**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

Distributional effects of *OPORTUNIDADES* on early child development

José Luis Figueroa

May 2013

2013/840

D/2013/7012/11

Distributional effects of *Oportunidades* on early child development

José Luis Figueroa[a]

Abstract:

Adequate health, nutrition, and education during childhood are essential for human development. Deficits in these realms undermine the capacity to acquire the necessary skills to perform in life. Social policies addressing the causes of disadvantages in child development take up an important place in the social agenda. The Mexican Oportunidades program is such a policy. Investments in children's health, nutrition, and education by the program are expected to facilitate children's development. Previous studies found little effect of Oportunidades on child's cognition and positive effects on noncognitve development. However, the majority of these studies take the average outcome as the relevant indicator of the effect of the program which overlooks the effect on the "non-average" child. A methodology capable of unveiling effects along the outcome's distribution is proposed here. Such methodology, originally proposed by Davidson and Duclos (2013), is based on tests of stochastic dominance and is suitable for observing effects beyond the mean. Four indicators of cognitive development and one of behavioral problems (noncognitve development) are analyzed in a sample of 2,595 children aged 2 to 6 years. The sample was collected in rural communities in Mexico in 2003 as part of the evaluation of the program. Oportunidades decreases behavioral problems experienced by children exposed to the program. The ranges where the effect is found cover a large part of the distribution of the outcomes and a large proportion of the children in the sample. In comparison to other studies, additional effects by gender and ethnicity are now found. Only one indicator of cognitive development (short-term memory) shows positive effects. Nevertheless, the results for this indicator show that children with lower values of cognitive development benefitted from the program while children with higher values did not. These heterogeneous effects highlight the importance of going beyond the average effect approach.

Davidson, R., & Duclos, J. (2013). Testing for restricted stochastic dominance. *Econometric Reviews*. 32(1): 84-125.

Keywords: Child development, *Oportunidades,* Distributional effects, Mexico.

JEL classification codes: I18, I38, J13

[a]Sherppa – Ghent University
Tweekerkenstraat 2
B-9000 Gent
Belgium
E-mail: Joseluis.figueroaoropeza@UGent.be

I. Introduction

Investments in human capital during the first years of life are essential for social and economic advancement. Adequate levels of health, nutrition, and education during childhood foster performance in adulthood, and are as such, conditions *sine qua non* for human development. In contrast, deficits in these realms during infancy and childhood predict malfunction in other social and economic domains: among others, they determine poor educational attainment, crime participation, poverty, and participation in risky behaviors (Heckman, 2007). In Mexico, the *Oportunidades* program is nowadays the most relevant public intervention aiming at raising the human capital of the most disadvantaged. The program is designed to improve day-to-day conditions of impoverished families, especially of children, and in the long-term, it aims at breaking the intergenerational cycle of poverty by making investments in health, nutrition, and education. Additionally, the set of actions carried out by the program in these domains may bring unintended benefits for the children enrolled in the program, namely higher cognitive and noncognitve development. Previous studies analyzing the effect of the program found little improvements on cognition and favorable results for noncognitve development. However, their conclusions rely on the analysis of average outcomes which ignores effects for the non-average child. The objective of this article is thus to use an alternative methodology suitable to observe effects along the distribution, i.e., effects at different percentiles and not only at the mean, and that otherwise are difficult to note if the researcher relies solely on the average outcome. The methodology, originally proposed by Davidson and Duclos (2013), relies on tests of stochastic dominance that performs comparisons between cumulative distributions of outcomes for children in and out of the program. The procedure has the main advantage of showing the outcomes' ranges where children in the program have higher or lower developmental indicators than children not enrolled.

The role of health and nutrition interventions on child development.

Abilities start to form very soon, thus inadequate conditions during the first stages of life diminish the capacity to acquire and develop skills that are in turn necessary for well-being.

Among the different domains that determine adult's capacities, health and nutrition during childhood are especially important. Health conditions such as undernourishment and stunting affect the development of the brain and the immunological system, and are associated with poor cognition and school performance, problems with conduct, and difficulties to develop social relations at school (Walker, Wachts, et al., 2007; Walker, Chang. Powell, & Grantham-McGregor, 2005; and Chang, Walker, Grantham-McGregor, & Powell, 2002). Similarly, episodes of diarrhea before age 2 years are associated with delays in school entry and inadequate cognitive performance (Lorntz, Soares, et al. 2006; and Patrick, Oria, et al., 2005), and repeated episodes of malaria have neurological and cognitive sequelae (Walker, Wachts, et al., 2007). Malnutrition and poor health conditions have also long-lasting consequences for wage earnings, education, and cognitive skills (Duc, 2011; Case &Paxon, 2010; Schick & Steckel, 2010). Therefore, interventions to prevent and remedy harmful conditions during the first years of life are crucial for fostering human development, and at the same time, are the most effective and cost-efficient manner to prevent disparities (for more on this see the revision made in Engle, Fernald, et al., 2011 and Walker, Wachs, et al., 2011).

Indeed, investments in human capital are more effective if they are carried out during the first years of life. Cunha and Heckman (2008) provide a framework explaining the process of skill formation along the life cycle. According to them, there are "sensitive" and "critical" periods in this process, the former referring to periods where investments are more effective at producing certain skills, while the latter points at those sensitive periods when skills can only be produced if investments are done during these periods. Furthermore, skills are diverse and thus affect development in different manners. In that sense, the relevance of cognitive skills for child development is as important as the role of non-cognitive ones. Psychological aspects of personality such as self-regulation, self-esteem, and motivation are as important as purely cognitive factors, and human capital interventions should take these aspects of skill formation into account (See for example the study by Schick and Steckel 2010 on the effect of height on wages). Cunha and Heckman (2008) also analyze how investments in different domains reinforce and promote development in other domains. In that sense, they define *self-productivity* as the process by which skills acquired at a particular period remained throughout the life cycle. Self-productivity also refers to the fact that investments in one dimension could foster abilities in other dimensions. Take for example emotional security in childhood which fosters curiosity and eventually promotes learning capacity. They also referred to *complementarity* of skills and define it as the power of present investments to raise the effectiveness of skills in the future. In sum, human capital interventions like *Oportunidades* are thus expected to improve children's immediate conditions in health and nutrition, but in the long-term, and attending to the framework by Cunha and Heckman (2008), they could also reflect on skills formation, although the program is not specifically designed to impact on these domains.

II. The Oportunidades program.

Oportunidades is a Conditional Cash Transfer program (CCT) and like other CCTs supplies monetary transfers to disadvantaged families on the condition that they will carry out regular investments in health, nutrition, and education. In order to receive the transfer, family members have to attend regular medical checkups. Medical attention starts during pregnancy with compulsory prenatal care and continue after birth until age 19. Before age 2, checkups are frequent (9 in total) and after that age they are provided every six months. Pregnant women should attend at least five checkups for prenatal control and nutritional supplementation. They also have to continue with medical follow-ups during the lactating period. The other adults in the family should attend checkups every year until age 49 and once after that (Fernald, Gertler, & Neufeld, 2008). Additionally, mothers have to attend regular workshops where they receive information on health and nutritional positive habits. Finally, school-aged children are entitled to receive scholarships if they attend classes at least 85% of the time and do not repeat more than twice the same level. Nowadays, the program covers more than 5.8 million poor families in Mexico and is the most important policy of this type in the country.

III. Evidence of transfer programs on skill formation

The effect of *Oportunidades* and other CCTs on skills formation is divergent. Gertler and Fernald (2004) analyze medium term impacts of *Oportunidades* in rural areas after 3-6 years of exposure. They find no improvements on cognition using three tests of cognitive development but they do find improvements on the Achenbach scale of behavioral problems for girls (9 percent average decrease). A subsequent study by Fernald, Gertler, and Neufeld (2008) investigated the role of cash on cognition. They concluded that a doubling on the cash transfer is positively associated with better motor, language, and cognitive development of children aged 2-6 years. However, Attanasio, Meghir, and Schady (2010) challenged these results arguing that the analysis suffers from reverse causality, as it is difficult to know if the effect runs from money to better skills, or if higher-skilled families attract more money coming from the program. An additional study by Fernald, Gertler, and Neufeld (2009) on the effects of the program after 10 years of implementation finds significant effects on behavioral problems, but no effects on cognition for children aged 8-10 years. They find, however, positive and independent associations between the accumulated transfers with cognition and negative ones for behavioral development. Finally, Ozner, Fernald, Manley, and Gertler (2009) concentrate on behavioral problems finding that children aged 4-6 years who have been exposed to the program between 3.5 and 5 years show a 10% reduction in a scale on aggressive/oppositional symptoms, but no decrements in anxiety/depressive ones.

Beyond the Mexican context, two studies in Ecuador and Nicaragua also look at the effect of social programs on child development. Paxon and Shady (2010) carried out and intent-to-treat analysis to analyze the effect of The *Bono de Desarrollo Humano (BDH)* in Ecuador. Their results suggest that development outcomes for poorer children are better than those of slightly richer ones. The other study is by Macours, Schady, and Vakis (2012) who report the effect of the *Atención a Crisis* in Nicaragua after 9 months of exposition using a randomization procedure to assign treatment and control groups. They find little evidence of the effect of the program on development outcomes, especially cognitive ones. However, they find evidence of socioeconomic gradients for child development.

IV. Beyond average effects.

A common trend among the studies previously described is their confidence on mean outcomes to establish the effect of the program. Although these programs are designed to reach children with similar socioeconomic characteristics, and many of them have been implemented in the rural context, differences in intensity, kind, and quality of the intervention are likely to affect each child differently. Furthermore, children enrolled in the same program may not react in the same fashion; therefore the mean effect may have underestimated (or overestimated) the impact of the program for different participants. In that sense, the mean effect describes how the average child responded to the intervention but says little about how it affected the non-average one. The larger effect among the poorest children found for example in Macour, Schady, and Vakis (2012), seems to reinforce this supposition.

Following the same reasoning, some voices have risen in favor of moving beyond the average effect approach that has become standard rule in the impact evaluation literature. An example from economics is Angus Deaton who illustrates the shortcomings of this approach arguing that the mean does not provide information on the effects at different quantiles of the distribution. According to him, relevant policy questions like the fraction of the population that benefits from the program cannot be adequately addressed by looking exclusively at the mean. He elaborates on his claim by saying that "the trial might reveal an average positive effect although nearly all of the population is hurt with a few receiving very large benefits" (Deaton, 2009; pp. 142). Moreover, there might be benefits for part of the children that the mean is not expressing, and therefore, concluding that the absence of an average effect implies that the program has no effect can be misleading.

In the same spirit, Heckman, Smith and Clemens (1997) discuss the assumptions behind the mean impact approach. For them, taking the difference in means of outcomes between participants and nonparticipants as a sufficient characteristic to describe a program's impact is only plausible under the strong assumption that the effect is homogeneous across persons. Thus, if this assumption does not hold true both desirable and undesirable distributional aspects of the program can be overlooked. Furthermore, the authors appeal to the importance of fully characterizing the distribution of impacts in order to infer the true effect of a program, and to do that, they argue on the necessity to know the joint distribution of outcomes under treatment and non-treatment conditions. However, given the difficulty of observing the joint distribution, because of the impossibility to observe individuals simultaneously in the treated and untreated states, they propose to "exhaust" all information in the marginal distributions. In an empirical example of this approach, Djebbari and Smith (2008), evaluate heterogeneous impacts of *Oportunidades* on consumption. Using the marginal distributions they construct limits of the variance of the treatment effects and find evidence of systematic variation of impacts by subgroups. Given the importance for policy analysis, they also argue in favor of a more careful analysis that takes into account heterogeneous treatment effects as well as "the number made worse off by the program and the extent of their losses".

A more recent paper by Conti and Heckman (2010) studies the role of early life endowments on adults' health and education. Family background characteristics together with cognitive and noncognitive skills developed before age 10 are found to be decisive factors of health disparities at age 30. Interestingly, they also defend the need for going beyond the mean which "can hide gains and losses for different individuals". Their results show miscellaneous effects of education on health across individuals who are similar in their observed characteristics.

The methodology proposed in this article echoes the claims about the need to go beyond the mean effect. The use of stochastic dominance criteria allow to extract information contained in the marginal distributions and is suitable for establishing gains and losses different participants to the program. To date, few studies have applied stochastic dominance methods in the context of program evaluation. Verme (2010) for example, evaluated the effect of a randomized experiment on consumption by establishing poverty dominance. Also Naschold and Barrett (2010) use stochastic dominance methods to compare treatment and control samples before and after treatment. The paper by Van de gaer, Vandenbossche, and Figueroa (2013) is perhaps the only example that looks at the effect of *Oportundiades* on children's health opportunities relying on the same test of nondominance applied in this paper. Their analysis however is different in spirit because it builds on equality of opportunity theories in which comparisons are made between individuals sharing the same background characteristics. Interestingly, their conclusions pointed to differential effects for more disadvantaged children in comparison with those in a slightly better social situation. Impacts on indigenous children are found to be larger than those found on non-indigenous children, and the effects within indigenous are larger among those with less educated parents in comparison to those with parents with at least primary education.

V. Methodology.

Imagine that θ represents a random variable measuring skill status of a child in a way that larger values of θ represent better skill levels. Skills can be of cognitive or noncognitive nature just as previously described. Consider two groups of children which before the intervention are similar in a set of observable characteristics X accounting for an array of characteristics that determine skills' status of children (for example, but not only, socioeconomic, demographic, and environmental characteristics). Define also the cumulative conditional distribution function of θ for children in the program as $F^T(\theta|X)$ and for children out of the program as $F^C(\theta|X)$. If $F^C(\theta|X) \geq F^T(\theta|X)$ for all θ and $F^C(\theta|X) > F^T(\theta|X)$ for some θ, then $F^T(\theta|X)$ first-order stochastically dominates $F^C(\theta|X)$ and a positive effect of the program can be established. In other words, first-order dominance implies that at each percentile in the distribution, children in the program achieve a higher value of skill development θ than children in control.

Traditional tests pose the null hypothesis as dominance, implying that whenever rejection occurs non-dominance can be inferred but if rejection does not occur dominance cannot be established. The analysis aims at looking at the effect of a program by comparing cumulative distributions; therefore establishing dominance and not merely the existence of non-dominance is crucial. The test proposed in Davidson and Duclos (2013) is designed to deal with this inconvenience because their procedure places non-dominance as the null hypothesis. In that way, rejection of the null necessarily implies that dominance exists. Such test offers a more logical manner to look for dominance criteria, and at the same time, results in a suitable methodology in the context of program evaluation.

However, in the presence of continuous variables, dominance never occurs for the entire range of the distribution. Take for example the lower bound of the common support where $F^C(\theta|X) = F^T(\theta|X) = 0$, such that at this point the cumulative distribution function for both groups is the same. In addition, the amount of information that is often available in the tails is very scarce which makes inference difficult in this part of the distribution. The nature of the tests implies that rejection of the null, i.e. dominance of treatment over control, is only possible over restricted ranges along the distribution of θ.

Hence, the test proposed by Davidson and Duclos (2013) is formally defined as follows:

Let $U \subseteq \Theta$ be the union of the supports of the cumulative distributions of children not enrolled in the program (referred from now on as control) and children in the program (from now on treatment), respectively $F^C(\theta|X)$ and $F^T(\theta|X)$. We test the null hypothesis of nondominance of $F^C(\theta|X)$ by $F^T(\theta|X)$,

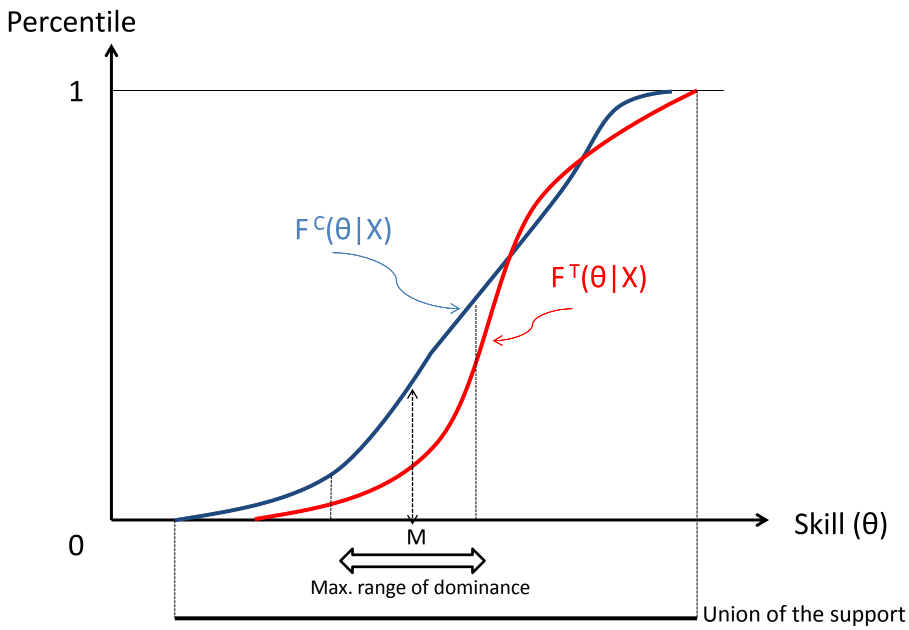
$$\max_{\theta \in U}((F^T(\theta|X) - F^C(\theta|X)) \geq 0,$$

against the alternative hypothesis that $F^T(\theta|X)$ first-order stochastically dominates $F^C(\theta|X)$,

$$\max_{\theta \in U}(\left(F^T(\theta|X) - F^C(\theta|X)\right) < 0.$$

The test identifies the maximal range U over the distribution where dominance arises by first detecting the point where the difference between both curves is the most significant (Point "M" in Figure 1). Around that point the test extends an interval U until the point where further extending the interval no longer allows rejection of the null. This is an alternative procedure to evaluate social programs because it identifies where in the distribution the effect occurs.

Figure 1. Test of first-order stochastic dominance



<u>Comparability between groups</u>

As pointed out by Van de gaer, Vandenbossche, and Figueroa (2013), the composition of the treatment and control groups in terms of the set of socioeconomic, demographic, and environmental characteristics (X) before the intervention should be similar for both groups. Given that skills are partly determined by these characteristics and the assumed effect of the program on skills' development, there is a risk of "confounding" the true effect of the program. For instance, one could erroneously infer a positive effect if treated children had initially higher levels of skills. To clarify this point, imagine for a moment that children in treatment come from wealthier families or have parents with higher levels of education in comparison with children in control. In this case, treated children would have characteristics that will make them to be more inclined to develop skills in comparison with untreated children, and thus, the impact of the program on treated children would be higher than the impact the program would have had on untreated children. This phenomenon is known in the impact evaluation literature as "selectivity bias" and results from the lack of independence between participation in the program and the set of demographic, socioeconomic and environmental characteristics that determine skill formation.

Controlling for such differences between groups is thus indispensible to correctly infer the effect of the intervention.

One way to deal with this problem is to deliberately assign the treatment in a random fashion. If groups are sufficiently large, the randomization assignment guarantees that the treatment and control groups are comparable in terms of pre-intervention characteristics. *Oportunidades'* original design followed such a randomization procedure at locality level but not at household or individual level, thus there might be important differences remaining between groups (See Section VI). To overcome this problem a Propensity Score Matching procedure (PSM) was carried out. The procedure consists in weighting observations according to the probability that a child in the sample belongs to the treatment group. The weighted observations in treatment and the corresponding non-weighted observations in control are then used to construct the distribution functions that are needed for the test of stochastic dominance explained above (for a detailed explanation of the weighting procedure see Appendix 3 in Van de gaer, Vandenbossche, & Figueroa, 2013; for an exposition on the Propensity Score Matching method see Blundell and Costa Dias, 2009).

VI. Data sample and outcomes

Data Sample

The data sample was collected in rural communities in Mexico as part of the external evaluation of the program after 6 years of implementation. A subsample of children aged two to six years was collected with the purpose to assess their cognitive and non-cognitive development in 2003. The sample contains information for three groups: children whose families were incorporated since the onset of the program (treatment-1998), children from families incorporated two years later (treatment-2000), and children from households that in 2003 had not been incorporated (control). The selection of program participants and controls proceeded in several steps. In a first stage in 1997, highly deprived communities were randomly assigned into two groups: treatment-1998 (group that was planned to be enrolled in 1998) and the treatment-2000 group that originally served as control (to be enrolled two years later). 560 localities with at least 500 and at most 2500 inhabitants were selected during this stage (320 for treatment-1998 and 186 for treatment-2000). During the second stage, socioeconomic and demographic conditions in the selected communities were assessed in order to identify which households where eligible for the program within each community. A marginality index based on income, demographic composition, and dwelling conditions of the household indicated whether the family was eligible for receiving the program or not.

By 2003, when the information was collected, all households that were initially assigned to the control group (treatment-2000) were already incorporated and thus it was necessary to construct a new control group. In order to construct the new control group, *Oportunidades*'s authorities implemented a PSM procedure according to characteristics at locality level. In that manner, deprived localities where the program did not operate at that time were matched to treatment localities with similar characteristics (Todd, 2004). However, as explained by Van de gaer, Vandenbossche, and Figueroa (2013) the procedure has two problems: firstly, information on the set of characteristics used to categorized deprived localities in 2003 was obtained from the National Census in 2000 when both treatment groups were already receiving the program, therefore it is possible that the program had already changed the composition of the groups with the consequent problems for the estimation of the true effect. Secondly, the matching procedure was made at locality instead of household level, thus differences between treatment and control households might still exist. Table A1 in the Appendix indicates that this is indeed a problem for the sample analyzed in this article. As can be observed in the table, there was a higher probability that the head of the household was older and was a male in the treatment group. Similarly, the head of the household was more likely to be indigenous in the treatment group but less educated and less likely to have a job. There are also differences in terms of the demographic composition, dwelling conditions, and the kind and quality of assets available in the household.

Additionally, there is a disadvantage about using the group incorporated in 1998 due to the lack of information on whether the households in this group effectively received the transfers from the program. Therefore, there is no certainty that households from the original treatment (treatment-1998) in the sample were effectively incorporated or just eligible for participation.

Given the role that cash seems to have on children's development (see the discussion presented in Section I) only those households for which this information was available were used for the analysis. The same criterion was applied to the original control (treatment-2000). Hence, the final sample was restricted to households incorporated in 2000 and for which information on transfers was available.

Having in mind the unbalanced composition in terms of pre-program characteristics of treatment-2000 and control, the PSM procedure was carried out to match individual children, as described at the end of the previous section. The new PSM procedure was performed on the basis of children's characteristics in 1997 when the program was not yet in place. Pre-program characteristics for treatment 1998 and 2000 were obtained from baseline data collected in 1997 by *Oportunidades'* authorities, while information for the new control came from retrospective questions about 1997 households' characteristics collected in 2003 (INSP, 2005). Additionally, the matching was carried out between treatment and control groups and among subgroups defined according to gender and ethnicity. In that way, 4 treated groups divided by race (indigenous or not) and gender (boys and girls) and the complete sample, were matched to the corresponding groups in control. That is, boys in control with boys in treatment, indigenous children in control with indigenous in treatment, and so forth. The estimated propensity scores for treatment and controls and for the 4 groups in the sample can be found in Figure A2 in the Appendix. The matching procedure was effective in balancing the sample in terms of pre-intervention characteristics for all groups except for children of indigenous origin. The results for this group should be taken with caution due to possible bias in the results.

Finally, the division of the sample into the four groups described above and the subsequent matching among them aims at analyzing heterogeneous effects across these groups. In Mexico, indigenous people remain socially and economically relegated, and there are studies documenting differential effects of *Oportunidades* between indigenous and non-indigenous people (See Behrman, Fernald, Gertler, Neufeld, & Parker, 2008 for an analysis of *Oportunidades* on health, nutrition and cognition of children after 10 years of exposition, and Gonzalez de la Rocha, 2008 for an analysis on labor and demographic patterns after 10 years of exposition to the program). In the same fashion, and as documented by Gertler and Fernald

(2004), the program affects boys and girls differently and so, heterogeneous effects are likely to be observed for the children in the sample used here. The composition of the final sample is shown in Table 1.

Table 1. Composition of the samples

	All		Boys		Girls		Indigenous		Non-indigenous	
Complete sample										
	N	%	N	%	N	%	N	%	N	%
Treatment	1,086	41.8	541	41.8	539	41.6	405	52.9	634	35.7
Control	1,509	58.2	752	58.2	756	58.4	361	47.1	1,144	64.3
Total	2,595	100.0	1,293	100.0	1,295	100.0	766	100.0	1,778	100.0
Achen Index (Behavioral problems)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	1,086	41.8	541	41.8	539	41.6	405	52.9	634	35.7
Control	1,509	58.2	752	58.2	756	58.4	361	47.1	1,144	64.3
Total	2,595	100.0	1,293	100.0	1,295	100.0	766	100.0	1,778	100.0
Woodcock-Johnson (Short Term Memory)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	1,047	41.4	522	41.3	519	41.8	382	52.3	619	35.3
Control	1,485	58.6	742	58.7	724	58.2	348	47.7	1,133	64.7
Total	2,532	100.0	1,264	100.0	1,243	100.0	730	100.0	1,752	100.0
Woodcock-Johnson (Visual integration)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	1,013	41.0	511	41.3	496	40.5	370	52.3	597	34.8
Control	1,458	59.0	727	58.7	730	59.5	337	47.7	1,117	65.2
Total	2,471	100.0	1,238	100.0	1,226	100.0	707	100.0	1,714	100.0
Communicative Development Inventories (CDI)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	186	37.2	102	39.1	84	35.1	61	42.4	117	33.6
Control	314	62.8	159	60.9	155	64.9	83	57.6	231	66.4
Total	500	100.0	261	100.0	239	100.0	144	100.0	348	100.0
Peabody Picture Vocabulary Test (PPVT)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	853	42.4	418	42.1	429	42.3	322	54.9	494	35.6
Control	1,160	57.6	575	57.9	584	57.7	264	45.1	892	64.4
Total	2,013	100.0	993	100.0	1,013	100.0	586	100.0	1,386	100.0

Outcomes

Four indicators of cognitive development are used in the analysis. The first one is the Spanish version of the Peabody Picture Vocabulary Test (PPVT) which is a measure of vocabulary recognition. The test has been extensively used as proxy of language development among pre-schooled children aged 3-6 years (See the study of Paxon & Schady, 2007 in Ecuadorian children, and the study of Gertler & Fernald, 2004 for children in *Oportunidades*). Children in the study were exposed to cards containing 4 images each. Next, the interviewer mentioned at loud a word that the child had to recognize among the set of images in front of her. If the answer was correct, another set of images corresponding to a more difficult word was presented. This procedure ran until the child made six mistakes in 8 consecutive questions or until de set of images was completed. To my knowledge, there are no norms for children with similar characteristics and age ranges to the ones in the sample used here. Thus, raw scores to construct the conditional distribution functions were used. In addition, a variable indicating age of the child in months was included in the PSM in order to avoid bias in the estimation.

The second measure of cognition development is the Spanish version of the MacArthur Communicative Development Inventories (CDI). The test evaluates language and communication skills of children aged 24-35 months and measures early language milestones in young Spanish-speaking children. The test consists in asking parents to identify words that their children know (understand and pronounce from a list). The number of words selected by the parents is then summed and used as an index of language ability. In comparison with other laboratory observation methods, this test has been proven to be more effective in assessing early language development (Jackson-Maldonado & Bárcenas Acosta, 2006). Also, children receiving nutritional supplementation have been found to score higher in the CDI test (Gertler & Fernald, 2004; O'Connor & Adamkis, 2001).

Two additional indicators contained in the Spanish language version of the Woodcock-Johnson-Muñoz test were also analyzed. The first indicator is designed to measure short-term memory by testing whether children are able to remember words that an interviewer mentions at loud. The second evaluates visual integration skills, i.e. the ability of the child to recognize figures shown in a piece of paper. Both tests are designed for children older than 3 years and are suitable for assessing interventions similar to *Oportunidades* given their sensitivity to changes in nutritional and health-related patterns, as well as to income interventions (Fernald, Gertler, & Neufeld, 2008). Similar to the case of the indicators mentioned before, raw scores to construct the conditional distribution functions were used.

Finally, our measure of non-cognitive development is the Achenbach Child Behavior Checklist (Achen Index). The test is suitable for evaluation of early behavioral and socio-emotional development among children aged 24-72 months. Based on parental answers, child's problems like hyperactivity, bullying, bad conduct, violence at home, and responsiveness are directly rated by parents from a list. As well as in the CDI case, the index is constructed adding the number of positive answers by the parents, but contrary to the tests of cognitive development, larger values indicate poorer development. The test has been previously used to assess the effects of parental background and environmental conditions of children on behavioral development (Kahn, Brandt, & Whitaker, 2004; and Pachter, Auinger, Palmer, & Weitzman, 2006).

The analysis of our five outcomes provides a picture of early cognitive and non-cognitive development in children. At the same time, the continuous nature of the variables allows us to construct distributions that provide a better insight in the effects of the program for different groups.

VII. Results

Table 2 shows the results of the tests of non-dominance for the entire sample and each of the groups. The table is divided into 5 panels (A to E), each showing the results for every outcome of skill development. The ranges over which each outcome is defined and the ranges for which an effect is found, i.e. where dominance is established, are presented in columns III and IV respectivelty. Column I indicates the direction of test: a favorable effect of the program is established if the CDF of treatment dominates the corresponding CDF of control (except in the case of the Achen Index, as a higher value of the index indicates more behavioral problems and a worse outcome). Column II lists the significance of the test which is either 5 or 10%. Since the test performed at 10% is quicker to reject the null hypothesis the ranges are wider than those at a 5% level. Finally, the corresponding percentiles and the proportion of the population where significant effects were found are shown in Columns V to VIII for both the treatment and the control groups.

Panel A shows the results for the outcome of non-cognitive development. The Achen Index is constructed in such a way that children who manifest more signs of psychological stress report larger values of the index. Therefore, a positive effect is established whenever the CDF for control dominates the one for treatment. The ranges where the null can be rejected cover almost the entire distribution of the variable. The effect is favorable in all cases for more than 90% of the population in each of the groups, and in the case of girls we found that practically all of them benefited. As mentioned before, the results for the group of indigenous children, although favorable, should be taken with caution since the matching procedure was not effective for balancing the composition between treatment and control children. Nonetheless, the results presented here are in line with previous analysis with children in *Oportunidades*: the program decreases behavioral problems in children. Furthermore, the analysis of distributions permits to observe effects that are overlooked when relying exclusively on average effects. In that sense, the analysis by Ozner, Fernald, Manley, and Gertler (2009) find no differences between gender or ethnicity, and Gertler and Fernald (2004) find only effects for girls and not for boys.

Moving down to Panel B and C, we find the results for the cognitive outcomes. In comparison with the evidence on behavior, the results for short term memory ability in Panel B are mixed. On the one hand, positive effects are found for the complete sample and for all groups for lower values of cognitive ability. On the other hand, the effect is negative for 17.5% of boys and 22.5% of non-indigenous children with high values of cognitve ability. Once again, the analysis of distributions unveils effects that are not reported in previous studies. In Gertler and Fernald (2004) for example, no effects on cognition for girls and only weak effects for boys are found while effects by ethnic group are not analyzed. Moreover, the methodology used here proves to be useful to know where in the distribution the effect exists as in this case where positive effects are found at the bottom-line of the distribution. Results for visual integration are presented in Panel C. *Oportunidades* has no positive effects for the complete sample nor for any of the groups. On the contrary, the results shows that the control group outperformed the treatment group for a small fraction of the girls (8%) in the range [18,19] at 10% level of significance. Similarly, non-indigenous children in control score better on visual integration than those in treatment: a fifth of the control group dominates treatment at 5% and close to half of the population at 10% level of significance.

Finally, the results for the two measures of language development are shown in Panel D and E. Positive effects of the program for the CDI index are only observed for 3.2% of girls in the range [72,73]. Negative effects are only found for boys in 17% of the population in the range [88,99]. In general, the CDI index shows no positive effects for the complete sample, while negative effects are observed for 17% of the population at 5% level of significance. Similarly, no positive effects are observed when looking at the Peabody Picture Vocabulary Test, and effects are negative for the vast majority of the population (98.9%) but only at 10%. The observation of such effects across groups shows that detrimental effects appear only for girls at 10% and that these effects are lessened because the ranges are smaller in comparison with the results for the complete sample.

Sensitivity Analysis

The superiority of the analysis of distributions over the mean effect approach remains as long as the composition of the sample used here and that of the samples used for the studies mentioned above is comparable. If as explained before, children from a particular sample possess characteristics that make them to response to the program differently from the children in other samples, then the effect of the program could be given by such differences and not by the robustness of the methodology proposed. The samples used in Gertler and Fernald (2004) and Ozner, Fernald, Manley, and Gertler (2009) differ from the one used in this analysis in terms of age of the children selected. There are also differences in terms of the time children (and their families) have been exposed to the program. For that reason, two additional analyses were carried out using samples that were constructed to be as comparable as possible with the samples in these studies.

The first analysis compares the combined treatment groups (1998 and 2000) with the control constructed in 2003, which are the groups used by Gertler and Fernald (2004). The composition of the sample and the results of the tests for all groups can be found in Appendix 3. As can be observed the results for the index on Behavioral problems are very similar to the results of our main analysis. Positive results for the all groups and similar percentages where dominance occurs are observed. Also, the results for short-term memory ability follow a similar pattern: effects are positive for children with low cognition and negative for children with higher values. On the other hand, the second analysis focuses on the same sample of children used by Ozner, Fernald, Manley, and Gertler (2009) in which the effect of *Oportunidades* on children's behavior is analyzed, and as well as in their analysis, the results are in favor of the program (Appendix 4). In contrast with the analysis by Ozner, Fernald, Manley, and Gertler (2009), a PSM procedure was carried out in the subgroups as explained above, however, the procedure was not successful to balance the characteristics between treatments and controls in this case, thus the results for the subgroups are likely to be biased.

Table 2. Stochastic dominance results by gender and race.

Panel A. Achen Index (Behavioral problems)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,50]	[1,37]	0.014	0.948	0.014	0.948	0.93	0.93
	Positive	10%	[0,50]	[1,37]	0.014	0.948	0.014	0.948	0.93	0.93
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
BOYS	Positive	5%	[0,50]	[1,37]	0.015	0.939	0.015	0.939	0.92	0.92
	Positive	10%	[0,50]	[1,37]	0.015	0.939	0.015	0.939	0.92	0.92
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
GIRLS	Positive	5%	[0,50]	[0,44]	0.000	0.998	0.000	0.999	0.999	0.999
	Positive	10%	[0,50]	[0,44]	0.000	0.998	0.000	0.999	0.999	0.999
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
INDIGENOUS	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	[2,30]	0.022	0.960	0.028	0.961	0.94	0.93
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
NON-INDIGENOUS	Positive	5%	[0,50]	[1,37]	0.013	0.962	0.010	0.949	0.95	0.94
	Positive	10%	[0,50]	[1,41]	0.013	0.983	0.010	0.983	0.97	0.97
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						

Panel B. Woodcock-Johnson (Short Term Memory)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,55]	[5,17]	0.076	0.342	0.083	0.358	0.266	0.275
	Positive	10%	[0,55]	[5,19]	0.076	0.393	0.083	0.392	0.316	0.309
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	[28,36]	0.647	0.962	0.614	0.951	0.315	0.337
BOYS	Positive	5%	[0,55]	[7,14]	0.132	0.280	0.144	0.294	0.148	0.150
	Positive	10%	[0,55]	[7,15]	0.132	0.305	0.144	0.321	0.172	0.177
	Negative	5%	[0,55]	[30,34]	0.753	0.929	0.704	0.879	0.176	0.175
	Negative	10%	[0,55]	[28,36]	0.665	0.966	0.628	0.946	0.301	0.318
GIRLS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	[1,21]	0.048	0.434	0.061	0.430	0.385	0.369
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	none						
INDIGENOUS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	[9,17]	0.202	0.442	0.276	0.546	0.241	0.270
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	none						
NON-INDIGENOUS	Positive	5%	[0,55]	[5,13]	0.058	0.191	0.069	0.222	0.132	0.154
	Positive	10%	[0,55]	[5,19]	0.069	0.328	0.058	0.313	0.259	0.255
	Negative	5%	[0,55]	[30,34]	0.661	0.892	0.643	0.868	0.231	0.225
	Negative	10%	[0,55]	[30,34]	0.661	0.892	0.643	0.868	0.231	0.225

Panel C. Woodcock-Johnson (Visual integration)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	none						
	Negative	10%	[0,42]	none						
BOYS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	none						
	Negative	10%	[0,42]	none						
GIRLS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	none						
	Negative	10%	[0,42]	[18,19]	0.857	0.927	0.821	0.900	0.071	0.079
INDIGENOUS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	none						
	Negative	10%	[0,42]	none						
NON-INDIGENOUS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	[15,19]	0.700	0.905	0.669	0.880	0.204	0.211
	Negative	10%	[0,42]	[11,20]	0.484	0.933	0.484	0.933	0.449	0.449

Panel D. Communicative Development Inventories (CDI)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	[89,98]	0.742	0.930	0.748	0.920	0.188	0.172
	Negative	10%	[0,100]	[88,99]	0.715	0.946	0.736	0.943	0.231	0.207
BOYS	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	[89,99]	0.725	0.951	0.780	0.950	0.225	0.170
	Negative	10%	[0,100]	[88,99]	0.696	0.951	0.774	0.950	0.255	0.176
GIRLS	Positive	5%	[0,100]	[72,73]	0.429	0.476	0.516	0.548	0.048	0.032
	Positive	10%	[0,100]	[71,75]	0.417	0.500	0.503	0.568	0.083	0.065
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	none						
INDIGENOUS	Positive	5%	[0,100]	[49,82]	0.262	0.672	0.277	0.735	0.410	0.458
	Positive	10%	[0,100]	[47,84]	0.262	0.705	0.265	0.747	0.443	0.482
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	none						
NON-INDIGENOUS	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	none						

Panel E. Peabody Picture Vocabulary Test (PPVT)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	[1,63]	0.009	0.999	0.002	0.997	0.989	0.995
BOYS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	none						
GIRLS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	[2,57]	0.026	0.995	0.014	0.993	0.970	0.979
INDIGENOUS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	[2,19]	0.034	0.835	0.023	0.848	0.801	0.826
NON-INDIGENOUS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	none						

VIII. Discussion

The present analysis stressed the need to go beyond the traditional mean approach that is typical in the impact evaluation literature. The methodology (Davidson & Duclos 2013) reveals important features of the distribution and thus offers a more detailed perspective of the effects of an intervention. I illustrate this by analyzing *Oportunidades*, one of the largest anti-poverty programs in the world, on four outcomes of cognitive and one of non-cognitive development for children aged two to six years.

Oportunidades provides monetary incentives as well as direct assistance on health and nutrition, placing special emphasis on children. By doing so, the program is expected to enhance human capital and thus eventually to break the intergenerational cycle of poverty. Although the program does not explicitly addresses skills development, the package of services and the information provided by the program (e.g. educational workshops on health and nutrition for mothers) may improve children's readiness to acquire these skills. Previous studies that document improvements on children's health and nutrition conditions thanks to *Oportunidades* suggest that this could be the case. To our knowledge, only three papers have studied the impact of *Oportunidades* on skill development for children (see Section III); the results in these studies exhibit scant effects on cognitive and some positive results on non-cognitive development. The scope of these studies is limited though given their reliance on mean effects. By contrast, the results in this analysis show that effects exist in some ranges while in others not.

The most striking result is the presence of positive effects in all groups on children's behavioral problems, which is our measure of non-cognitive development. Furthermore, effects are present in ranges that cover a substantial part of the distribution and for the majority of children in our sample. Contrary to the analysis by Gertler and Fernald (2004), who find large mean effects for girls and no effects for boys, our methodology proves to be powerful enough to unveil effects for all groups. Similarly, Ozner, Fernald, Manley and Gertler (2009) find favorable effects of the program on children's behavioral problems but effects do not differ by gender and race as in our analysis. The fact that children incorporated to the program experience fewer symptoms of behavioral problems is especially relevant given the importance that noncognitive development factors have for children's future well-being. According to Heckman and Kautz (2012), personality traits (non-cognitive abilities) are highly predictive of success in life, and thus, interventions aiming at improving such abilities occupy a central role in the design of public policies to foster human development. They also argue in favor of interventions to foster "soft skills" because their effects remain throughout life while effects on cognition eventually disappear without further investments in other domains.

Among the four outcomes of cognitive skill, short-term memory is the only one for which effects are in favor of the program. The results indicate that positive effects appear in lower parts of the distribution while the evidence against the program appears in ranges higher in the distribution for boys and non-indigenous. To the extent that the program should aim at the bottom end of the distribution of skills, these results could be interpreted as favorable for the program. Contrary to language development (Peabody and CDI tests) where parental stimulation plays a significant role, the effect on short-term memory depends more on factors associated with brain development. That is, the program does not provide specific information to teach parents on how to stimulate their children, but it does improve health and nutritional development, therefore, the positive results on short-term memory and absence of results on language development should not come as a surprise.

IX. References

Attanasio, O., Meghir, C., & Schady, N. (2010). Mexico's conditional cash transfer programme. *Lancet*. 375: 980.

Behrman, J. R., Fernald, L.C.H., Gertler, P.J., Neufeld, L.M., & Parker, S. (2008). *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007).* Instituto Nacional de Salud Pública.

Blundell, R., & Costa Dias, M.(2009). Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources.* 44(3): 565-640.

Case, A., & Paxson, C. (2008). Causes and consequences of early-life health. *Demography.* 47(Supp): S65.

Chang ,S.M. Walker, S.P., Grantham-McGregor, S.M., & Powell, C.A. (2002). Early childhood stunting and later behaviour and school achievement. *Journal of Child Psychology and Psyhiatry*.43:775-783.

Conti, G., & Heckman, J. (2010). Understanding the early origins of the education-health gradient: a framework that can also be applied to analyze gene-environment interactions. *Perspectives on Psychological Science*. 5:585-605.

Cunha, F., & Heckman, J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*. 43(4): 738-782.

Davidson, R., & Duclos, J. (2013). Testing for restricted stochastic dominance. *Econometric Reviews*. 32(1): 84-125.

Deaton, A. (2009). *Instruments of development: randomization in the tropics, and the search for the elusive keys to economic development*. Proceedings of the British Academy. 162: 123-160. Oxford University Press. Oxford.

Djebbari, H., & Smith, J. (2008). Heterogeneous impacts in PROGRESA. *Journal of Econometrics*. 145: 64-80.

Duc, L. T. (2011). Height and cognitive achievement of Vietnamese children. *World Development*. 39(12): 2211-2220.

Engle, P.L., Fernald, L.C.H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., et al. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *Lancet*. 378:1339-1353.

Fernald, L.C.H., Gertler, P.J., & Neufeld, L.M. (2008). Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. *Lancet.*371: 828-837.

Fernald, L.C.H., Gertler, P.J., & Neufeld, L.M. (2009). 10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behavior: a longitudinal follow-up study. *Lancet.*374: 1197-2005.

Gertler, P., & Fernald, L. C. (2004). *The medium term impact of Oportunidades on child development in rural areas*. National Institute of Public Health; Cuernavaca, Mexico.

González de la Rocha, M. (2008). *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007).* Instituto Nacional de Salud Pública.

Heckman, J. & Kautz, T. (2012). Hard Evidence on soft skills. *Labour Economics*, 19(4): 451-464.

Heckman, J. (2007). The economics, technology, and neuroscience of human capability formation. *National Academy of Sciences.* 104(33):13250-13255.

Heckman, J., Smith, J., & Clemens, N. (1997). Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts. *Review of Economic Studies*. 64(4): 487-535.

INSP (2005). *General Rural Methodology Note*. Instituto Nacional de Salud Pública. Cuernavaca, Mexico. INSP2005.

Jackson-Maldonado, D. & Bárcenas Acosta, R.P. (2006). Assesing early communicative abilities in spanish-speaking children from low-SES families. *Zero to Three*. 27: 22-27.

Kahn, R.S., Brandt, D., & Whitaker, R.C. (2004). Combined effect of mothers' and fathers' mental health symptoms on children's behavioral and emotional well-being. *Archives of Pediatrics and Adolescent Medicine.* 158: 721-729.

Lorntz, B., Soares, A.M., Moore, S.R., Pinkerton, R., Gansneder, B., Bovbjerg, B.E., et al. (2006). Early childhood diarrhea predicts impaired school performance. *Pediatric Infectious Disease Journal*. 25: 513-520.

Macours, K., Schady, N., & Vakis, R. (2012). Cash transfers, behavioral changes, and cognitive development in early childhood: Evidence from a randomized experiment. *American Economic Journal: Applied Economics*. 4(2): 247-273.

Naschold, F. & Barrett, C.B.(2010). *A Stochastic Dominance Approach to Program Evaluation with an Application to Child Nutritional Status in Kenya*. Working Paper.

O'Connor, D., Hall, R., & Adamkis, D. (2001). Growth and development in preterm infants fed long-chain polyunsaturated fatty acids: a prospective, randomized controlled trial. *Pediatrics*. 108(2): 359-371.

Ozner, E. J., Fernald, L.C.H., Manley, J.G., & Gertler, P.J. (2009). Effects of a conditional cash transfer program on children's behavior problems. *Pediatrics*. 123(4): e630-e637.

Pachter, L.M., Auinger, P., Palmer, R., & Weitzman, M. (2006). Do Parenting and the Home Environment, Maternal Depression, Neighborhood, and Chronic Poverty Affect Child Behavioral Problems Differently in Different Racial-Ethnic Groups? *Pediatrics*. 117(4):1329-1338.

Patrick, P.D., Oria, R.B., Madhavan, V., Pinkerton, R., Lorntz, B., Lima A. A., et al. (2005). Limitations in verbal fluency following heavy burdens of early childhood diarrhea in Brazilian shantytown children. *Child Neuropsycology*. 11:233-244.

Paxon, C., & Schandy, N. (2010). Does money matter? The effects of cash transfers on child development in rural Ecuador. *Economic Development and Cultural Change*. 59(1):187-229.

Paxson, C., & Schady, N. (2007). Cognitive development among young children in Ecuador: the roles of wealth, health, and parenting. *Journal of Human Resources*. 42(1): 49-84.

Schick, A., & Steckel, R. H. (2010). *Height as a proxy for cognitive and non-cognitive ability*. NBER Working Paper 16570.

Todd, P. E. (2004). *Design of the Evaluation and Method used to Select Comparison Group Localities for the Six Year Follow-Up Evaluation of Oportunidades in Rural Areas*. Technical report, International Food Policy Research Institute.

Van de gaer, D., Vandenbossche, J., & Figueroa, J.L. (2013). Children's health opportunities and program evaluation: Mexico's Oportunidades program. *World Bank Economic Review*. Forthcoming, dio: 10.1093/wber/lhs032.

Verme, P.(2010). Stochastic Dominance, Poverty and the Treatment Effect Curve. *Economics Bulletin* 30 (1): 365–373.

Walker, S.P., Wachs, T.D., Grantham-McGregor, S.M., Black, M. M., Nelson, C.A., Huffman, S.L., et al. (2011). Inequality in early childhood: risk and protective factors for early child development. *Lancet*. 378: 1325-1338.

Walker, S. P., Wachs, T.D., Gardner, J.M., Lozoff, B., Wasserman, G.A., Pollitt, E., et al. (2007). Child development: risk factors for adverse outcomes in developing countries. *Lancet*, 369: 145-167.

Walker, S.P., Chang, S.M., Powell, C.A., & Grantham-McGregor, S.M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study. *Lancet*, 366:1804-1807.

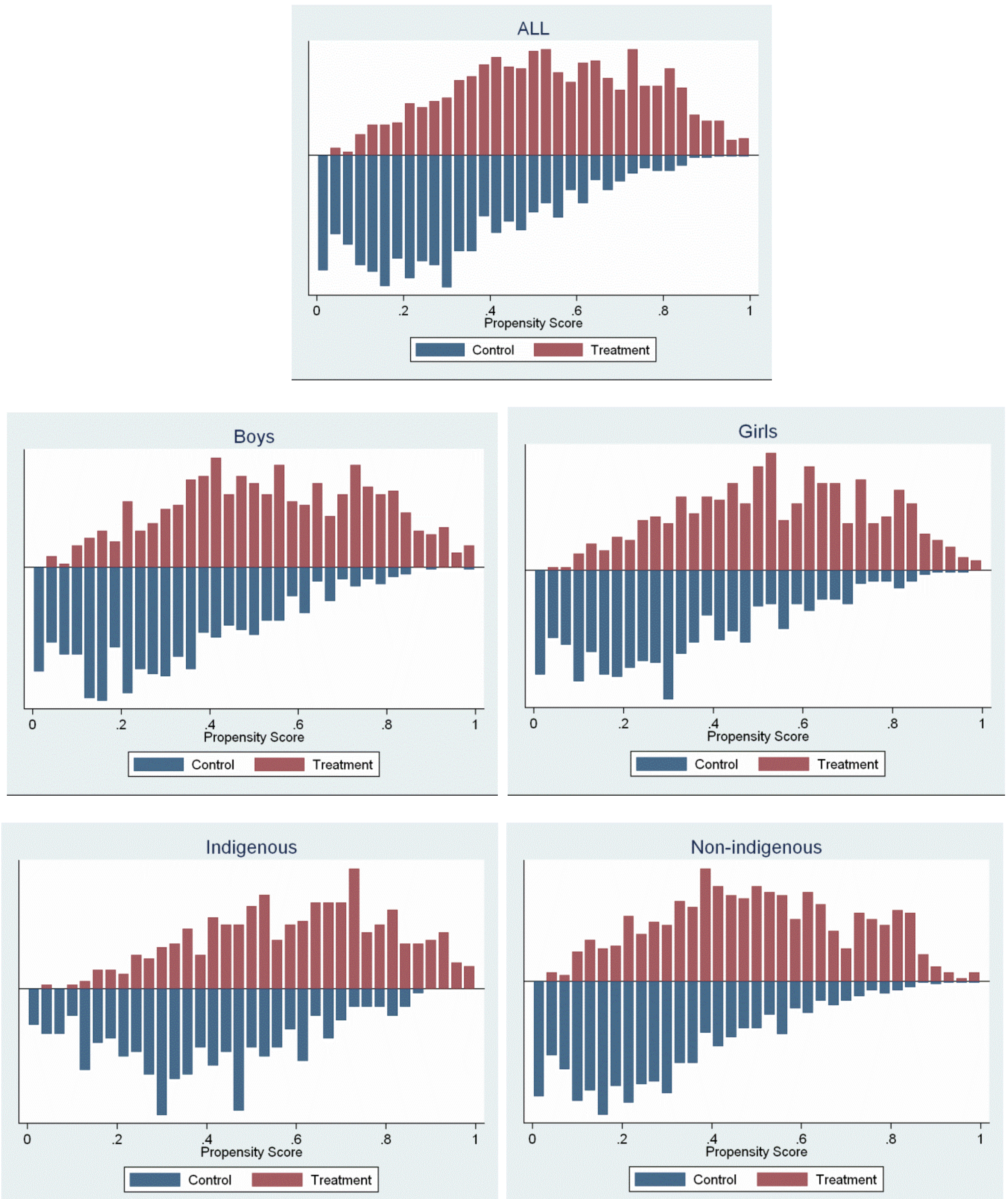
Appendix 1

Table A1. Characteristics of the households in the sample. Logistic regression with dependent variable as 1 if observation belongs to treatment, 0 otherwise.

Variable	Coefficient	SE	z	Variable	Coefficient	SE	z
Height mother	-0.021	0.009	-2.24	Draft animals	0.170	0.114	1.49
PPVT score mother	-0.001	0.002	-0.46	Land ownership	0.531	0.106	5.01
age of child (months)	0.004	0.003	1.09	Blender	0.176	0.133	1.33
Age HH head	0.013	0.007	1.91	Fridge	-0.112	0.194	-0.58
Age spouse	0.011	0.007	1.53	Gas stove	-0.358	0.146	-2.44
Sex HH head	2.215	0.352	6.30	Gas heater	-0.655	0.363	-1.81
Indigenous HH head	0.664	0.274	2.42	Radio	0.592	0.100	5.87
Indigenous spouse	-0.256	0.278	-0.92	Hifi	0.351	0.252	1.39
Educ HH head	0.244	0.115	2.13	TV	0.643	0.119	5.41
Educ spouse	0.410	0.118	3.46	Video	-0.515	0.345	-1.49
Work HH head	-1.085	0.264	-4.11	Wash machine	0.038	0.329	0.11
Work spouse	-0.611	0.163	-3.76	Car	-1.213	0.468	-2.59
# Children 0-5 years	0.087	0.048	1.81	Truck	-0.208	0.285	-0.73
# Children 6-12 years	0.206	0.042	4.91	Guerrero	0.537	0.190	2.83
# Children 16-20 years	0.025	0.074	0.33	Hidalgo	0.945	0.211	4.49
# Children 13-15 years	0.166	0.084	1.96	Michoacán	0.646	0.179	3.60
# Women 20-39 years	0.025	0.120	0.21	Puebla	1.085	0.150	7.21
# Women 40-59 years	-0.022	0.156	-0.14	Querétaro	-0.131	0.220	-0.60
# Women 60+ years	-0.020	0.184	-0.11	San Luis Potosí	0.458	0.154	2.97
# Men 20-39 years	-6.777	0.236	-2.87	Missing Age Spouse	4.332	0.716	6.05
# Men 40-59 years	-0.367	0.161	-2.27	Missing Indg Hh head	-0.421	1.776	-0.24
# Men 60+ years	-0.677	0.236	-2.87	Missing Indg Spouse	1.878	1.695	1.11
#Rooms	0.007	0.047	0.15	Missing Work Hh head	-3.094	1.669	-1.85
Electricity	-0.017	0.115	-0.15	Missing Work Spouse	-3.593	1.637	-2.19
Running water land	-0.885	0.115	-7.67	Missing water land	-1.155	1.735	-0.67
Running water house	0.435	0.209	2.08	Missing water house	-0.634	0.824	-0.77
Dirtfloor	-0.118	0.119	-0.99	Missing height mother	-3.135	1.444	-2.17
Poor quality roof	0.004	0.109	0.04	Missing PPVT score mother	0.225	0.299	0.75
Poor quality walls	0.471	0.127	3.71	Constant	-0.787	1.470	-0.53
Number of observations	2,712			Pseudo R2	0.1951		
LR Chi2(57)	712.61			Log Likelihood	-1469.78		
Prob. > chi2	0.000						

Appendix 2

Figure A2. Estimated Propensity Scores by group



Appendix 3

Table A2 . Composition of the sample*

	All		Boys		Girls		Indigenous		Non-indigenous	
Complete sample										
	N	%	N	%	N	%	N	%	N	%
Treatment	2,825	64.4	1,443	64.6	1,361	63.7	1,089	74.7	1,663	58.2
Control	1,565	35.6	791	35.4	776	36.3	369	25.3	1,196	41.8
Total	4,390	100.0	2,234	100.0	2,137	100.0	1,458	100.0	2,859	100.0
Achen Index (Behavior problems)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	2,825	64.4	1,443	64.6	1,361	63.7	1,809	83.1	1,663	58.2
Control	1,565	35.6	791	35.4	776	36.3	369	16.9	1,196	41.8
Total	4,390	100.0	2,234	100.0	2,137	100.0	2,178	100.0	2,859	100.0
Woodcock-Johnson (Short Term Memory)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	2,749	64.0	1,404	64.3	1,326	63.5	1,050	74.7	1,629	57.9
Control	1,547	36.0	780	35.7	762	36.5	355	25.3	1,185	42.1
Total	4,296	100.0	2,184	100.0	2,088	100.0	1,405	100.0	2,814	100.0
Woodcock-Johnson (Visual integration)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	2,686	63.9	1,377	64.3	1,290	63.3	1,020	74.9	1,595	57.7
Control	1,517	36.1	764	35.7	748	36.7	342	25.1	1,168	42.3
Total	4,203	100.0	2,141	100.0	2,038	100.0	1,362	100.0	2,763	100.0
Communicative Development Inventories (CDI)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	526	61.9	265	61.8	259	61.8	196	69.8	319	57.2
Control	324	38.1	164	38.2	160	38.2	85	30.2	239	42.8
Total	850	100.0	429	100.0	419	100.0	281	100.0	558	100.0
Peabody Picture Vocabulary Test (PPVT)										
	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	2,194	64.4	1,128	65.0	1,050	63.7	838	75.6	1,297	58.1
Control	1,212	35.6	608	35.0	599	36.3	270	24.4	935	41.9
Total	3,406	100.0	1,736	100.0	1,649	100.0	1,108	100.0	2,232	100.0

** Sensitivity analysis comparing treatment-1998 and treatment-2000 with control 2003*

Table A3. Stochastic dominance results by gender and race comparing treatment-1998 and treatment-2000 with control 2003.

Panel A. Achen Index (Behavioral problems)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,50]	[3,36]	0.037	0.954	0.029	0.940	0.92	0.91
	Positive	10%	[0,50]	[1,37]	0.017	0.959	0.014	0.947	0.94	0.93
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
BOYS	Positive	5%	[0,50]	[3,38]	0.033	0.963	0.027	0.954	0.93	0.93
	Positive	10%	[0,50]	[3,38]	0.033	0.963	0.027	0.954	0.93	0.93
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
GIRLS	Positive	5%	[0,50]	[1,33]	0.018	0.927	0.014	0.909	0.91	0.89
	Positive	10%	[0,50]	[1,33]	0.018	0.927	0.014	0.909	0.91	0.89
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
INDIGENOUS	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	none						
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
NON-INDIGENOUS	Positive	5%	[0,50]	[4,36]	0.049	0.957	0.038	0.938	0.91	0.90
	Positive	10%	[0,50]	[1,42]	0.016	0.992	0.010	0.987	0.98	0.98
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						

Panel B. Woodcock-Johnson (Short Term Memory)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,55]	[13,15]	0.242	0.303	0.247	0.310	0.06	0.06
	Positive	10%	[0,55]	[6,19]	0.114	0.399	0.114	0.388	0.29	0.27
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	[30,36]	0.723	0.962	0.693	0.952	0.24	0.26
BOYS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	[1,22]	0.066	0.486	0.067	0.464	0.42	0.40
	Negative	5%	[0,55]	[31,34]	0.772	0.918	0.721	0.879	0.15	0.16
	Negative	10%	[0,55]	[28,51]	0.653	0.999	0.614	0.997	0.35	0.38
GIRLS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	none						
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	none						
INDIGENOUS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	none						
	Negative	5%	[0,55]	none						
	Negative	10%	[0,55]	none						
NON-INDIGENOUS	Positive	5%	[0,55]	none						
	Positive	10%	[0,55]	[1,15]	0.042	0.252	0.050	0.257	0.21	0.21
	Negative	5%	[0,55]	[30,34]	0.656	0.884	0.640	0.870	0.23	0.23
	Negative	10%	[0,55]	[30,34]	0.656	0.884	0.640	0.870	0.23	0.23

Panel C. Woodcock-Johnson (Visual integration)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	[22,24]	0.963	0.993	0.958	0.992	0.03	0.03
	Negative	5%	[0,42]	[9,20]	0.423	0.945	0.407	0.932	0.52	0.53
	Negative	10%	[0,42]	[9,20]	0.423	0.945	0.407	0.932	0.52	0.53
BOYS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	[2,21]	0.167	0.960	0.161	0.958	0.79	0.80
	Negative	10%	[0,42]	[1,21]	0.155	0.960	0.158	0.958	0.81	0.80
GIRLS	Positive	5%	[0,42]	[22,24]	0.966	0.990	0.957	0.989	0.02	0.03
	Positive	10%	[0,42]	[21,22]	0.951	0.994	0.936	0.991	0.04	0.05
	Negative	5%	[0,42]	[12,20]	0.605	0.951	0.586	0.936	0.35	0.35
	Negative	10%	[0,42]	[9,20]	0.417	0.951	0.422	0.936	0.53	0.51
INDIGENOUS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	none						
	Negative	5%	[0,42]	none						
	Negative	10%	[0,42]	none						
NON-INDIGENOUS	Positive	5%	[0,42]	none						
	Positive	10%	[0,42]	[22,25]	0.956	0.992	0.952	0.991	0.04	0.04
	Negative	5%	[0,42]	[9,20]	0.381	0.932	0.364	0.921	0.55	0.56
	Negative	10%	[0,42]	[9,20]	0.381	0.932	0.364	0.921	0.55	0.56

Panel D. Communicative Development Inventories (CDI)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	[1,71]	0.004	0.490	0.003	0.491	0.49	0.49
BOYS	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	[5,76]	0.030	0.592	0.037	0.585	0.56	0.55
	Negative	10%	[0,100]	[4,77]	0.023	0.604	0.030	0.610	0.58	0.58
GIRLS	Positive	5%	[0,100]	[72,74]	0.459	0.510	0.519	0.563	0.05	0.04
	Positive	10%	[0,100]	[72,75]	0.459	0.537	0.519	0.569	0.08	0.05
	Negative	5%	[0,100]	[3,4]	0.027	0.042	0.006	0.006	0.02	0.00
	Negative	10%	[0,100]	[3,4]	0.027	0.042	0.006	0.006	0.02	0.00
INDIGENOUS	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	none						
NON-INDIGENOUS	Positive	5%	[0,100]	none						
	Positive	10%	[0,100]	none						
	Negative	5%	[0,100]	none						
	Negative	10%	[0,100]	[2,16]	0.016	0.091	0.004	0.088	0.08	0.08

Panel E. Peabody Picture Vocabulary Test (PPVT)

	I Effect	II Significance	III Range	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	[65,68]	0.999	1.000	0.998	0.999	0.00	0.00
	Negative	5%	[0,77]	[6,58]	0.288	0.997	0.257	0.994	0.71	0.74
	Negative	10%	[0,77]	[5,64]	0.215	0.999	0.196	0.998	0.78	0.80
BOYS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	[5,64]	0.214	0.998	0.189	0.998	0.78	0.81
GIRLS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	[6,58]	0.294	0.997	0.287	0.995	0.70	0.71
	Negative	10%	[0,77]	[4,58]	0.137	0.997	0.135	0.995	0.86	0.86
INDIGENOUS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	none						
	Negative	5%	[0,77]	none						
	Negative	10%	[0,77]	none						
NON-INDIGENOUS	Positive	5%	[0,77]	none						
	Positive	10%	[0,77]	[65,68]	0.998	0.999	0.997	0.999	0.00	0.00
	Negative	5%	[0,77]	[5,64]	0.193	0.998	0.167	0.997	0.81	0.83
	Negative	10%	[0,77]	[5,64]	0.193	0.998	0.167	0.997	0.81	0.83

Appendix 4

Table A4. Composition of the sample*

	All		Boys		Girls		Indigenous		Non-indigenous	
	N	%	N	%	N	%	N	%	N	%
Treatment	712	78.5	343	77.1	359	80.7	257	86.8	405	73.9
Control	195	21.5	102	22.9	86	19.3	39	13.2	143	26.1
Total*	907	100.0	445	100.0	445	100.0	296	100.0	548	100.0

Sensitivity analysis comparing treatment-1998 and treatment-2000 with control 2003. Only children between 4 and 5 years old included.

Table A5. Stochastic dominance results by gender and race comparing treatment-1998 and treatment-2000 with control 2003. Only children between 4 and 5 years old included.

Achen Index (Behavioral problems)

	I Efffect	II Significance	III Rage	IV Range of effect	V Percentile range T		VI Percentile range C		VII Prop. in T	VIII Prop. in C
ALL	Positive	5%	[0,50]	[16,41]	0.488	0.980	0.431	0.969	0.49	0.54
	Positive	10%	[0,50]	[6,41]	0.099	0.980	0.082	0.969	0.88	0.89
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
BOYS	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	[20,36]	0.651	0.946	0.539	0.912	0.30	0.37
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
GIRLS	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	[30,42]	0.886	0.993	0.826	0.977	0.11	0.15
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
INDIGENOU S	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	none						
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						
NON-INDIGENOUS	Positive	5%	[0,50]	none						
	Positive	10%	[0,50]	[7,41]	0.126	0.978	0.098	0.965	0.85	0.87
	Negative	5%	[0,50]	none						
	Negative	10%	[0,50]	none						