, we find a white box over the small range [9.7,9.9] containing very few observations of the treatment type, a solid black box further up in the distribution and, testing for second order stochastic dominance, a small white box. On balance, the evidence for this type again'st treatment is not strong.

The other panels in figure 1 can be interpreted similarly. 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 (standardized height below -2) parts of the support. For the NL type we find a first order stochastic dominance effect in favor of the treatment in an important part of the distribution (standardized height below -2) and an adverse effect higher up in the distribution. There is evidence of a marginal (at 10%) perverse first order treatment effect on standardized height for the NP type over a very small range [-2.11,-2.00] containing only 3% 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 non-indigenous types. In the bottom left panel we concentrate on what happens at the right of the dotted vertical line, which are the children at risk of being overweighted. We see positive first order stochastic dominance effects at 5% level of significance for the IL type and some evidence of marginally significant perverse treatment effects for IP and NP types. The bottom right panel shows first order improvements for the NL, NP type for the IL type. The intervals reported here, except IL, contain few observations. This is due to the high frequency of zero reported sick days (see table 2).

The results in table 3 and figure 1 are consistent. The stochastic dominance results provide more detail and pick up effects in important parts of the distribution that would go unnoticed otherwise, such as the positive first order stochastic dominance effect on standardized height for NL children. If first order improvements cannot be found, then, if genetic luck is a compensation characteristic, in principle, second order stochastic dominance provides a way to determine whether the program has positive effects. This rarely happens in our application: the only case is for hemoglobin concentration of the NP type. In summary, we find strong evidence of positive treatment effects for children from indigenous origin, especially for those without a parent that completed primary education. The evidence for childen from non-indigenous origin is less strong, but, on balance, if anything, also for these children enrollment in the program seems to have positive effects on their health opportunities.

4.2 Comparison to previous studies

Diaz and Handa (2006) use propensity score matching (PSM) techniques to construct alternative control samples from the Mexican national household survey. They compute average treatment effects by comparing the immediate treatment sample after 8 months of receiving program benefits with, at the one hand, the delayed treatment sample (which did not yet receive benefits) and, at the other hand, their newly constructed control samples. They

conclude that "The PSM technique requires an extreemly 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 was constructed following as closely as possible the beneficiary selection process. Our selection of covariates for the PSM follows closely Behrman et al. (2009b). Behrman and Parker (2010) use almost identical covariates and compare the effects on schooling outcomes of the short run differential exposure (between the immediate and delayed treatment sample) with the long run differential exposure (between the immediate treatment and control sample) and found that longer exposure produces larger effects, and the differences between the order of magnitude of the short and long run effects was reasonable. This suggests that the PSM technique we used can produce reliable estimates of average treatment effects.

The interpretation of the difference between the distributions of the weighted treatment and control sample as a treatment effect crucially depends on the extent to which the weighting procedure manages to correct for the, possibly unobserved heterogeneity due to the imperfect randomness of the assignment to treament and control. It is, of course not possible to test this directly, but we can compare our results to the findings in the literature that look at differences in children's health outcomes between the immediate and delayed treatment sample. Rivera et al. (2004) compare the health outcomes of children younger than 12 months old in 1997. They found that in 1999, after 12 months of treatment, children in the immediate treatment sample had higher mean hemoglobin values than the children from the, up to that point untreated, delayed treatment sample. When the immediate treatment sample received 24 months of treatment, and the delayed treatment sample received about 6 months of treatment, children from the immediate treatment sample had grown more than those of the delayed treatment sample and the differences in height were significantly larger for households with low socio-economic status (a score based on dwelling characteristics, posession of durable goods and access to water and sanitation). Gertler (2004) finds similar results for children aged 0 to 35 months in 1997: " ... 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 these differences are statistically significant at the 1-percent level. Unfortunately he does not report whether the effect is different for different subgroups like our types. Hemoglobin levels, contrary to height are not observed before the program started. Therefore the results on hemoglobin, as opposed to the growth effects, do not control for child fixed effects. This is important, as pointed out by Behrman and Hoddinott (2005). They investigate the effect on child height of children aged 4-48 months when treatment started in August 1998. They find that, without the inclusion of child fixed effects, treatment has a significant negative impact on child height (children aged between 4 and 36 months), but, if controlled for child fixed effects (by looking at the difference between 1999 and 1998), the treatment effect becomes significantly positive, and is about 1 centimeter, like in Gertler (2004)[12]. Interestingly, program effects are larger for children in households where the household head speaks an indigeneous language and the mother is better educated.

We compare the health outcomes of immediate and delayed treatment in appendix D, for the

[12]Behrman and Hoddinott (2005) obtain the same pattern when looking at standardized height-for-age scores.

children born between the beginning of the initial and the beginning of the delayed treatment. This seriously limits the size of the sample. Moreover, as all these children received at least 3 years of treatment by the time their health outcomes were measured, few significant effects can be found, especially on hemoglobin concentration and reported days sick, showing that these variables are more sensitive to nutritional status in the immediate past than in the more distant past. We find 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 non-indigenous children with parental primary education over a limited support of the distribution. Again, therefor, the evidence is rather in favor of the program.

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 aged between 24 and 68 months in 2003, the time their 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 overweight. Neither whether the household head spoke an indigenous language, nor father's or mother's education were significant controls in their model.

Overall these findings are very much 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 where an indigenous language is spoken. This is compatible with Fernald et al. (2008), provided that indigenous families receive larger cash transfers than non-indigenous families, for instance because they have more children. What our results add is that we can visualize where in the distribution the program is most effective for the different types, and that the program is most powerful for the most disadvantaged types (the children from indigenous origin).

5 Conclusion

By now there is a growing literature on the measurement of inequality of opportunity. For an overview see, e.g., Ramos and Van de gaer (2011). So far, the ideas in this literature have not been applied to evaluate social programs. We propose a methodology to do so.

We bring together insights from three diferent literatures: the literature on equality of opportunity, on program evaluation and on testing for stochastic dominance. Roemer (1993)'s normative approach to equality of opportunity says that we should focus on types, and that, if responsibility characteristics are unobserved, those that are at the same percentile of the distribution of the outcome within their type have exercised a comparable degree of responsibility. This provides a normative foundation for the comparison of cumulative distribution functions of corresponding treatment and control types. The literature on program evaluation stresses that we should be careful that the treatment and control samples are comparable in terms of pre-program characteristics. If they are not, propensity score matching techniques can be used to make the samples more comparable. Given that we compare the distribution of corresponding treatment and control types, we test whether these sample are comparable in terms of pre-program characteristics and propose a weighted sampling method based on standard propensity score matching techniques to make the samples more comparable. Fi-

nally, Davidson and Duclos (2009) and Davidson (2009) proposed a new technique to test for stochastic dominance, taking non-dominance as the null, such that rejection of the null implies dominance. Their test procedure is particularly suited in our context, as it allows us to see where in the distribution dominance can be established.

We applied our procedure to study the effect of the Mexican *Oportunidades* program on children's health opportunities. Concerning the proposed methodology two conclusions can be drawn. First, in our application (just like 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 add much to the conclusions drawn from first order stochastic dominance. Hence, it does not matter much for the conclusions whether children are considered responsible for their genetic make-up or not. Second, the treatment and control samples differed substantially in terms of pre-program characteristics. It is therefor important to use weighted sampling based on techniques such as propensity score matching that make the samples (more) comparable. Concerning the actual effects of the program, our results indicate that the *Oportunidades* program has a substantial favorable impact on the health opportunities of the most disadvantaged children from poor parents, children of indigenous origin without a parent that completed primary education. The effect on children from indigenous origin with a parent that completed primary education are still sizeable and important. The effect on non-indigenous children growing up in poor households is less outspoken, but, overall, if anything, the evidence in this paper indicates that for them also the program results in better health opportunities.

References

Alderman, H., J. Hoddinott, and B. Kinsey (2006). Long term consequences of early childhood malnutrition. *Oxford Economic Papers 58*(3), 450–474.

Aslam, M. and G. Kingdon (2010). Parental education and child health-understanding the pathways of impact in Pakistan. *RECOUP Working Paper No. 30*.

Backstrand, J., L. Allen, G. Pelto, and A. Chávez (1997). Examining the gender gap in nutrition: An example from rural Mexico. *Social Science & Medicine 44*(11), 1751–1759.

Behrman, J. and J. 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), 547–569.

Behrman, J. and S. Parker (2010). Do contional cash transfers generate lasting benefits? a five year follow-up of PROGRESA/Oportunidades. *mimeo*.

Behrman, J., S. Parker, and P. Todd (2009a). Medium-term impact of the oportunidades conditional cash transfer program on rural youth in Mexico. In S. Klasen and N. F (Eds.), *Poverty, Inequality and Poverty in Latin America*, pp. 219–270. Cambridge: MIT Press.

Behrman, J., S. Parker, and P. Todd (2009b). Schooling impacts of conditional cash transfers on young children: Evidence from Mexico. *Economic Development and Cultural Change 57*(3), 439–477.

Behrman, J., P. Sengupta, and P. Todd (2005). Progressing through PROGRESA: an impact assessment of a school subsidy experiment in rural Mexico. *Economic Development and Cultural Change 54*(1), 237–275.

Behrman, J. and P. Todd (1999). Randomness in the experimental samples of PROGRESA –Education, Health, and Nutrition Program. *International Food Policy Research Institute*.

Behrman, J., P. Todd, B. Hernández, J. Urquieta, O. Attanasio, M. Angelucci, and M. Hernández (2006). *Evaluación externa de impacto del programa Oportunidades 2006*. Instituto Nacional de Salud Pública.

Black, S., P. Devereux, and K. 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 –439.

Bossert, W. (1995). Redistribution mechanisms based on individual characteristics. *Mathematical Social Sciences 29*(1), 1–17.

Bourguignon, F., F. Ferreira, and M. Menéndez (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth 53*(4), 585–618.

Branca, F. and M. Ferrari (2002). Impact of micronutrient deficiencies on growth: The stunting syndrome. *Annals of Nutrition and Metabolism 46*(Suppl. 1), 8–17.

Breierova, L. and E. Duflo (2003). The impact of education on fertility and child mortality: Do fathers really matter less than mothers? *OECD Development Centre Working Papers. OECD Publishing 217*.

Case, A. and C. Paxson (2008). Stature and status: Height, ability and labor market outcomes. *Journal of Political Economy 116*(3), 499–532.

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources 39*(2), 451–474.

Cawley, J. and C. Spiess (2008). Obesity and skill attainment in early childhood. *Economics and Human Biology 6*, 388–397.

Chen, W. and J. Duclos (2008). Testing for poverty dominance: an application to Canada. *IZA Discussion Paper N 2829*.

Davidson, R. (2009). Testing for restricted stochastic dominance: Some further results. *Review of Economic Analysis 1*(1), 34–59.

Davidson, R. and J. Duclos (2009). Testing for restricted stochastic dominance. *GREQAM Document de Travail 2009-38* (06-09).

de Barros, R., F. Ferreira, J. Molinas Vega, and J. Chanduvi (2009). *Measuring Inequality of Opportunities in Latin America and the Caribbean*. The World Bank.

Desai, D. and S. Alva (1998). Maternal education and child health: Is there a strong causal relationship? *Demography 35(1)*, 71–81.

Diaz, J. and S. 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), 319–345.

Fernald, L., P. Gertler, and L. 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), 828–837.

Fernald, L. and L. Neufeld (2006). Overweight with concurrent stunting in very young children from rural Mexico: prevalence and associated factors. *European Journal of Clinical Nutrition 61*(5), 623–632.

Ferreira, F. H. G. and J. Gignoux (2011). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth forthcoming*.

Fiszbein, A., N. Schady, F. Ferreira, M. Grosh, N. Kelleher, P. Olinto, and E. Skoufias (2009). Conditional Cash Transfers: Reducing present and future poverty , a World Bank policy research report. *The World Bank, Washington*.

Fleurbaey, M. (1995). The requisites of equal opportunity. In M. Salles and N. Schofield (Eds.), *Social Choice, Welfare and Ethics*, pp. 37–53. Cambridge University Press.

Fleurbaey, M. (1998). Equality among responsible individuals. In J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy (Eds.), *Freedom in Economics: New Perspectives in normative economics*, pp. 206–234. London: Routledge.

Fleurbaey, M. (2008). *Fainess, Responsibility and Welfare*. Oxford: Oxford University Press.

Foster, J., J. Greer, and E. Thorbeke (1984). A class of decomposable poverty measures. *Econometrica 52*(3), 761–766.

Gertler, P. (2004). Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *American Economic Review 94*(2), 336–341.

González de Cossío, T., J. Rivera, D. González-Castell, and E. Monterrubio (2009). Child malnutrition in Mexico in the last two decades: prevalence using the new WHO 2006 growth standards. *Salud Pública de México 51*.

Grantham-McGregor, S. and C. Ani (2001, February). A review of studies on the effect of iron deficiency on cognitive development in children. *The Journal of Nutrition 131*(2), 649S –668S.

Heckman, J. (1992). Randomization and social policy evaluation. In C. Manski and I. Garfinkel (Eds.), *Evaluating Welfare and Training Programs*, pp. 201–230. Cambridge: Harvard University Press.

Heckman, J., S. Jeffrey, and N. 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*. INSP2005.

Lefranc, A., N. Pistolesi, and A. Trannoy (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? *Review of Income and Wealth 54*(4), 513–546.

Lefranc, A., N. Pistolesi, and A. 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, F. and C. Barrett (2010, October). A stochastic dominance approach to program evaluation with an application to child nutritional status in kenya. Working Paper.

Nozick, R. (1977). *Anarchy, State and Utopia*. Basic Books.

Olaiz, G., J. Rivera, T. Shamah, R. Rojas, S. Villalpando, and M. Hernández (2006). Encuesta Nacional de Salud y Nutrición 2006 [National Health and Nutrition Survey 2006]. *Instituto Nacional de Salud Pública*.

O'Neill, D., O. Sweetman, and D. Van de gaer (2000). Equality of opportunity and Kernel density estimation: An application to intergenerational mobility. In T. Fomby and R. C. Hill (Eds.), *Advances in Econometrics*, Volume 14, pp. 259–274. Stanford: JAI Press.

Peragine, V. and L. Serlenga (2008). Higher education and equality of opportunity in Italy. In J. Bishop and B. Zheng (Eds.), *Inequality of opportunity: papers from the Second ECINEQ Society Meeting, Research on Economic Inequality*, Volume 16, pp. 67–97. Bingley: Emerald Group Publishing.

Psacharopoulos, G. and H. Patrinos (1994). Indigenous people and poverty in Latin America. *The World Bank*.

Ramos, X. and D. Van de gaer (2011). Empirical approaches to inequality of opportunity: principles, measures and evidence. *Mimeo. Faculty of Economics and Business Administration, Ghent University.*

Rawls, J. (1971). *A Theory of Justice.* Oxford: Oxford University Press.

Reilly, J., E. Metheven, Z. McDowell, B. Hacking, D. Alexander, L. Stewart, and C. Kelnar (2003). Health consequences of obesity. *Archives of desease in childhood 88*(9), 748–752.

Rivera, J., E. Monterrubio, T. González-Cossío, A. García-Feregrino, A. García-Guerra, and J. 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–S476.

Rivera, J. and J. Sepúlveda (2003). Conclusions from the Mexican National Nutrition Survey 1999: Translating results into nutrition policy. *Salud Pública de México 45*, S565–S575.

Rivera, J., D. Sotres-Alvarez, J. Habicht, T. Shamah, and S. 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 –2570.

Roemer, J. (1993). A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs 22*(2), 146–166.

Roemer, J. (1998). *Equality of opportunity.* Cambridge MA: Harvard University Press.

Rosa Dias, P. (2009). Inequality of opportunity in health: evidence from a UK cohort study. *Health Economics 18*(9), 1057–1074.

Schick, A. and R. Steckel (2010). Height as a proxy for cognitive and Non-Cognitive ability. *NBER Working Paper N 16570*.

Schultz, P. (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, M., D. Ivery, R. Coates, D. Freedman, D. Williamson, and T. Byers (1993). Do obese children become obese adults? a review of the literature. *Preventive Medicine 22*, 167–177.

Silverman, B. W. (1986). *Density Estimation for Statistics ad Data Analysis.* London: Chapman & Hall.

Todd, P. (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, A., T. Sandy, F. Jusot, and M. Devaux (2010). Inequality of opportunities in health in France: a first pass. *Health Economics 19*(8), 921–938.

Verme, P. (2010). Stochastic dominance, poverty and the treatment effect curve. *Economics Bulletin 30*(1), 365–373.

A Appendix

Appendix 1: sampling procedure

Table A.1: Sampling process.

	Original number of children (a)	matched children number (b)	% of (a)	1997 data available number	% of (b)
C	2247	1871	83	1871	100
T	2615	2200	84	1128	51
Total	4862	4071	84	2999	73

Note: the acronyms refer to samples : C = Control sample; T = Treatment sample.

Comparing the sample sizes in the column "1997 data available" with those in table 1 in the main text, one observes that 12 (3) observations dropped out in the final control (treatment) sample. This is due to missing observations on circumstance characteristics.

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 1 if the observation comes from the control sample and 0 otherwise. Explanatory variables are based on pre-program characteristics of the treatment sample and recalled 1997 characteristics of the control sample. We have five kinds of explanatory variables.

(1) Household characteristics: the ages of the household head and spouse (in years), sex of household head, whether the household head and spouse speak an indigenous language, whether they completed primary education, whether they worked, and composition of the household (number of children, women and men of different ages).

(2) Dwelling conditions of the household: number of rooms in the house and a list of dummy variables indicating the presence of electrical light, running water on the property, running water in the house (which implies of course presence of running water on the property), a dirtfloor and whether the roof and wall were of poor quality.

(3) Asset information: dummy variables indicating whether the family owned animals or land and whether the family possessed a blender, fridge, fan, gas stove, gas heater, radio, hifi, TV, video, washing machine, car or truck.

(4) State of residence: a list of dummy variables indicating the state where the family lived. The reference state (all state of residence dummies equal to zero) is Veracruz.

(5) Following Behrman et al. (2009b) and Behrman and Parker (2010), we include dummy variables for missing characteristics, provided they could be meaningfully estimated. 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.

Table A.2: Logistic regression results.

Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	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
IndigHhHead	-0.718	0.272	-2.64	Gas stove	0.377	0.145	2.60
IndigSpouse	0.249	0.278	0.90	Gas heater	0.709	0.360	1.97
EducHhHead	-0.229	0.114	-2.01	Radio	-0.600	0.100	-5.96
EducSpouse	-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	0.118	-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	Wash 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	Michoacan	-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	Queretaro	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 Indig HH	0.799	1.959	0.41
Electrical light	0.036	0.115	0.32	Miss Indig 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
Dirtfloor	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	0.4223	9.13
Land	-0.545	0.105	-5.17				
Number of Obs		2741					
LR Chi2 (54)		730.0		Pseudo R2		0.198	
Prob>Chi2		0.000		Log Likelihood		-1478.75	

Appendix 3: matching estimator and construction of the corresponding distribution function.

Table A.3: Propensity score matching: common support and number of observations in the common support.

	Common support	Control #	Treatment #	Band-width
IP	[0.158,0.957]	155	193	0.074
IL	[0.106,0.868]	228	260	0.074
NP	[0.063,0.949]	668	318	0.071
NL	[0.017,0.952]	586	318	0.071
Total		1637	1089	

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

STEP1 : 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 pre-program characteristics x_i. Figure A.1 depicts the estimated propensity scores. As we matched for each of the 4 combinations of race and parental level of education the treatment into the control sample, we determined the common support for each of these four comparisons as the overlap of the supports of the control and treatment sample. 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 test. In 14 cases the balancing property was rejected. As an additional test, we rerun the logistic equation from table A.2 using the weighted sample. Only four coefficients out of 54 turned out to be significant. These results are quite ecouraging.

STEP 2: Construction of the cumulative distribution function.

Let I_1 denote the set of individuals in the treatment sample, I_0 the set of individuals in the control sample and S_P 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 don't 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 then becomes

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} \left[E\left(Y_{1j} \mid D = 1, P_j\right) - Y_{0j} \right],$$
$$\text{with } E\left(Y_{1j} \mid D = 1, P_j\right) = \sum_{i \in I_1} W\left(i, j\right) Y_{1i}.$$

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

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

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

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

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 ω_i to construct the cumulative distribution function.
There exist many possible ways to determine the weights $W\left(i, j\right)$. We use a Kernel estimator, such that

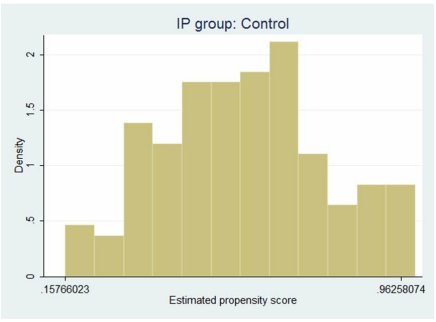
$$W\left(i, j\right) = \frac{G\left(\frac{P_i - P_j}{\alpha}\right)}{\sum_{k \in I_1} G\left(\frac{P_k - P_j}{\alpha}\right)},$$

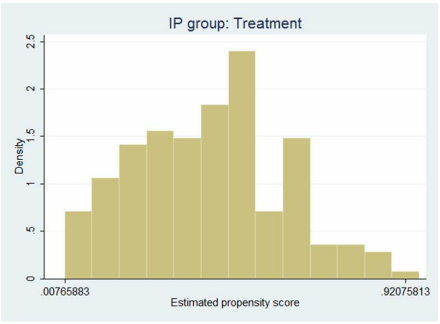
where $G\left(.\right)$ is the Epanechnikov kernel function and α is a bandwidth parameter. The bandwidth parameter was chosen in an optimal way, using the formula in Silverman (1986), page 45-47:

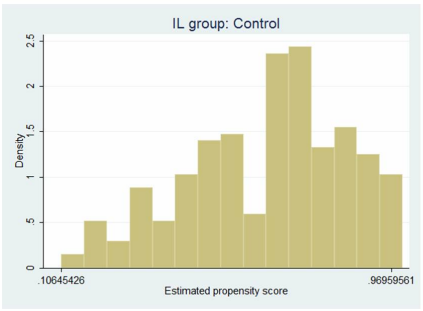
$$\alpha = 1.06 \min\left(\sigma, \frac{\rho}{1.34}\right),$$

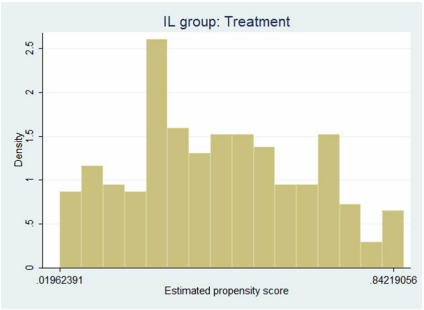
where σ is the standard deviation and ρ the interquartile range of the distribution of propensity scores. The resulting bandwiths for each of the types are given in the last column of table A.3.

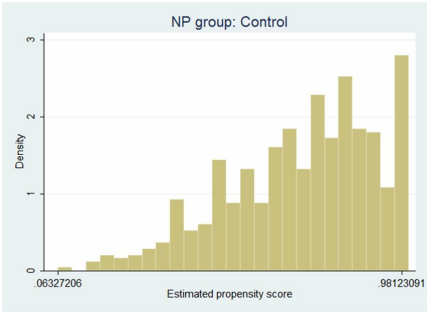
Figure A.1: Estimated propensity scores.

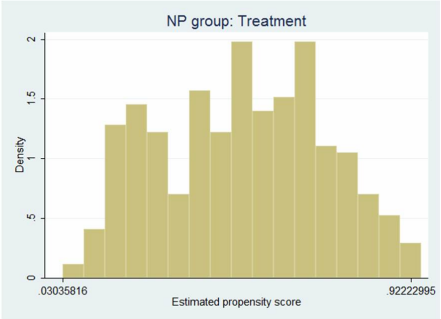


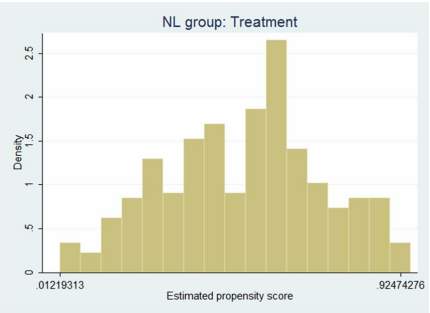












Appendix 4: treatment and control effects in matched samples

Table A.4: Health outcomes of 2-6 year old children in 2003.

(a) Control sample

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

(a) Treatment sample

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

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

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.

Appendix 5: 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.[13] 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 Chen and Duclos (2008). Our samples contain $C_1^X, \ldots, C_c^X, \ldots, C_{n^X}^X$ clusters (villages), $X = C, T$. Each cluster in the sample contains n_c^X children $(c = 1, \ldots, n^X)$. We mimic this sample design as follows. First, define for each cluster

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

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 π_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

[13] 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).

get an idea of the distribution of the minimal LR under the frontier of the null hypothesis. 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 $[\widehat{r}^{-}, \widehat{r}^{+}]$ for which the hypothesis

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

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

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

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

Sensitivity Analysis

In addition to the results discussed at length in the main part of the paper, we present three sensitivty analyses by modifying inclusion criteria to the program and by modifying the definition of parental education. In the analysis presented in sections 3 and 4, our base case, two conditions were necessary for inclusion to the treatment group: (i) the household should be part of a treatment community (communities were the program was operating) and (ii) information on monetary transfers received by the household should be available. Children's types were defined on the basis of indigenous origin and whether at least one parent completed primary education or not.

The analysis in appendix B that follows incorporates *all* children living in treatment communities independently of whether information on transfers received by the household were available or not. As a result, the treatment sample for this analysis contains 219 additional observations, as can be seen upon simple comparison of tables 1 (in the main text) and B1 (in appendix B).

Comparing the results in tables 3 and B4, it is striking that all estimated program effects have the same sign. Most significantly estimated effects in table 1 also turn up significant in table B4 and the other way around, with few exceptions. All significantly estimated effects in both tables 3 and B4 are in favor of the program. Looking at the stochastic dominance results in figure B.2 we find very similar arrangements as in the base case in figure 1. Indigenous children seem to benefit most from receiving *Oportunidades*, although the effect now is somewhat weaker for indigenous children without parental primary education background (IL group) and stronger for indigenous children with parental primary education background (IP group). Except for the negative effect observed on standardized BMI for non-indigenous children with parental primary education, the effect on non-indigenous is similar to the base case. Overall, the effects are very close to the effects in the base case and very much in favor of the program.

The analysis in appendix C changes the definition of type. Althoug the contribution of parental education to child health is generally recognised, education effects of both parents separately are still disputed (Breierova and Duflo (2003), Aslam and Kingdon (2010)). In particular, it has been suggested that education of the mother could have a major influence on child well-being (Desai and Alva (1998)). Based on this hypothesis, appendix C defines types on the basis of indigenous background (as before) and on whether the *mother* has completed primary education or not. Table C1 shows that this diminishes the sizes of both control and treatment samples compared to the base case (table 1). This is due to the fact that in the base case, some observations for which mother's educational level was not but father's educational level was observed, could be classified as having at least one parent that completed primary education.

The comparison of tables 3 and C4 reveals that all estimated program effects have the same sign (except the effect on the fraction of anemic children for the NL group, which changes from being marginally positive to -0.02). Most significantly estimated effects in table 1 also turn up significant in table C4 and the other way around, with only few exceptions. All significantly estimated effects in table C4 are in favor of the program. The stochastic dominance tests in figure C2 show the same pattern as in figure 1. The most noteworthy difference is that the positive effects on hemoglobin concentration and standardized height of the IL group become

less pronounced. From this sensitivity analysis, we conclude again that, overall, the effects are very close to the effects in the base case and very much in favor of the program.

The analysis in appendix D compares the effect of *Oportunidades* between the *immediate and delayed treatment* groups. As mentioned in section 3.2, the original sample design followed a random procedure to allocate the treatment to two comparable groups. One group received the program immediately (original treatment) while the other was phased-out 18 months in order to operate as control (delay treatment). Lack of information on the amount of transfers for the original treatment motivated the use of the latter for the main analysis. Here we aim at assessing the effect of having been exposed longer to the program, by comparing the health outcomes of children in the original and delayed treatment. The main advantage is the randomization of households over these two groups. In order to make the comparison meaningfull, we limit the sample to children that were born after april 1998 (when the original treatment started) and before october 1999 (when delayed treatment started). As can be seen in table D1, this decreases the number of observations that can be used drastically.

The logistic regression in table D5 reports much fewer significant coefficients than the regressions in tables A2, B5 and C5. This is due to the much better randomization of households between initial and delayed treatment and the smaller sample size. Table D4 shows the limitation of the exercise: it shows only one positive treatment effect: for indigenous children with a mother that completed primary education, the fraction of children reporting zero sickdays increased by 18 percent. Also the stochastic dominance tests find fewer significant effects, especially on hemoglobin concentration and days sick. The reason for this is probably that both the children in the initial and delayed treatment samples received the program during the three years preceding the collection of the health data in 2003, and these two health indicators are more influenced by what happens during the period immediately before they are measured. What is quite remarkable, however is the substantial impact on standardized height of longer exposure to the program, exposure in womb and during the first months of life. Here we do find significant effects for children from indigenous origin without a parent that completed primary education and for non-indigenous children with a parent that completed primary education. Hence, we conclude that program exposure at a very young age can have significant positive effects on standardized health three years later.

B Entire delayed treatment group versus Control

Table B.1: Composition of the samples (delay vs control).

	Control sample		Treatment sample	
	#	%	#	%
All	1859	100	1344	100
IP	173	9.3	227	16.9
IL	241	13.0	329	24.5
NP	824	44.3	395	29.4
NL	621	33.4	393	29.2

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

Table B.2: Health outcomes of 2-6 year old children in 2003 (delay vs control).

(a) Control sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.36	11.6	0.50	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.40	0.30	0.64	0.13
NP	0.18	12.2	0.20	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.47	0.25	0.58	0.18

(b) Treatment sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.23	12.1	0.33	-1.53	0.20	0.66	0.12
IP	0.27	12.0	0.36	-1.70	0.14	0.62	0.13
IL	0.29	11.7	0.45	-1.87	0.18	0.70	0.10
NP	0.14	12.5	0.24	-1.17	0.25	0.67	0.11
NL	0.26	12.2	0.30	-1.47	0.22	0.64	0.14

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

Table B.3: Health outcomes of 2-6 year old children in 2003 (delay vs control): Matched samples.

(a) Control sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.18
IP	0.37	11.6	0.49	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.37	0.30	0.64	0.13
NP	0.18	12.2	0.19	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.46	0.25	0.58	0.18

(a) Treatment sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.22	12.0	0.33	-1.51	0.19	0.67	0.11
IP	0.20	12.0	0.29	-1.54	0.13	0.67	0.10
IL	0.29	11.7	0.48	-1.90	0.18	0.72	0.09
NP	0.11	12.4	0.24	-1.12	0.25	0.62	0.11
NL	0.28	12.2	0.32	-1.45	0.20	0.65	0.15

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

Table B.4: Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs control).

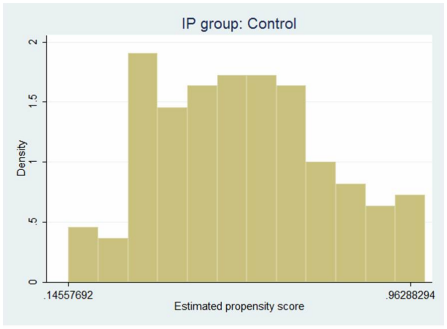
	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.02	0.01	-0.05*	0.09**	-0.06**
IP	-0.16**	-0.21**	-0.10	0.10	-0.08*
IL	-0.01	-0.16*	-0.12**	0.08	-0.05
NP	-0.07*	0.05	0.04	0.06	-0.07**
NL	0.03	-0.00	-0.05*	0.07	-0.03

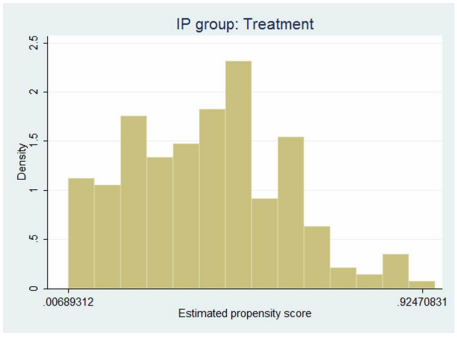
Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) "*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

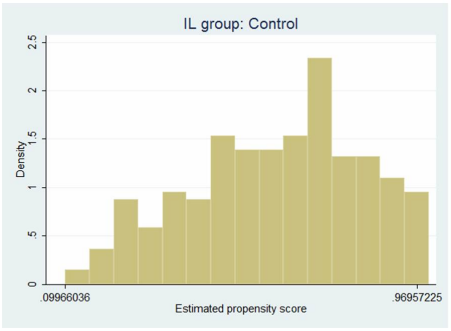
Table B.5: Logistic regression results (delay vs control).

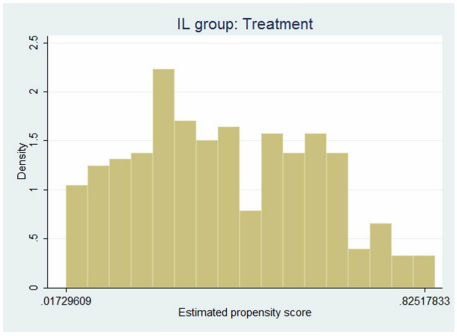
Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	-0.016	0.006	-2.44	Blender	-0.186	0.126	-1.48
Age spouse	-0.014	0.007	-2.12	Fridge	-0.032	0.184	-0.17
Sex Hh head	-2.32	0.335	-6.94	Fan	0.167	0.185	0.90
IndigHhHead	-0.600	0.255	-2.34	Gas stove	0.276	0.136	2.03
IndigSpouse	0.109	0.260	0.42	Gas heater	0.707	0.310	2.28
EducHhHead	-0.234	0.110	-2.13	Radio	-0.546	0.096	-5.67
EducSpouse	-0.487	0.110	-4.39	Hifi	-0.360	0.230	-1.56
Work Hh head	1.024	0.244	4.20	Tv	-0.646	0.114	-5.67
Work spouse	0.490	0.147	3.32	Video	0.412	0.293	1.41
# Children 0-5	-0.077	0.045	-1.71	Wash machine	-0.200	0.290	-0.69
# Children 6-12	-0.182	0.040	-4.58	Car	0.877	0.358	2.45
# Children 13-15	-0.156	0.080	-1.98	Truck	-0.243	0.236	-1.03
# Children 16-20	-0.085	0.067	-1.27	Guerrero	-0.841	0.171	-4.91
# Women 20-39	-0.126	0.105	-1.20	Hidalgo	-0.863	0.196	-4.40
# Women 40-59	-0.083	0.142	-0.59	Michoacan	-0.422	0.167	-2.52
# Women 60+	0.016	0.171	0.09	Puebla	-1.061	0.142	-7.43
# Men 20-39	-0.200	0.096	-2.07	Queretaro	0.290	0.207	1.40
# Men 40-59	0.420	0.152	2.76	San Luis	-0.460	0.144	-3.18
# Men 60+	0.757	0.221	3.42	Miss Age Sp	-4.65	0.712	-6.54
# Rooms	-0.006	0.010	-0.68	Miss Indig HH	0.911	2.08	0.44
Electrical light	0.083	0.110	0.75	Miss Indig Sp	-2.30	2.030	-1.13
Running water land	0.812	0.108	7.51	Miss Work HH	3.65	2.004	1.82
Running water house	-0.350	0.191	-1.84	Miss Work Sp	3.990	1.984	2.01
Dirtfloor	0.060	0.112	0.54	Miss Water land	1.090	1.601	0.68
Poor quality roof	-0.002	0.104	-0.02	Miss Water house	0.354	0.643	0.55
Poor quality wall	-0.380	0.123	-3.09	Miss Assets	-3.912	2.070	-1.89
Animals	-0.191	0.107	-1.78	Constant	4.232	0.406	10.42
Land	-0.505	0.100	-5.06				
Number of Obs		2959					
LR Chi2 (54)		815.6		Pseudo R2		0.201	
Prob>Chi2		0.000		Log Likelihood		-1624.23	

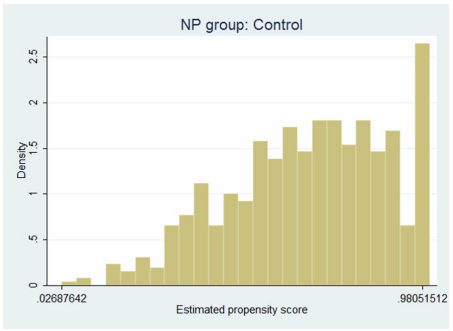
Figure B.1: Estimated propensity scores (delay vs control).

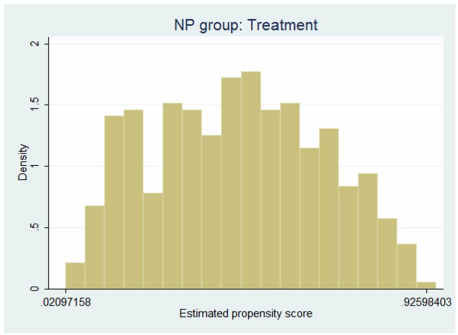












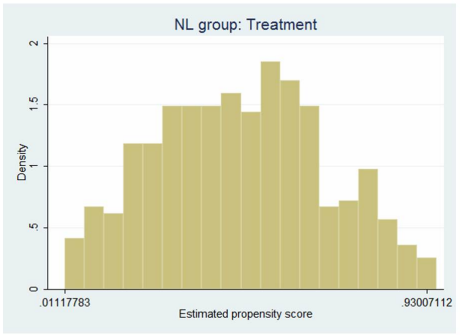


Figure B.2: Stochastic dominance results (delay vs control).

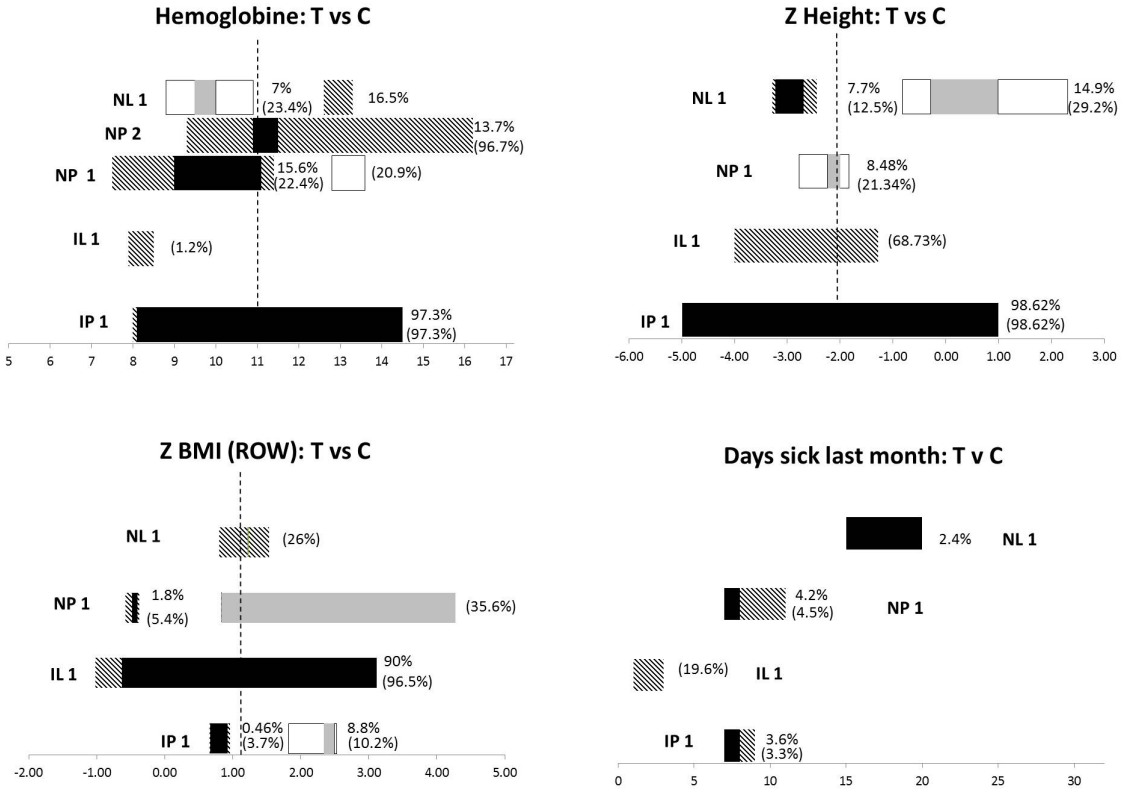


Table B.6: Propensity score matching: common support and number of observations in the common support (delay vs control).

	Common support	Control #	Treatment #	Band-width
IP	[0.145,0.959]	148	209	0.072
IL	[0.099,0.850]	191	312	0.069
NP	[0.027,0.950]	596	392	0.068
NL	[0.011,0.950]	551	390	0.069
Total		1486	1303	

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

C Mother's education as circumstance criterion

Table C.1: Composition of the samples (Mother's education case).

	Control sample		Treatment sample	
	#	%	#	%
All	1808	100	1079	100
IP	121	6.7	150	13.9
IL	278	15.4	310	28.7
NP	680	37.6	255	23.6
NL	729	40.3	364	33.8

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education.

Table C.2: Health outcomes of 2-6 year old children in 2003 (Mother's education case).

(a) Control sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.38	11.9	0.48	-1.88	0.24	0.56	0.18
IL	0.30	11.8	0.63	-2.36	0.28	0.63	0.14
NP	0.17	12.3	0.17	-1.04	0.23	0.57	0.17
NL	0.25	12.0	0.33	-1.52	0.23	0.58	0.18

(b) Treatment sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.22	12.1	0.33	-1.58	0.20	0.67	0.12
IP	0.31	11.9	0.34	-1.64	0.12	0.66	0.14
IL	0.27	11.8	0.43	-1.83	0.17	0.72	0.10
NP	0.10	12.6	0.23	-1.25	0.22	0.68	0.11
NL	0.23	12.2	0.32	-1.58	0.24	0.64	0.14

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education.

Table C.3: Health outcomes of 2-6 year old children in 2003 (Mother's education case): Matched samples.

(a) Control sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.38	11.9	0.48	-2.00	0.24	0.56	0.18
IL	0.30	11.8	0.63	-2.36	0.28	0.63	0.14
NP	0.17	12.3	0.17	-1.04	0.23	0.57	0.17
NL	0.25	12.0	0.33	-1.51	0.23	0.58	0.18

(a) Treatment sample

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.20	12.1	0.32	-1.47	0.19	0.67	0.11
IP	0.25	12.0	0.26	-1.46	0.12	0.66	0.17
IL	0.25	11.9	0.50	-1.97	0.17	0.74	0.07
NP	0.07	12.4	0.21	-1.09	0.23	0.61	0.11
NL	0.23	12.3	0.32	-1.46	0.22	0.66	0.12

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education.

Table C.4: Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (Mother's education case).

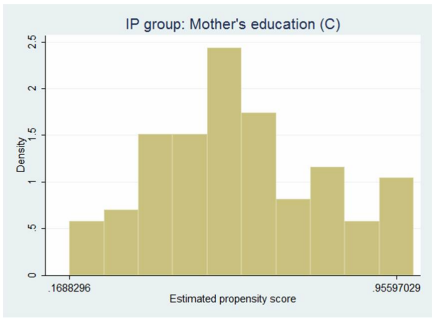
	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.04	0.00	-0.05*	0.09**	-0.06**
IP	-0.12	-0.21**	-0.12*	0.10	-0.02
IL	-0.05	-0.13	-0.12**	0.11**	-0.08**
NP	-0.10**	0.05	0.00	0.04	-0.06**
NL	-0.02	-0.00	-0.01	0.08	-0.07**

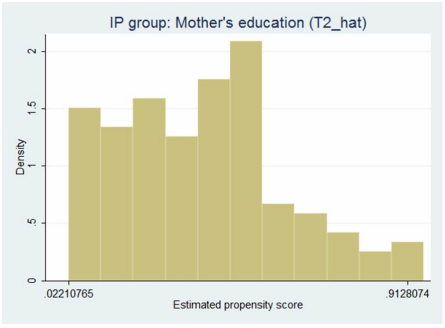
Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education. One (two) "*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

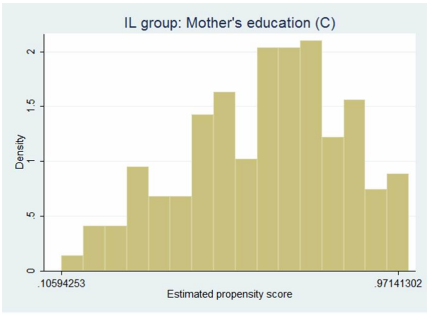
Table C.5: Logistic regression results (Mother's education case).

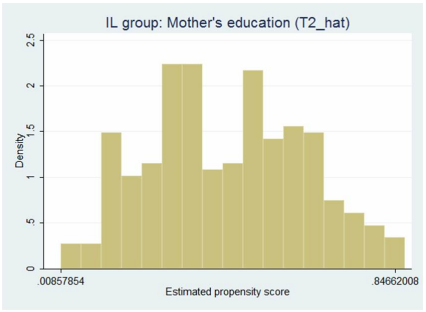
Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	-0.017	0.007	-2.40	Blender	-0.180	0.134	-1.34
Age spouse	-0.005	0.007	-0.63	Fridge	0.073	0.204	0.36
Sex Hh head	-2.354	0.380	-6.19	Fan	0.125	0.202	0.62
IndigHhHead	-0.691	0.288	-2.40	Gas stove	0.364	0.147	2.48
IndigSpouse	0.213	0.292	0.73	Gas heater	0.609	0.365	1.67
EducHhHead	-0.222	0.115	-1.92	Radio	-0.590	0.101	-5.79
EducSpouse	-0.398	0.117	-3.39	Hifi	-0.357	0.254	-1.41
Work Hh head	1.199	0.280	4.27	Tv	-0.626	0.120	-5.22
Work spouse	0.575	0.164	3.50	Video	0.572	0.354	1.62
# Children 0-5	-0.090	0.048	-1.84	Wash machine	-0.139	0.337	-0.41
# Children 6-12	-0.218	0.042	-5.13	Car	1.214	0.468	2.59
# Children 13-15	-0.144	0.085	-1.70	Truck	0.262	0.287	0.91
# Children 16-20	-0.218	0.042	-5.13	Guerrero	-0.535	0.191	-2.80
# Women 20-39	-0.008	0.121	-0.70	Hidalgo	-0.864	0.218	-3.96
# Women 40-59	0.036	0.157	0.23	Michoacan	-0.576	0.178	-3.23
# Women 60+	-0.000	0.189	-0.00	Puebla	-1.103	0.151	-7.29
# Men 20-39	-0.178	0.107	-1.65	Queretaro	0.108	0.222	0.49
# Men 40-59	0.036	0.157	0.23	San Luis	-0.441	0.155	-2.83
# Men 60+	0.682	0.241	2.83	Miss Age Sp	-3.85	0.723	-5.33
# Rooms	-0.005	0.010	-0.55	Miss Indig HH	0.649	1.919	0.34
Electrical light	0.059	0.116	0.51	Miss Indig Sp	-2.015	1.850	-1.09
Running water land	0.844	0.116	7.26	Miss Work HH	3.510	1.827	1.92
Running water house	-0.412	0.209	-1.97	Miss Work Sp	3.733	1.795	2.08
Dirtfloor	0.097	0.120	0.81	Miss Water land	0.794	1.627	0.49
Poor quality roof	-0.002	0.109	-0.02	Miss Water house	0.707	0.828	0.85
Poor quality wall	-0.506	0.127	-3.95	Miss Assets	-3.990	2.289	-1.74
Animals	-0.177	0.114	-1.55	Constant	3.900	0.442	8.80
Land	-0.549	0.107	-5.12				
Number of Obs		2635					
LR Chi2 (54)		671.61		Pseudo R2		0.190	
Prob>Chi2		0.000		Log Likelihood		-1434.70	

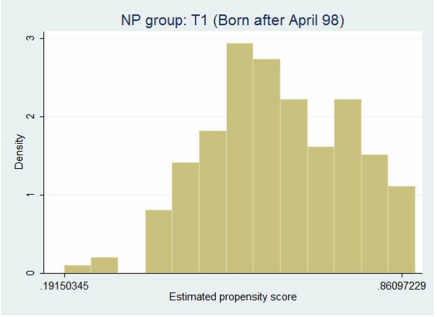
Figure C.1: Estimated propensity scores (Mother's education case).

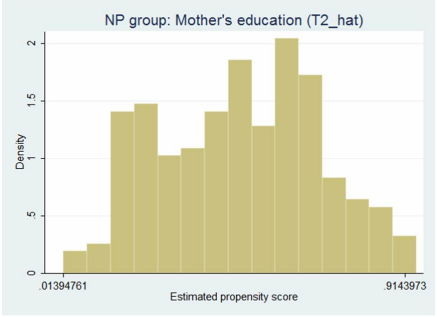


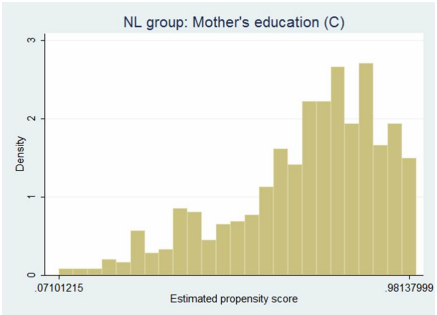












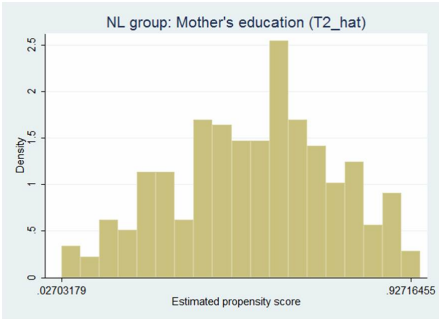


Figure C.2: Stochastic dominance results (Mother's education case).

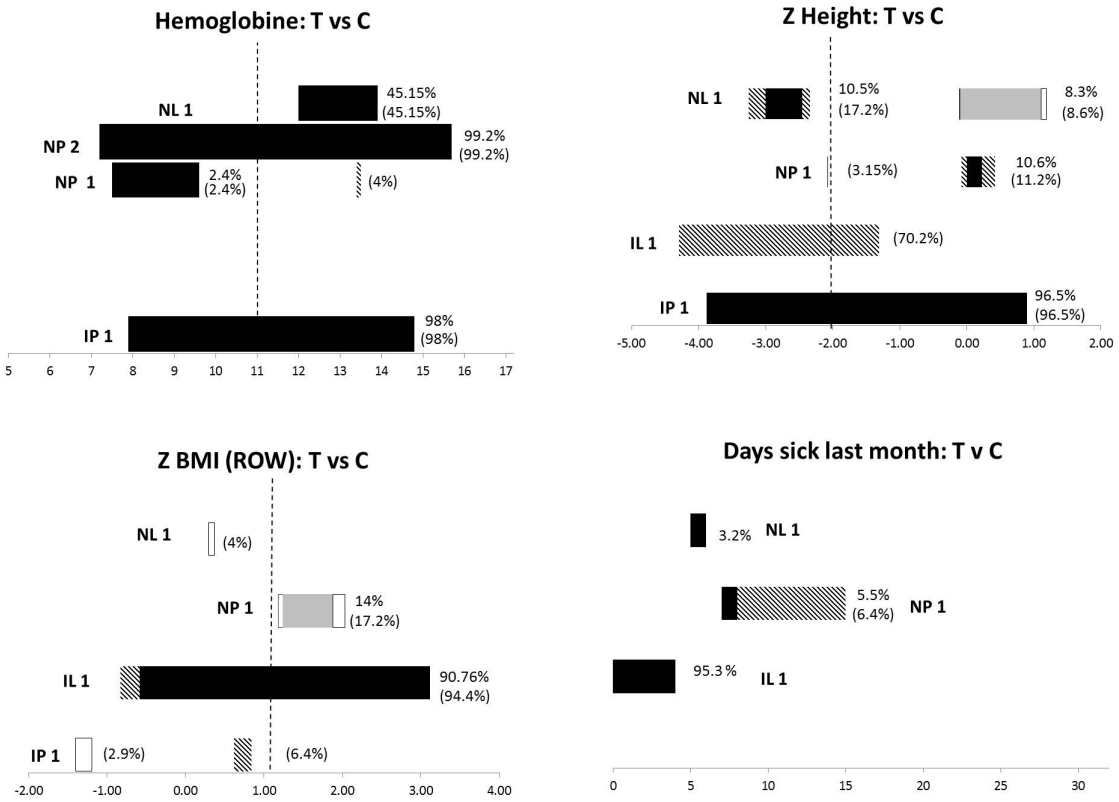


Table C.6: Propensity score matching: common support and number of observations in the common support (Mother's education case).

	Common support	Control #	Treatment #	Band-width
IP	[0.169,0.955]	104	139	0.073
IL	[0.106,0.872]	264	290	0.073
NP	[0.023,0.945]	552	250	0.070
NL	[0.071,0.951]	668	363	0.071
Total		1588	1042	

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education.

D Original versus Delay treatment

Table D.1: Composition of the samples (delay vs original treatment).

	Initial treatment		Delayed treatment	
	#	%	#	%
All	730	100	527	100
IP	110	15.1	69	13.2
IL	227	31.1	156	29.5
NP	186	25.5	156	29.5
NL	207	28.3	146	27.7

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

Table D.2: Health outcomes of 2-6 year old children in 2003 (delay vs original treatment).

(a) Initial treatment

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.17	12.4	0.35	-1.51	0.17	0.68	0.11
IP	0.27	12.0	0.41	-1.62	0.14	0.72	0.12
IL	0.19	12.4	0.50	-2.00	0.19	0.71	0.11
NP	0.14	12.5	0.24	-1.23	0.18	0.60	0.12
NL	0.13	12.5	0.26	-1.42	0.17	0.71	0.11

(b) Delayed treatment

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.18	12.3	0.33	-1.53	0.17	0.67	0.13
IP	0.31	11.9	0.34	-1.67	0.13	0.61	0.09
IL	0.25	12.0	0.41	-1.81	0.15	0.71	0.09
NP	0.10	12.5	0.23	-1.13	0.21	0.67	0.14
NL	0.14	12.7	0.35	-1.57	0.16	0.64	0.16

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

Table D.3: Health outcomes of 2-6 year old children in 2003 (delay vs original treatment): Matched samples.

(a) Initial treatment

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.17	12.4	0.35	-1.51	0.17	0.68	0.11
IP	0.27	12.0	0.40	-1.60	0.14	0.72	0.12
IL	0.19	12.4	0.50	-1.99	0.19	0.71	0.11
NP	0.13	12.5	0.24	-1.23	0.18	0.60	0.12
NL	0.12	12.6	0.26	-1.41	0.17	0.71	0.11

(a) Delayed treatment

	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.19	12.4	0.37	-1.65	0.17	0.63	0.14
IP	0.29	12.3	0.42	-1.71	0.16	0.54	0.08
IL	0.25	12.0	0.51	-2.03	0.14	0.71	0.10
NP	0.12	12.5	0.23	-1.14	0.20	0.60	0.22
NL	0.11	12.8	0.34	-1.57	0.17	0.65	0.15

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

Table D.4: Difference between initial and delayed treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs original treatment).

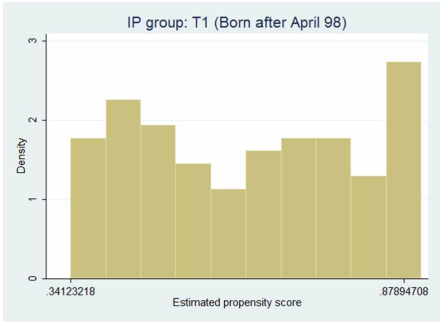
	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.02	-0.02	0.01	0.06	-0.03
IP	-0.02	-0.01	-0.02	0.18**	0.03
IL	-0.06	-0.01	0.04	0.00	0.01
NP	0.02	0.02	-0.02	0.00	-0.10
NL	0.02	-0.07	0.00	0.07	-0.05

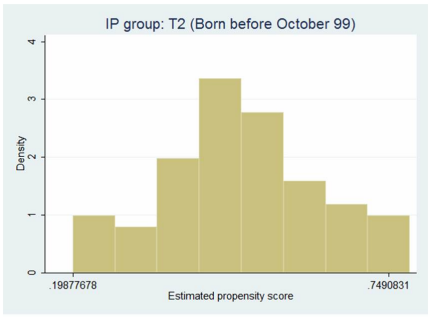
Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) "*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

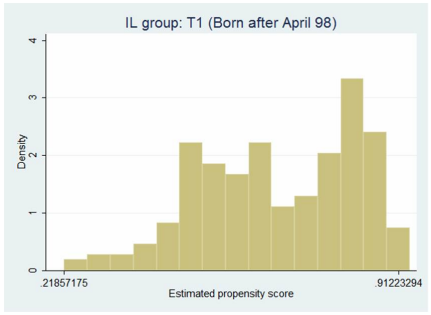
Table D.5: Logistic regression results (delay vs original treatment).

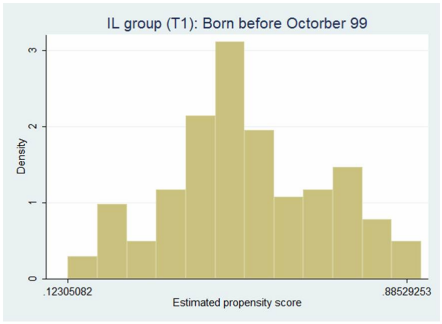
Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	0.015	0.010	1.43	Blender	-0.165	0.180	-0.92
Age spouse	-0.002	0.011	-0.18	Fridge	0.523	0.244	2.15
Sex Hh head	-0.561	0.500	-1.12	Fan	-0.400	0.280	-1.39
IndigHhHead	-0.004	0.314	-0.01	Gas stove	-0.074	0.190	-0.39
IndigSpouse	0.332	0.323	1.03	Gas heater	0.360	0.412	0.87
EducHhHead	-0.085	0.158	-0.54	Radio	-0.040	0.145	-0.28
EducSpouse	-0.136	0.165	-0.82	Hifi	0.044	0.337	0.13
Work Hh head	0.823	0.300	2.74	Tv	-0.110	0.157	-0.70
Work spouse	0.222	0.205	1.08	Video	0.401	0.395	1.02
# Children 0-5	-0.007	0.064	-0.11	Wash machine	-0.236	0.426	-0.55
# Children 6-12	-0.050	0.060	-0.86	Car	0.513	0.589	0.87
# Children 13-15	-0.106	0.119	-0.88	Truck	-0.165	0.296	-0.56
# Children 16-20	-0.025	0.094	-0.27	Guerrero	1.024	0.226	4.52
# Women 20-39	-0.074	0.151	-0.49	Hidalgo	1.596	0.245	6.50
# Women 40-59	-0.347	0.219	-1.58	Michoacan	0.385	0.267	1.44
# Women 60+	-0.208	0.240	-0.87	Puebla	0.630	0.199	3.16
# Men 20-39	-0.099	0.142	-0.70	Queretaro	-0.281	0.349	-0.81
# Men 40-59	-0.150	0.223	-0.65	San Luis	0.506	0.212	2.38
# Men 60+	-0.400	0.315	-1.27	Miss Age Sp	-0.516	1.811	-0.29
# Rooms	-0.018	0.013	-0.42	Miss Indig HH	-0.557	1.792	-0.31
Electrical light	0.066	0.165	0.40	Miss Indig Sp	0.230	2.079	0.11
Running water land	0.356	0.155	2.30	Miss Age Sp	-0.516	1.811	-0.29
Running water house	-0.654	0.292	-2.24	Miss Work Sp	-0.049	1.771	-0.03
Dirtfloor	-0.092	0.161	-0.58	Miss Water land	0.278	1.461	0.19
Poor quality roof	-0.184	0.146	-1.26	Miss Assets	0.822	0.956	0.86
Poor quality wall	-0.175	0.172	-1.02				
Animals	0.100	0.150	0.67	Constant	-0.464	0.650	-0.71
Land	0.407	0.142	0.29				
Number of Obs		1252					
LR Chi2 (56)		148.97		Pseudo R2		0.087	
Prob>Chi2		0.000		Log Likelihood		-776.97	

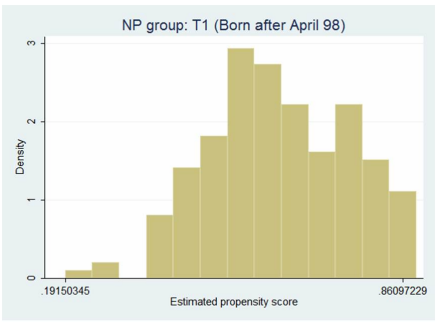
Figure D.1: Estimated propensity scores (delay vs original treatment).

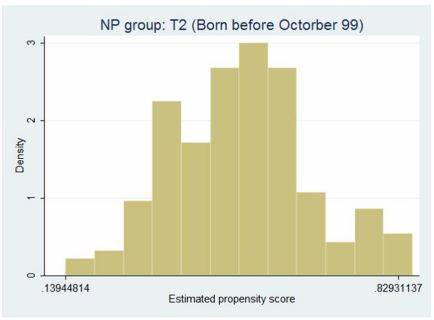


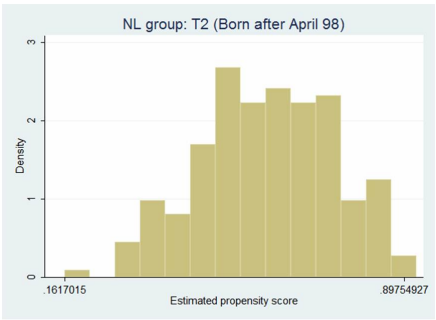












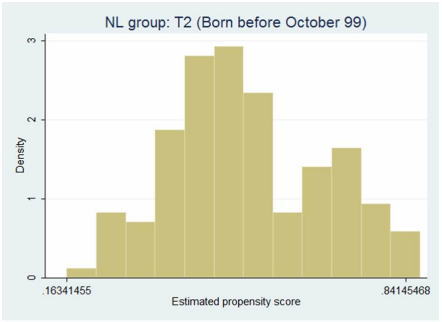


Figure D.2: Stochastic dominance results (delay vs original treatment).

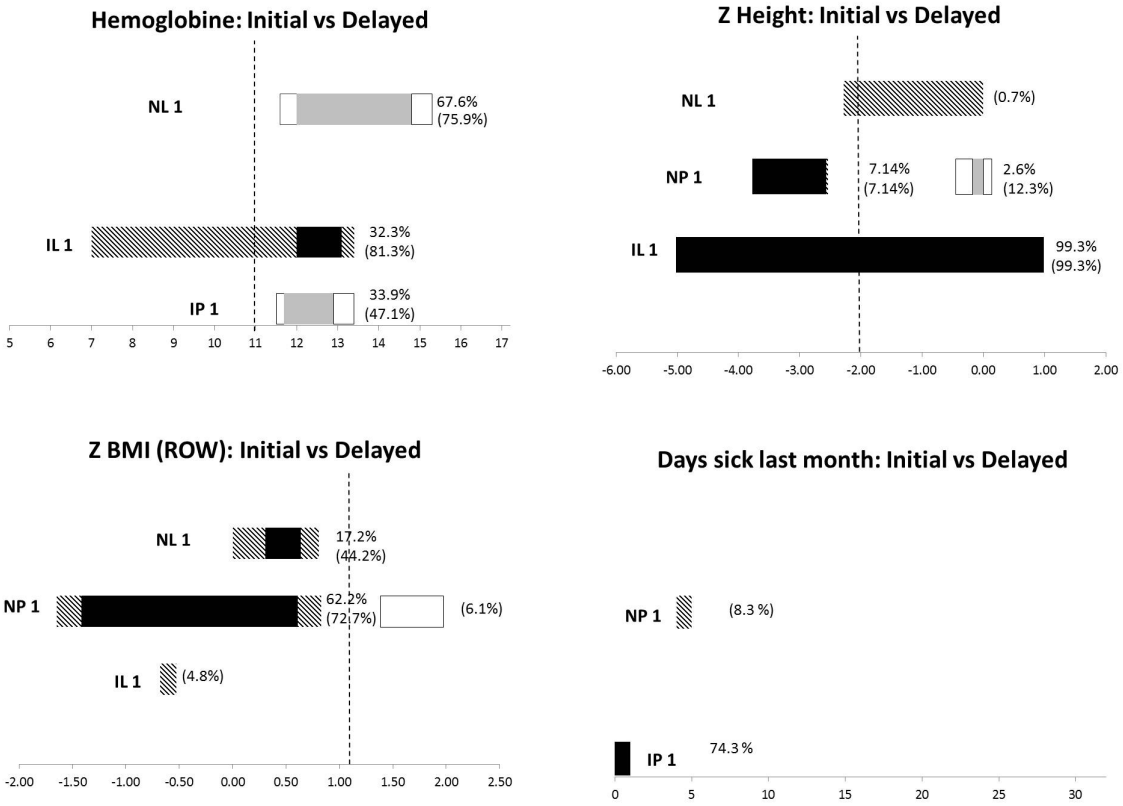


Table D.6: Propensity score matching: common support and number of observations in the common support (delay vs original treatment).

	Common support	Initial #	Delayed #	Bandwidth
IP	[0.341,0.785]	110	69	0.057
IL	[0.218,0.918]	226	152	0.066
NP	[0.191,0.859]	185	155	0.056
NL	[0.161,0.870]	206	145	0.060
Total		727	521	

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