Experimental Evidence from a Conditional Cash Transfer Program: Schooling, Learning, Fertility, and Labor Market Outcomes After 10 Years

By TANIA BARHAM, KAREN MACOURS AND JOHN A. MALUCCIO[1]

February 2021

Abstract: Conditional cash transfer (CCT) programs are the anti-poverty program of choice in many developing countries, aiming to improve human capital and break the intergenerational transmission of poverty. A decade after a randomized CCT program began, earlier and greater exposure during primary school led to higher labor market participation for young men and women and higher earnings for men. The analysis highlights the role of the different program components with variation in timing of access to nutrition, health and education translating into substantial differential impacts on learning for men and reproductive health outcomes for women.

JEL Codes: I25, I38, I18, I28, J13
Key words: CCT, long-term effects, labor markets, education, learning, fertility, age of menarche

---

[1] Barham: Department of Economics and IBS, University of Colorado at Boulder, Boulder, CO 80309-0256 (tania.barham@colorado.edu); Macours: Paris School of Economics and INRA, 48 Boulevard Jourdan, 75014 Paris, France (karen.macours@psemail.eu); Maluccio, Department of Economics, Middlebury College, Middlebury VT 05753 (maluccio@middlebury.edu). This paper combines and extends CEPR Discussion Papers 11937 and 13165. Acknowledgments: This research would not have been possible without the support of Ferdinando Regalia of the Inter-American Development Bank (IDB). We gratefully acknowledge generous financial support from IDB, the Initiative for International Impact Evaluation (3ie: OW2.216) and the National Science Foundation (SES 11239945 and 1123993). See also AEA RCT registry AEARCTR-0001572. We are indebted to Veronica Aguilera, Enoe Moncada and the survey team from CIERUNIC for excellent data collection and for their persistence in tracking. We also acknowledge members of the *Red de Protección Social* program team (in particular Leslie Castro, Carold Herrera and Mireille Vijil) for discussions regarding this research and Emma Sanchez Monin for facilitating the data collection process. We thank Teresa Molina Millán, Olga Larios, Jana Parsons and Gisella Kagy for help with data preparation and Vincenzo di Maro for numerous contributions to the tracking survey. Finally, we are grateful for comments received from Esther Duflo, Norbert Schady, Guillermo Cruces, Berk Özler, several anonymous referees and others during presentations at the IDB, Northeast Universities Development Conference 2012, Allied Social Sciences Association 2013, Colby College, Middlebury College, European Development Network Conference 2013, Population Association of America meetings, UCL, Louvain-la-Neuve, Pompeu Fabra and at the Impact Evaluation Network conference of LACEA. All remaining errors and omissions are our own.

I. Introduction

Interventions aimed at increasing the nutrition, health and education of children are often motivated by their potential to break the intergenerational transmission of poverty. The logic is that investment in human capital can improve future outcomes, in particular through higher income generating potential in adulthood. There is substantial evidence of short-term effects on human capital for a variety of such interventions in low- and middle-income countries. Less is known about whether and how interventions live up to their full promise in the longer term, improving the labor market outcomes and economic well-being of the next generation.[2]

A prominent example in this class of interventions is the conditional cash transfer (CCT) program. Started in Mexico and Brazil in 1997, CCTs spread to more than 60 countries worldwide (World Bank, 2015), reaching 25 percent of the population in Latin America by 2013. Numerous rigorous evaluations demonstrate their short-term effects on poverty alleviation, improved nutrition and health for young children, and increased school attainment for older children (Fiszbein and Schady, 2009; Bastagli et al., 2016). Evidence on their long-term effects, however, is mixed (Molina Millán *et al*., 2019). We contribute to filling the evidence gap by analyzing the effect of a randomized CCT program in Nicaragua on a comprehensive set of human capital and labor market outcomes, using a research design in which it is possible to address several key methodological challenges inherent to estimating long-term impacts and permits assessment of underlying mechanisms.

We find that greater exposure to the CCT during critical primary school ages increased education, labor market participation and earnings when the beneficiaries were young adults a decade later. To understand what underlies these effects, we examine multiple intermediate outcomes along the causal pathway by gender. First generation CCT programs like the one we study promoted nutrition, health and education, and targeted children throughout different stages of their lives.[3] Much of the literature has focused on their educational components. But there are several possible mechanisms by which CCTs can affect beneficiaries and all of the program

[2] Important exceptions include evidence on long-term effects of early childhood stimulation (Gertler *et al*., 2014), early childhood nutrition (Hoddinott *et al*., 2008), early childhood health and family planning interventions (Barham *et al*., 2018), deworming (Baird *et al*., 2016), education subsidies and HIV prevention education (Duflo *et al*., 2015) and school vouchers (Bettinger *et al*., 2016).

[3] More recent CCT programs, in contrast, are increasingly designed to focus more narrowly on a single specific objective (e.g., nutrition or health or education) or single specific target group.

components—rather than the education component alone—are potentially important for long-term outcomes. This paper notably highlights the impact of improved food availability and nutrition during primary school ages, which can influence health, and for females the timing of the onset of puberty and subsequent reproductive health outcomes. This is particularly relevant for understanding impacts on young women, since their education, reproductive health and economic outcomes are closely linked.

To analyze the long-term effects in Nicaragua, we exploit the eligibility rules and randomized phase-in of a CCT and estimate differential intent-to-treat (ITT) effects 10 years after the start of the program. All households randomly assigned to the early treatment group were eligible for transfers for three years from late 2000 to late 2003, while all those randomly assigned to late treatment were eligible for modestly lower transfers for three years from 2003 to 2005. In both treatment groups, all households were eligible for conditional transfers for nutrition and health, but only households with children ages 7–13 years who had not completed grade four were eligible for conditional transfers for education. Therefore, the CCT provided incentives to remain in school, as well as means to improve food availability and nutrition.[4] It also increased exposure to reproductive healthcare information and services during young adolescence. Each of these pathways potentially influenced later education, learning, fertility and marriage outcomes, and consequently labor market outcomes.

We collected data in 2010 on individuals and households originally interviewed at baseline prior to randomization and the start of the program in late 2000. Given the potential relationship between program exposure and migration in this mobile group of young adults, all migrants were intensively tracked throughout Nicaragua and Costa Rica, the main international migration destination. Final attrition is between 4 and 22 percent (depending on the outcome), balanced between early and late treatment groups, and results in analysis samples that are balanced on baseline observables. We consider a variety of complementary estimation strategies to gauge the importance of any remaining selectivity.

We estimate differential experimental ITT impacts exploiting the randomized phase-in of the CCT program, which provides experimental variation in the timing and extent of exposure to different program components, even if it does not provide estimates of the absolute long-term

[4] The short-term evaluation results (Maluccio and Flores, 2005) confirmed that the program led to relatively large increases in enrollment, grade attainment and household food expenditures.

program effects. The follow-up survey measured outcomes approximately ten years after the program start, which was seven years after households eligible for the program in the early treatment group had stopped receiving transfers. At that point, individuals 7–13 years old at baseline were transitioning into early adulthood and at an early stage of their labor market trajectories. Evidence on outcomes in early adulthood is key for understanding the mechanisms underlying the long-term effects of CCTs. Throughout the analyses we account for multiple hypothesis testing by grouping outcomes into families, and also test the sharpened familywise error rate (Anderson, 2008).

Results pooled by gender show long-term positive differential impacts in labor market outcomes; we then disaggregate by gender and demonstrate the different underlying mechanisms for males and females pointing to differences between genders in educational experiences, labor market opportunities and the role of fertility for labor market outcomes of young adults in Nicaragua. By the time of the 10-year follow-up, most young men had completed their schooling and virtually all were working. Therefore, this analysis offers a rare opportunity to examine the sustainability of CCT program effects, assessing whether individuals who have received more schooling because of the program performed better on achievement tests and in their initial years in the labor market well after the program ended. The differential ITT estimates indicate exposure to the CCT at ages when they were likely to drop out of primary school led to sustained education and learning gains of nearly 0.2 standard deviations (SD).[5] In addition, they were more likely to have wage work, migrate temporarily for better paid jobs and they earned ~15 percent more per month worked. The pattern of results for young men is in line with standard human capital theory in which higher investment in human capital leads to more education, more learning and subsequent changes in labor market outcomes including higher earnings. In this sense, the males arguably provide straightforward evidence regarding the long-term returns to CCTs and their potential to reduce the intergenerational transmission of poverty through improved educational human capital.

For young women we also find positive and significant differential ITT estimated effects on labor market participation. Education differentials, however, are modest, and there are no

[5] The results indicate that in some contexts CCTs have a role to play in promoting schooling and in helping children learn, a finding in line with some results in the literature (Baird *et al*., 2011; Duque *et al*., 2018), but in contrast to others (Filmer and Schady, 2014; Baird *et al*., 2019).

differential learning effects. The modest differential impacts on educational outcomes for females can in part be explained by prior enrollment patterns, with girls typically remaining in school longer than boys during their early teenage years in Nicaragua, allowing late treatment females to benefit from the program as well and catch up in schooling, even though they were not directly exposed to the education component. Separate non-experimental estimates of the 5-year absolute impact of the program using national census data support this interpretation and show gains in education alongside reductions in early pregnancy and marriage, for both experimental groups. The 10-year experimental impacts also show that the average age of menarche for the early treatment girls, as well as the onset of sexual activity for them, was later than for the late treatment girls. These differences can be linked to the random variation in the timing of the transfer-related nutrition shocks during ages important for physical development. Consistent with the timing, early treatment girls started childbearing later and had lower young adult BMI compared to those in the late treatment. Although the interpretation of the differential effects along the causal pathway for females is hence more complex than for males, taken together the pattern of results for the young women suggests the CCT long-term impacts on their labor market outcomes in part reflect changes in reproductive health outcomes.

This paper bridges two literatures. First, it complements experimental evidence on long-term impacts of exposure to CCTs during school-going ages including Behrman *et al*. (2009a, 2011), Barrera-Osorio *et al*. (2017), Baird *et al*. (2019), Cahyadi *et al*. (2018) and Molina *et al.* (2020).[6] Existing research generally finds that CCTs lead to higher levels of schooling, but is far less conclusive for subsequent outcomes. Most notably, with the exception of Parker and Vogl (2018) who use census data and the non-experimental national rollout of *PROGRESA* in Mexico to analyze differential long-term impacts, there is little evidence of income or labor market gains.

Our paper contributes to these literatures in a number of ways. The extensive tracking of migrants and rich information on the diversified portfolio of economic activities of young adults enables a thorough examination of the long-term impacts on labor market outcomes and income. We then go further by providing experimental evidence on a wide set of outcomes along the causal pathway to young adult labor market and earning outcomes, including education and

[6] More broadly, this paper also relates to evidence of long-term impacts of cash transfers during early childhood (Behrman *et al*., 2009b; Fernald *et al*., 2009; Barham *et al*., 2013; Araujo *et al*., 2018); and to ongoing debates on the longer-term evidence of unconditional cash transfers (Bandiera *et al*., 2017; Banerjee *et al*., 2016; Handa *et al*., 2018; Haushofer and Shapiro, 2018).

nutrition, as well as learning, socio-emotional, and fertility outcomes. This provides a more comprehensive picture of the ways in which CCT exposure during primary-school age years can lead to longer-term outcomes, and highlights the roles of the education and the nutrition and health components of the programs, and how they differ by gender.

To date the longer-term impacts of primary-school age exposure to the nutritional and reproductive health components common to most large CCT programs have received relatively little attention in comparison to exposure to the education components. One reason this is an important gap is that medical and nutritional research make clear that the different ages at which females in particular are exposed to transfers (and the associated effects on household spending and food availability) can have implications for the onset of puberty, and therefore affect later fertility and other outcomes.[7] Indeed, research shows biological sexual maturation may be particularly sensitive to nutritional shocks for young adolescent girls who had poor nutritional status earlier in life (INSERM, 2007), as was almost certainly the case for our study population.

As such, the paper also relates to literature examining the links between schooling, fertility, marriage and labor market outcomes. This includes Field and Ambrus (2008), who exploit cross-sectional variation in the age of menarche to demonstrate that later age of menarche is related to higher schooling and later marriage. And it closely relates to research examining the longer-term effects of scholarships or educational subsidies on education, fertility and labor market outcomes for young women. In this literature, impacts on fertility outcomes are often directly linked to the educational outcomes and incentives because inducing girls to remain in school longer can postpone childbearing (Filmer and Schady, 2014; Duflo *et al*., 2015, 2017; Baird *et al*., 2011). Relatedly, Buchman *et al*. (2018) show that financial incentives for delaying marriage also can translate into increased schooling. Our paper contributes to this literature by highlighting that the nutrition and health related components of transfer programs are additional channels through which fertility outcomes of teenagers may be affected, which is important for later labor market outcomes.[8]

With its findings on learning, fertility and earnings, the paper has important policy implications. Indeed, the lack of conclusive evidence on the long-term effects of CCTs has led

[7] See appendix G for a detailed review of the relevant medical and nutrition literature.

[8] As such, the evidence also complements studies examining the short-term reproductive health and fertility effects of related interventions, most of which focus on the adult beneficiaries (Avitabile, 2012; Lamadrid-Figueroa *et al*., 2010; Stecklov *et al*., 2007; Todd *et al*., 2012). See also Khan *et al*. (2016) for a systematic review.

some to conclude that their potential to reduce the intergenerational transmission of poverty may be limited (Levy and Schady, 2013). The evidence we present for a program closely resembling many other CCTs arguably means there is room for more optimism.

II. The Nicaraguan CCT Program and Its Experimental Design

The *Red de Protección Social* was a CCT implemented by the government of Nicaragua and designed to address both current and future poverty of rural households. Transfers were made every two months and on average were 18 percent of pre-program total household expenditures. The transfers were provided to female caregivers when possible and always came with both formal enforced conditionalities and a strong social marketing message that the money was intended for food, health and educational investments. The household representative signed an agreement about how the money would be spent though household spending was not directly monitored or an explicit program conditionality. Like most first-generation CCTs, the program had two core components.

The first component aimed to improve nutrition and health through a universal bimonthly transfer for which all households were eligible, regardless of size or composition. The transfer was conditional on preventive healthcare visits for children under five years old and on the household representative attending regular health information workshops. Local supply of the necessary preventive health services was increased in parallel with the program. In addition, during the years in which the late treatment group received the program, adolescents were required to attend newly introduced workshops covering reproductive health and contraception, and contraceptive methods were made available to beneficiaries through the healthcare providers.[9] Otherwise, eligibility criteria and conditionality requirements were the same across the early and late treatment groups.

The second component focused on education. All households with a child age 7–13 years old who had not yet completed the fourth grade of primary school were eligible for this additional component. Each eligible household received a fixed bimonthly school attendance cash transfer, about half the size of the nutrition and health transfer and conditional on enrollment and regular

[9] Reproductive and general health information sessions for the (adult) household representative were a conditionality throughout the program. Sessions targeted to adolescents were added in the 2003. Similar information sessions were incorporated into many other CCTs, including Mexico's *PROGRESA* (Fiszbein and Schady, 2009).

attendance of all eligible children. For each eligible child, the household also received a small annual cash transfer intended for school supplies condtional on enrollment at the start of the school year. We refer to the combined school attendance and school supplies transfers as the education transfer.

The scheduled nominal transfers for the late treatment group were modestly lower, about 88 percent of the amounts in early treatment. Actual average transfers fell by about twice that amount with late treatment households receiving 73 percent of the average transfers made to early treatment. This was still a substantial 13 perent of pre-program total household expenditures. The larger reduction in average compared to scheduled transfers was due to a combination of lower program take-up and children aging out of eligibility for the education component in some households. Appendix C provides further details.

Although the Nicaraguan CCT was modeled after Mexico's original CCT *PROGRESA*, two differences between the programs are particularly important for the analyses. First, in Nicaragua beneficiary households were only eligible to receive the program for a fixed period of three years, after which it was not possible to renew eligibility, which was ongoing in *PROGRESA*. Second, education conditionalities and transfers applied only to the first four grades of school, targeting the lower levels of education and higher primary school dropout rates in Nicaragua.

The randomized evaluation was incorporated into the design of the CCT in six rural municipalities from three regions in central and northern Nicaragua, chosen based on their low health and education outcomes and high poverty rates. In these municipalities, 42 of 59 rural localities were selected for evaluation based on a marginality index.[10] A program census of all residents in the 42 localities was collected in May 2000 prior to randomization. A locality-level marginality index was used to categorize them into seven strata of six localities each, and randomization done at the strata level in a public lottery in July 2000. The 21 early treatment localities received their first transfers in November 2000 and were eligible to receive up to three years of transfers, with the last one delivered in late 2003. Households in the 21 late treatment localities were informed the program would start later in their localities and were phased in at the beginning of 2003, also eligible to receive up to three years of transfers. Households in the early

[10] Localities, or census *comarcas* from the 1995 national census, are administrative areas within municipalities that included as many as 10 small communities together totalling approximately 250 households. The marginality index was constructed from the national census including indicators of literacy, family size, and water and sanitation.

treatment did not receive transfers, and were not subject to conditionalies, after 2003. This included attendance at the adolescent workshops which although available to both groups starting in 2003, comprised an ongoing condition beyond 2003 only for the late treatment group. As a result, these services were less completely taken up in early treatment localities. At the end of 2005, the program ended in these municipalities.[11]

Previous short-term analysis of the CCT shows that the randomized sample was balanced at baseline and that there was minimal contamination of the late treatment group (Maluccio and Flores, 2005). Appendix A demonstrates balance for the samples used in this paper. Compliance with the experimental design was high, with approximately 85 percent of households taking up of at least one component of the program.

III. Data

We draw on several data sources for the long-term analysis.

Program Census and Short-term Panel Evaluation Surveys—The 2000 program census provides baseline data on all households in the 42 localities including schooling and household demographics, characteristics and assets. A random sample of 42 households in each of the 42 localities was drawn from the 2000 program census for the short-term evaluation survey; because locality sizes differed, we constructed sample weights for population estimates of the study area. A comprehensive household survey was collected in September 2000, with subsequent rounds in 2002 and 2004. There was no intensive tracking in the short-term evaluation surveys and attrition was approximately 10 percent per round (Maluccio and Flores, 2005).

2010 Long-term Evaluation Survey—The 2010 evaluation survey was collected between November 2009 and November 2011, 9–11 years after the start of the program. The sample frame consisted of all households in the baseline 2000 short-term evaluation panel survey sample and, to increase power, a random oversample of households from the 2000 program census with children who had been 11 years old at the start of the program (specifically, born between January and June 1989) and therefore at relatively high risk of dropping out of school in the years immediately following the start (see section IV.A). All estimates account for the short-term evaluation survey sampling methodology and oversampling using sampling weights constructed

---

[11] Nationally, all CCT programs in Nicaragua were discontinued after 2006. Overall, they benefitted approximately 30,000 households. As such, general equilibrium effects on labor or other markets are likely to be limited.

to provide population estimates of the study area. Although the more detailed data collected in the short-term evaluation surveys is not available for oversampled individuals, there is information on them in the 2000 program census. The target sample included 2,200 individuals 9–12 years old in 2000, the main cohort of interest (see Section IV.A).

The 2010 survey expanded on the short-term evaluation household survey instrument and added a separate, individual-level instrument. The household instrument was answered by the best-informed person available for the interview, so that responses were from the young adults themselves, or the household head or spouse. It included educational outcomes and detailed participation and earnings measures for all economic activities over the last 12 months, distinguishing between work activities conducted while residing in the current home community versus those performed during periods of temporary migration. An additional module also collected the full labor market history for young adults, with questions on participation, location and earnings for all non-agricultural wage jobs and self-employment. Given the seasonal and temporary nature of many of the economic activities carried out, this comprehensive approach was key to accurately capturing labor market returns for this population.

The 2010 individual survey instrument was conducted through direct in-person interviews with the young adults in their homes, and designed to measure individual learning, cognition and socio-emotional outcomes, and marriage market and reproductive health outcomes, as well as to obtain standardized measures of height and weight for all females.[12] It was conducted for all individuals born after January 1, 1988. To measure learning we administered three Spanish language and two math tests. Spanish tests included word identification, spelling and reading fluency. Math tests included math fluency and problem solving with increasing levels of difficulty, similar to the *Peabody Individual Achievement Test* (Markwardt, 1989). In addition we administered two tests that could capture both learning and cognitive development: the *Test de Vocabulario en Imágenes Peabody* (TVIP, the Spanish version of the *Peabody Picture Vocabulary Test*; Dunn *et al*., 1986) assessing receptive vocabulary, and a forward and backward digit span test assessing memory. To measure cognition, we also implemented Raven's colored matrices (Raven *et al*., 1984). The Raven was included to help distinguish more general

[12] We could not collect anthropometrics for males 15 years old and older because of sensitivities of measurements for that age group related to potential military service. Consequently we are unable to evaluate nutritional pathways for males to the same extent as is possible for females in the analyses.

cognitive skills from more specific skills likely acquired in the classroom. An important advantage of all of the tests is that they provided observed, as opposed to self-reported, measures of learning and cognition regardless of schooling status, thereby substantially reducing concerns about social desirability or selectivity biases. We also measured socio-emotional outcomes using Center for Epidemiologic Studies of Depression (CESD) Scale and the Strengths and Difficulties Questionnaire (SDQ). Lastly, the individual survey included a retrospective module about fertility and prior attendance at, and learning in, reproductive health workshops, and age of first sex, if begun. Young women also reported their age of menarche. The fertility module was developed in light of qualitative interviews we carried out in preparation for the long-term evaluation survey that revealed a perception by program beneficiaries of increased early sexual activity of young teens during the program years. There was also a perception of earlier age of menarche and sexual maturity (CIERUNIC, 2009). Appendix D provides a detailed description of the survey instruments.

Nicaraguan Population Censuses—Finally, we also use the two most recent Nicaraguan national censuses (1995 and 2005). They provided repeated cross sections of individuals including demographic, education and migration information.

IV. Methodology

Given the timing of the initial intervention starting in 2000 and of the subsequent data collection, there is no formal pre-analysis plan guiding the analysis in this paper. However, analyses adhere to the original experimental and program designs and the hypotheses investigated follow those outlined in proposals financing the 2010 follow-up evaluation survey that included the oversample.[13]

A. Identification of Experimental Long-term Differential Impacts

To estimate the long-term differential effects of the CCT, we use the randomized timing of exposure to the program, which created exogenous variation in the ages at which children were exposed leading to what is sometimes referred to as a cross-over design. Because both experimental groups received the program, we estimate differential rather than absolute program

[13] Proposals available at Nicaragua CCT Proposals.

effects. Although the education transfers specifically targeted primary school children, all households in both experimental groups were eligible to receive the nutrition and health transfers. In addition, individual children who did not meet the education transfer eligibility requirements themselves often lived in households with other children who were eligible. As a result, the program could have enabled children to stay in school even beyond the formally targeted ages and grade levels. Because the children lived in households exposed to both transfers, differential estimates capture the overall impact of greater exposure to the entire package offered by the CCT program during critical primary school ages. Appendix Table C1 summarizes differential exposure by age.

Following the program design and using the largest eligible cohort potentially exposed to the education conditionalities for which we have detailed survey information in 2010, we first examine all 7–12 year-olds at the start of the program.[14] This analysis most closely aligns with the original experimental design and the short-term evaluation (conducted before the late treatment group entered the program). We then focus in on the narrower 9–12 year-old cohort for whom there was a greater potential difference in exposure to education conditionalities between the early and late treatment groups (as illustrated in Appendix Table C1) as well as random variation in exposure at ages critical for physical development.

We take advantage of the fact that each treatment group received the program for a different fixed three-year period. The 9–12 year-old cohort corresponds to those who in the absence of the program were at risk of having dropped out of school by the time the CCT reached late treatment localities in 2003. Arguably, transfers and conditions at ages when most children are already enrolled in school and likely to continue, or after most have already been out of school for some time, will be less effective than those at ages when children are enrolled but at higher risk of dropping out or have only recently left (de Janvry and Sadoulet 2006).[15] Baseline evidence indicates important gender differences in schooling patterns. Pre-program enrollment rates suggest the ages at which the risk of dropout starts to increase was 11 years for boys and 13

---

[14] Due to their limited time of exposure to the education conditionality and to budgetary constraints in the 2010 evaluation survey, we did not intensively track those who were 13 years old at baseline beyond their original households, or administer the individual-level survey to them. Therefore, we do not have complete information for them and they are not included in the main analyses.

[15] Children who would have been in school even in the absence of the transfer also may have been affected, for example due to the requirement on number of days of attendance and the increase in household resources.

years for girls. Consequently, at some point between 2000 and 2003, prior to late treatment localities receiving the program, boys 9–12 years old at baseline in 2000 in late treatment would have been 12–15 years old and thus more likely to have already dropped out of school with the implication that a program beginning for them in January 2003 might have come too late to affect their enrollment. The same is true for girls 11–12 years old in 2000 who would have been 14–15 years old in 2003.

Analysis of enrollment rates using short-term evaluation data demonstrates program compliance was in line with the above insights (Table 1). In early treatment localities enrollment (and hence compliance with the education transfer conditionality) was nearly universal in 2002 for children ages 7–12 years at the start of the program, although modestly lower for boys 11–12 (92 percent). By 2004, when the program was no longer operating in early treatment, enrollment for children 7–8 years old at baseline in the late treatment, where it was now operating, was similarly close to 100 percent, suggesting high compliance in the late treatment group for this age cohort. It was lower, however, for 9–12 year olds at baseline who by 2004 were 13–16 years old. As a result, both early and late program exposure may have had similar effects on educational outcomes for the younger 7–8 year-old cohort, but this seems less likely for the 9–12 year-olds for whom the program may have arrived too late for those in the late treatment.[16]

Table 2 presents short-term ITT estimates for educational outcomes of the 9–12 year-old cohorts, before and after the phase-out of the early treatment and the phase-in of the late treatment.[17] In 2002—when the early treatment localities had received transfers for two years but late treatment localities had not yet benefited—boys in the 9–12 year-old cohort (panel A) had 0.36 additional grades attained. Moreover, they were 5 percentage points more likely to have completed fourth grade (the grade level implicitly targeted by the program through its conditionality), 18 percentage points more likely to be enrolled, 36 percentage points more likely to attend school regularly (more than 85 percent), and 15 percentage points more likely to be literate (able to read and write according to parental report). Although smaller, effects for girls in 2002 (panel C) were qualitatively similar and also significant.

Between 2002 and 2004 the difference in grades attained (as well as the percent completing fourth grade) between early and late treatment continued to widen, to 0.49 grades for boys (panel

[16] We discuss results for 7–8 year-olds in section V.E.

[17] Estimates are based on equation (1) described in section IV.B.

A) and 0.57 for girls (panel B). At the same time, in line with transfers having ended in the early but begun in the late treatment group, the signs of the effects on enrollment and attendance reverse. The (now negative) differences between early and late treatment for these two indicators are significant for girls, suggesting that the late treatment girls in particular had begun to catch up with the early treatment girls on schooling. Notably, this occurs even though most late treatment girls in the cohort had already aged out and were no longer themselves directly eligible for education transfers or subject to education conditions when the program reached their localities. The differential impact on enrollment for the late treatment girls in 2004, therefore, likely stems from other components of the program that their households were receiving. These short-term patterns point to the potential for substantial catch-up in grades attained for the late treatment girls, indicating that for them more so than for the boys, long-term differential impacts on schooling could underestimate the program's absolute impacts.[18]

Beyond education, for the 9–12 year-old girls, the CCT may have directly affected reproductive health outcomes differently in the early than in the late treatment group through two additional mechanisms—physical or behavioral. Research on physical mechanisms demonstrates that poor childhood nutrition is associated with delayed puberty for girls, and conversely that better childhood nutrition and health associated with earlier menarche (Garn, 1987; Cooper *et al*., 1996; INSERM, 2007). Moreover, sexual maturation may be particularly sensitive to nutritional shocks for young adolescent girls who had poor nutritional status earlier in life (INSERM, 2007), as is likely for our study population.[19] This is important because when the households of late treatment girls in the 9–12 year-old cohort began receiving transfers they were starting their early teenage years (11–14) and a positive shock in nutrition (corresponding with increased expenditures on food) may have accelerated sexual maturity and menarche. In contrast, early treatment girls in the cohort may have benefitted from better nutrition at younger ages and their access to that better nutrition may have been curtailed after the end of transfers to their households (when they were 12–15 year olds), potentially delaying menarche. Either possibility,

[18] Non-experimental matching results using a comparison sample of households from neighboring municipalities reported elsewhere suggest differential impacts are underestimates of the absolute program impact for the education and learning outcomes examined in this paper (Barham *et al*., 2017, 2018).

[19] This sensitivity to early life undernutrition is believed to explain, for instance, the high frequency of early puberty among migrant and adopted children (Mul *et al*., 2002; Parent *et al*., 2003; Gluckman and Hanson, 2006). Direct evidence on early childhood nutritional status for the females in the estimation samples is unavailable, but nearly 50 percent are stunted as young women in 2010 and girls under age three in 2000 from the same localities had a 40 percent prevalence of stunting, both indicating high undernutrition.

or the combination of both, leads to a positive differential between early and late treatment groups in the age of menarche and therefore potentially in the age of first sex. In addition to the physical, behavioral mechanisms also might have a role. All girls in this cohort were exposed to the reproductive health information sessions between 2003 and 2005. Girls in the late treatment group, where program conditionalities were in place, were exposed to a greater degree. Qualitative findings suggest this exposure may have led to an increase in early sexual activity.

In Table 3, we use short-term evaluation data to confirm that exposure to the CCT was indeed related to large experimentally induced shocks in the availability of nutritious food. Both the quantity and the quality of household food consumption changed significantly. In 2002, after households in the early treatment had received transfers for two years and before the late treatment had benefited, per capita food consumption expenditure in households with a girl in the 9–12 year-old cohort was 35 percent higher. Moreover, there was significantly improved nutritional quality with higher consumption shares of animal proteins and fruits and vegetables, and lower shares of staples and other (less nutritious) foods.[20] By 2004, when the program had been operating for more than a year in the late treatment group but had been phased out in the early treatment, the pattern reversed with 13 percent higher food consumption in late treatment (as shown by the negative ITT effect).[21] Average per capita food expenditures in the early treatment group were stable between 2002 and 2004, but in the late treatment increased. The negative sign of the estimated differential in 2004, therefore, mainly appears to reflect a positive nutrition shock for late treatment households rather than an (absolute) negative shock for early treatment.

B. Empirical Specification for Experimental Long-term Differential Impacts

We estimate

$$Y_{il} = \alpha T_l + \boldsymbol{\beta X}_{il} + \varepsilon_{il} \quad (1)$$

[20] Results are nearly identical for households with a 9–12 year-old boy or with either a boy or girl. Nutritional status information for these age cohorts is not available in the short-term evaluation surveys but evidence for a younger cohorts showed that by 2002, the CCT had significantly reduced stunting for those under 5 years old by five percentage points, a 12 percent decline (Maluccio and Flores, 2005).

[21] The absolute value of the differential impacts are smaller in 2004 than in 2002, possibly because transfer sizes were modestly smaller but also because the early treatment group may have continued to invest in better nutrition even after the end of the program as demonstrated for a related Nicraguan CCT program (Macours, Schady and Vakis, 2012).

where Y_{il} is the outcome for individual i in baseline locality l. T is an ITT indicator that equals one for localities randomly assigned to early treatment and zero for late treatment. Analyses use all respondents from both treatment groups in the age cohort being considered, regardless of initial levels of completed schooling or actual program participation. Given randomized assignment, our main specifications limit the set of control variables **X** to age of the individual when the program started in early treatment (binary indicators for 3-month age groups), binary indicators for whether the individual had 0, 1, 2, 3 or 4+ grades attained prior to the program starting and regional fixed effects. All regressions also include strata fixed effects to account for the fact that randomization was stratified by the marginality index level. Regressions are weighted to account for sampling and attrition (section IV.D), providing population estimates of the study area. We assess robustness of this main specification to alternative controls, weights and samples in appendix B.

Standard errors are adjusted for clustering at the locality level. With 42 localities, we also assess whether accidental imbalance related to the number of clusters changes the significance of the results, following Athey and Imbens (2017) and Young (2019) and estimating the exact p-value under the sharp null hypothesis that the treatment effect is zero for all participants.

We examine the effect of the CCT on measures of labor market, education, learning, fertility and socio-emotional outcomes in 2010. To reduce concerns regarding multiple hypothesis testing, we follow Kling *et al*. (2007) and organize specific outcomes into different domains or "families of outcomes." The index for each family is constructed by averaging the z-score of the specific components of the family. We also adjust for multiple hypothesis testing of all families of outcomes at once, using Anderson's (2008) familywise error rate. Finally, in the appendix, we present results where we instead construct the families using inverse covariance weighting.

C. Outcomes

To characterize the labor market we constructed two families: 1) labor market participation which captures participation and temporary migration for work; and 2) earnings which captures labor market returns and earnings for work off the family farm. We consider two different versions of the earnings family to account for the skewed nature of the distribution of earnings for the predominantly rural, young adult population with a large number of zeros and extended right-side tail: 1) the ranks of earnings (Athey and Imbens, 2017); and 2) actual reported earnings trimmed at the top five percent of values. Specific variables used in the earnings family are defined further in section V.A.

The education family includes an indicator of whether the respondent was enrolled in school, his or her highest grade attained (i.e., the number of grades successfully passed), and whether they had completed grade four, after which an individual was no longer eligible for the education transfer or subject to the CCT's education-related conditionalities. To analyze learning and cognition, we categorized the tests into three families. A learning family measures skills most likely acquired in the classroom and comprises the five achievement tests (word identification, spelling, reading and math fluency, and math problems). A learning-and-cognition family includes the tests likely to capture both learning and cognition (receptive vocabulary and memory test). And finally the cognition family has only one test, the Raven's colored matrices.

The fertility family has different components for males and females. Both include whether the individual had had sex by age 15, has any children, and has ever been married. We include marriage because in the rural Nicaraguan context it is common for couples to marry when the woman becomes pregnant. For women only, the fertility family also includes age of menarche and body mass index (BMI). Higher BMI can be indicative both of an earlier onset of puberty and of child bearing.

For socio-emotional outcomes, we conducted an exploratory factor analysis combining all items from both the CESD and SDQ.[22] The analysis pointed to four latent factors broadly capturing optimism, positive and negative self-evaluation, and stress (appendix D). The socio-emotional family is measured as the average of the z-scores of the four factors, after reversing signs for those that have opposite meaning (stress and negative self-perception) so higher values indicate better socio-emotional outcomes.

D. Attrition and Internal Validity

Attrition levels in this study are on par with or lower than those found in related longitudinal studies with similar time horizons and target populations (appendix E). Individuals who could not be found in their original locations were tracked to their new locations throughout Nicaragua and to Costa Rica, the destination of 95 percent of international migrants from the sample. For

[22] We use exploratory factor analysis, as the correlations between items within CESD (and within SDQ) suggested standardized scoring is unlikely to reflect intended latent traits, similar to findings in Laajaj and Macours (2020).

migrants who could not be located we collected information on individual characteristics related to educational, marital, migration and labor market status through proxy reports from the original household. For those select variables information is available for all but 4 percent of males and 6 percent of females 9–12 years old at baseline. Ten percent of males and 16 percent of females in this age cohort could not be tracked to their 2010 location and hence have missing information for outcomes coming from the household-level instrument, which includes labor market outcomes, schooling, migration and marital status. Attrition is higher, 19 percent for males and 22 percent for females, for information collected in the individual-level instrument, which required direct, in-person interviews and includes all tests of learning, cognitive, and socio-emotional outcomes and the reproductive health history.[23]

There are no significant differences in either permanent migration or attrition levels between early and late treatment groups. Attrition differences are smaller than |0.015| with p-values of the differences 0.6 or higher. Attrition also does not affect the balance of baseline observables (appendix Table A1).[24] A comprehensive analysis of attrition (see Appendix E), however, demonstrates it is correlated with some baseline characteristics that are associated with migration patterns, and that these correlations differ between early and late treatment groups. Consequently, even if attrition is balanced, sample selectivity is still a potential concern. Our preferred estimates, therefore, account for attrition selection using inverse probability weighting (IPW), allowing for differences between early and late treatment groups and incorporating information from the survey tracking process to give higher weight to individuals who were more difficult to find because they are more similar to those not found (Molina Millán and Macours 2017). Separate weights are calculated for variables from the household and individual surveys and for males and females, and all constructed weights also incorporate the relevant sampling weights. Appendix E provides further information on tracking and describes the construction of the attrition weights. Appendix Table B1 presents alternative approaches including without attrition weights, an alternative set of IPW weights and Lee bounds, demonstrating that the main results are not driven by the attrition correction.

---

[23] Attrition rates are lower for the 7–8 year-old cohorts.

[24] Attrition rates from the oversample households are not significantly different than for the rest of the sample and also equally balanced across early and late treatment (p-values of the differences 0.5 or higher).

V. Results

A. 2010 Differential Experimental ITT Impacts for the Pooled Sample

We first present the long-term differential experimental ITT impacts for each family of outcomes pooling males and females (Table 4) with results for the 7–12 year-olds at the start of the program in panel A, and 9–12 year-olds in panel B. For both cohorts, there are significant positive differential results for labor market participation (0.17–0.20 SD) and for earnings (0.12–0.14 SD). Greater exposure to the CCT at critical primary-school ages compared to exposure later translated into significantly better labor market outcomes after a decade, when individuals were beginning their adult working lives. Impacts on the education and learning families are also evident, though more modest (approximately 0.10 SD) and consistently significant only for the 9–12 year-old cohort. Point estimates are close to zero (and insignificant) for the socio-emotional outcomes.

B. 2010 Differential Experimental ITT Impacts 9-12 Cohort by Gender

We decompose the pooled results to aid interpretation and shed light on possible mechanisms, distinguishing between males and females who face substantially different constraints, particularly in the labor market.[25] The pooled results mask some important gender differences as evident in Panels C and D of Table 4, where we present the set of families by gender, each of which we examine in turn. We first focus attention on the 9–12 year-old cohorts for which there was greater differential exposure, returning to the wider age range by gender and pooled in Section V.E.

Labor market participation and earnings outcomes—Because 98 percent of males in this cohort are working, labor market effects for them cannot come from increased participation on the extensive margin. Most young men combine work on the family farm (89 percent) with work off the family farm (83 percent). Program impacts on labor market outcomes show participation in off-farm work increased by 6 percentage points (Table 5, panel A). In line with low average schooling (5.5 grades), off-farm work is mostly unskilled including work as agricultural laborers on others' farms or large plantations (not belonging to the household), salaried jobs in the non-

[25] Results for the wider 7–12 year-old cohort by gender are discussed in section V.E and qualitatively consistent with those presented in this subsection.

farm sector (such as construction workers or security guards), and non-agricultural self-employment. Opportunities for such remunerative off-farm work, however, are highly limited in the poor rural communities where the CCT operated. Therefore, we also include seasonal migration for work in the labor market family. Males in the early treatment are 9 percentage points more likely to have migrated temporarily for work in the last 12 months, nearly one-third higher than the average. Young men in the early treatment group are 8 percentage points (one-third) more likely to have held a salaried non-agricultural wage job at some point, and 7 percentage points (one-half) more likely to have held an urban wage job. Combined into a labor market participation family, these indicators show a sizeable 0.27 SD increase in the degree of temporary mobility and a shift in the type of economic activities the young men engage in.

In Table 6, we explore whether the changes in labor market participation and seasonal migration for this cohort are accompanied by differential impacts on earnings.[26] To avoid selectivity bias, all values are unconditional with zero earnings when an individual did not have earnings during the reference period.[27] Results are broadly consistent across the different indicators and methods, with the earnings family outcome showing an overall increase of about 0.2 SD for both the rank (panel A) and the five percent trim values (panel B). Density functions shown in appendix Figure A1 and quantile regressions in appendix Table A2 demonstrate there were positive gains across the earnings distribution. Differential exposure to the CCT during primary school led to substantive increased labor market returns for these young men.

Young women exposed to the CCT program starting when they were girls ages 9–12 also were doing markedly better in the labor market than those exposed three years later (Table 5, panel B). There is a 0.17 SD increase in the labor market participation family driven by increases in work off the family farm (7 percentage points) and a doubling of temporary migration for

[26] We use various measures to distinguish between earnings from any activity (agricultural or non-agriculture, self and salaried employment) and to capture the temporary nature of many activities. We do not, however, include earnings on the family farm, as person-specific individual returns could not be reliably quantified in this setting. Only 11 percent of this cohort is head of their own household and almost all work on the farm of an older household member, typically their father or father-in-law. Moreover, we do not have information on agricultural inputs, preventing us from calculating agricultural profits for household farms. Although in theory it is possible that the higher earnings we measure off-farm are offset by lower earnings on-farm, the ITT differential impacts on participating in on-farm activities are positive (though small and insignificant), indicating that at least on the extensive margin off-farm work does not substitute for on-farm work. Additionally, while total months worked off-farm did not change (and the point estimate is negative), time worked during seasonal migration increases (from about 30 to 40 days on average) and time worked off-farm in the current area of residence decreases.

[27] As relatively few have ever had a nonagricultural salaried job, the unconditional mean of the monthly nonagricultural salary is much lower than the average off-farm earnings in the last 12 months.

work (9 percentage points). Correspondingly, there is a differential effect on earnings of approximately 0.11 SD for both versions of earnings families (Table 6, panels C and D), significant at the 10 percent level.

Greater CCT exposure during critical primary school ages led, after 10 years, to positive and qualitatively similar outcomes for both young men and young women in the labor market. Although the magnitude of these effects is more modest for women than for the men, because all results are unconditional they incorporate women not participating in the (off-farm) labor force and therefore with no earnings, approximately half of the sample. Examination of other outcomes, however, suggests that there are differences by gender on the mechanisms underlying these impacts.

Education and Learning—The first mechanism we consider is education. Pooled estimates suggest relatively small differential impacts on education and learning for the 7–12 year-olds (Table 4, panel A), but larger and significant impacts (0.10 SD) for the 9–12 cohort (panel B). The smaller difference for the combined 7–12 year-old cohort is consistent with the high enrollment rates observed for the 7–8 year-olds in both treatment groups during the program (Table 1). Education and learning results by gender for 9–12 year olds are presented in Table 7.

Males experienced an approximately 0.10 SD program effect on the education family (panel A). This difference is driven by a significant impact of nearly 0.30 grades attained on an average of 5.5, indicating that two-thirds of the significant differential effect estimated in 2004 (Table 2, panel B) persisted through 2010. Early treatment males are also 4.5 percentage points more likely to still be in school. Overall, a substantial minority, 18 percent in the late treatment group, continues to study, typically taking weekend classes so they can both work and attend school.[28] Despite these gains, however, the overall level of completed schooling remains low, with only 75 percent having completed fourth grade.

Table 7, panel A shows that for males increases in schooling are reflected in learning gains of 0.18 SD. Significant differential impacts are observed precisely for the sorts of skills expected to be acquired in school: 0.16 SD for math and 0.20 SD for Spanish.[29] This result is corroborated by

[28] Of those still enrolled in school, only 15 percent are in tertiary education and nearly all others in secondary school, illustrating the extend of accumulated schooling delays.

[29] In fact, early treatment males performed significantly better than those in the late treatment on each of the five individual achievement tests that make up the Math and Spanish Learning family (appendix Table A3).

a significant increase in self-reported literacy (not included in the test-based learning family). Smaller and insignificant differentials are found on the mixed learning and cognition family that could capture both classroom and non-classroom skills (receptive vocabulary and memory).[30] The impact on the Raven, a cognitive test less likely to capture skills learned in the classroom, is close to zero (-0.02 SD), as is expected given that the intervention began in late childhood for this cohort. These results highlight that exposure to the CCT during critical ages in primary school for males led to significant learning. Moreover, the magnitude of the differential effects are non-negligible and in line with absolute results from education interventions in other settings. Together, the results for schooling and learning demonstrate that boys exposed to the CCT during these ages not only have higher grades attained and do better on achievement tests, but also are more likely still studying, suggesting the positive educational outcome differentials may continue to grow.

Examining females, the first important education finding is that both the early and late treatment groups continued to accumulate additional years of schooling after 2004 (when the average was 4.4) reaching 6.8 grades attained for the late treatment group in 2010 (Table 7, panel B). Overall, the education family indicator indicates a modest positive differential the same size as for males, 0.10 SD. The estimated differential effect on grades attained is relatively small (0.18) and not significant, likely a result of continued catch-up for the late treatment group first evident in their higher enrollment rates in 2004 (Table 2). Indeed, even when 19–22 years old in 2010, 30 percent of these young women are still in school, with nearly two-thirds of those in secondary and a third in tertiary education.

In contrast to the results for males, there are no significant differential impacts on learning for females, with point estimates close to zero (Table 7, panel B and appendix Table A3, panel B). The lack of a differential impact could indicate that the CCT did not improve learning for females in either early or late treatment groups, or that there were positive, but similar-sized absolute effects for each group. The latter interpretation is consistent with the observed catch-up in most schooling outcomes. We examine this possibility further in section V.C where we estimate non-experimental absolute program effects and find that it is likely that the early and late treatment groups had similarly-sized positive program impacts.

[30] Achievement (math and Spanish) and cognitive tests were separated as specified in the project funding proposal.

Fertility—In addition to education, fertility is another key potential mechanism underlying the labor market gains, particularly for females for whom child bearing imposes strong limitations on labor market participation and earnings in this context. For males, there is no overall significant differential impact on fertility (appendix Table A4).

For females, however, Table 8, panel A shows significant negative differential impacts on fertility-related outcomes with the overall fertility family 0.17 SD lower. In addition to 2010 fertility outcomes, the family includes variables reflecting reproductive health history. The early treatment group is 11 percentage points less likely to have had sex by age 15 years, a one-third reduction over the mean in the late treatment. This large differential effect is mirrored by a 6 percentage point (10 percent) lower probability of having had a child by 2010 when the young women were 19–22 years old (p-value 0.17).[31] There is a significant differential impact on the age of menarche, with the late treatment group reaching sexual maturity earlier. This pattern is consistent with better (relative) nutrition for them at potentially critical developmental years during which there were demonstrated increases in household expenditures (Table 3) and the possible interruption of good nutrition for the early treatment group at the time their program transfers ended. There is also a strong differential impact on BMI for non-pregnant women in the early treatment group, 0.6 kg/m^2 lower than in late treatment, consistent with later reported ages of menarche and lower fertility. The final columns in Table 8, not included in the fertility family, show that early treatment girls were less likely to have attended the CCT's reproductive health workshops or to know about the PAP test, a health topic emphasized in those sessions. This is consistent with the hypothesis that there was differential exposure to those sessions and the interpretation that the sessions may have influenced fertility behavior, a possible unanticipated result of increased information and temporary access to contraceptives.[32]

Socio-emotional Outcomes—Finally, we examine the effects on socio-emotional outcomes. Although these represent another potential pathway through which the program could have

[31] Consistent with this, by 2004, there was already a negative 3 percentage point, but insignificant, effect on childbirth.

[32] It is possible that respondents could not remember whether certain specific workshops were organized by the CCT or other entities, so that overall participation in the actual CCT-organized workshops may be overstated. Recall error also may affect the information collected on age at menarche. For both of these variables, however, it is unclear whether such measurement error would affect the early and late treatment groups differently. Results for BMI, objectively measured and therefore unlikely to suffer from differential measurement error, are significant and in line with the other reproductive health outcomes including age at menarche. The consistent findings for BMI and age at menarche suggest systematic recall measurement error on age at menarche is not driving the fertility findings.

affected labor market outcomes, the interpretation of them as possible mechanisms is less straightforward, as the causality plausibly runs in the opposite direction as well, with labor market or other outcomes affecting socio-emotional measures.

Table 9 shows small and insignificant differential impacts on socio-emotional outcomes for both males and females. Examining the four factors separately, however, reveals that for males this average impact masks offsetting effects on the different latent traits. Specifically, the differential effects on positive self-evaluation and optimism are positive, significant and large (0.25 SD), possibly reflecting higher learning and earnings for the early treatment group. At the same time, males exhibit more stress, and are more likely to agree with statements reflecting negative self evaluation, although the point estimates for these outcomes are smaller, ~0.16 SD. For females, the individual dimensions do not show differential effects except for a similar increase in negative self-evaluation. Overall, these findings suggest that the differential effects on labor market outcomes and earnings are unlikely to be the result of, or for that matter driving, differences in socio-emotional outcomes for either gender.

Robustness—In Table 4 we show that the results by gender for the different families of outcomes are robust to adjustments for multiple hypothesis testing (using Anderson's familywise error rate) and randomization inference. Appendix B demonstrates results are robust to: 1) various alternative definitions of weights or sample; 2) different assumptions related to attrition; and 3) an alternative family index using inverse covariance weighting that assigns less weight to individual indicators that are highly correlated within the family.

C. 2005 Non-Experimental Absolute Impacts

To further understand the differential program impacts after 10 years, we estimate non-experimental absolute program impacts after five years. At some point during the period 2000-05, the three-year CCT operated in *all* rural areas of the six evaluation municipalities covering over 90 percent of the rural population. In addition, over the period this was the only substantial education-related program operating in the region. Therefore, it is possible to use national census data to approximate absolute CCT program effects on selected outcomes.

Using 1995 and 2005 population census data, Table 10 presents double difference absolute impacts for the 9–12 year-old cohorts at the start of the program comparing changes in outcomes in rural areas of the six program municipalities to changes in outcomes in rural areas of six adjacent municipalities that never received the program, but include localities with marginality index scores similar to the treatment municipalities.[33] This complementary evidence provides absolute impacts after five years, and as individual ITT assignment can be determined using information on municipality of residence five years earlier, is unaffected by domestic migration-related attrition. The double difference results show improvements for the educational indicators in 2005 that are similar or larger than the 2010 differential experimental effects.

Males 9–12 at the start of the program (who would have been 14–17 in 2005) in treatment municipalities had 0.6 more grades than similarly aged males in comparison municipalities, were were nearly 4 percentage points more likely to be enrolled in school and 9 percentage points more likely to be literate. Similarly, females in treatment municipalities had an average of 0.5 more grades attained by 2005, translating into an 11 percentage point increase in having completed grade four, and a 7 percentage point increase in self-reported literacy. Moreover, at that point half of the females in this cohort were still enrolled in school (with an estimated absolute impact on enrollment of nearly 3 percentage points), pointing to the potential for further absolute gains in the years after 2005. While the lack of long-term differential impacts for female learning described in Section V.B could mean that the CCT did not lead to any absolute effects on those outcomes for either treatment group, these positive absolute effects estimated after five years indicate it is more likely that gains experienced by the early treatment group are offset by equivalently sized impacts in the late treatment group.

For the 9–12 year-old female cohort we also observe absolute impacts on early marriage and teenage pregnancy, with reductions of more than 2 percentage points for each. These results represent a relatively large absolute impact on early fertility decisions since less than 10 percent had given birth by the 2005 census. The combination of these negative 5-year absolute impacts alongside the 10-year negative differential effect on fertility suggests that although there are important absolute reductions in early marriage and pregnancy for both groups, the reductions are larger in the early treatment group, possibly due to the late treatment group experiencing an earlier age of menarche leaving them at greater risk of early pregnancy.

---

[33] Appendix F provides the detailed methodology for the double difference results, shows results for different age groups, demonstrates results are robust to alternative comparison groups, and provides placebo evidence indicating no program effect on education or marriage of adults in support of the common trends assumption.

Last, we also can use the 2005 national census data to gauge the extent to which the cohorts of interest in 2010 differ from same-age cohorts five years earlier. Overall, there are substantive improvements that potentially reflect absolute improvements for individuals in the program areas. Enrollment rates for females 19–22 years old in 2005 were 15 percent, half those for the late treatment group in 2010 shown in Table 7. Enrollment also was about one-third lower (13 versus 18 percent) for males. Consistent with these patterns, marriage rates for females 19–22 years old in 2005 were higher (64 versus 61 percent) than for females 19–22 old in 2010 and they were more likely to have had a child (62 versus 53 percent).

D. Interpretation and Mechanisms Underlying in the Labor Market Gains

Although the original randomized design does not allow fully disentangling the mechanisms, the gendered impacts on learning and fertility strongly suggest that the underlying mechanisms leading to the observed improved labor market outcomes differ for males and females.

For males, the CCT led to more education, more learning, and changes in labor market outcomes and higher earnings. This pattern supports a principal rationale for CCT programs and provides straightforward evidence on their potential to reduce the intergenerational transmission of poverty through improved human capital.

The CCT did not, however, lead to large gains in high skilled or formal wage jobs for these men. This is not surprising when one considers the improvements in basic skills were for a population with low average levels of education. Findings from qualitative work done in 2009 help explain how the basic skills attained might have led to higher labor market returns in seasonal migration. Although temporary migration has costs, there are important differences in wages across regions in Nicaragua and wages are substantially higher in Costa Rica. Better math skills may have helped the young men assess the cost-benefit trade-offs of seasonal migration, and better reading and writing skills may have enabled them to more easily complete any relevant paper work, particularly that required for international migration (CIERUNIC 2009).

The lack of findings on learning for females, on the other hand, suggests for them the negative differential effect on fertility played a role in their differential labor market outcomes. This explanation is consistent with the challenges women face combining early motherhood with labor market activities in rural Nicaragua. The scarcity of job opportunities for young women in most rural villages means that off-farm employment often requires commuting to local urban centers or migration to larger cities or Costa Rica. We present three additional pieces of evidence to support the interpretation that the fertility channel was important.

First, we divide the 9–12 year-old female cohort into 9–10 and 11–12 year-old age groups since these two cohorts of girls received nutrition shocks at different ages and that may have resulted in differences in the effects of the shocks on sexual maturity. The results make clear that impacts on the fertility family and its components are driven by girls 9–10 years old in 2000 (Table 8, panel B). These results are consistent with nutrition shocks in 2003 (related to the start of transfers in the late treatment group and, possibly, the end of the transfers in the early treatment) affecting the age of menarche more for the younger girls because most had not yet already reached menarche at the time of these shocks. The differential impacts are large with the probabilities of having had sex by age 15, marriage, or having a child when they were surveyed at ages 19–20 years old each reduced by about 10 percentage points. Impacts are smaller for girls ages 11–12 years at the start of the program but, with the exception of marriage, not significantly different from the 9–10 year-olds (panel C), possibly in part because of the small sample sizes. Results examining the same two narrower age cohorts for the other families of outcomes are presented in appendix Table A5. Point estimates for labor market participation differentials are not correspondingly larger for the 9–10 year-olds, likely because 40 percent in the early treatment group are still in school, compared to 24 percent for the 11–12 year-olds.

Second, we explore the importance of the fertility channel for labor market outcomes by examining the differential results for all families of outcomes for young women with, and separately without, children. Having children is endogenous and indeed we argue in this paper that fertility-related decisions are key intermediate outcomes for understanding long-term differentials. As such, these estimates are only descriptive. Nevertheless, they strongly suggest that the significant differentials on labor market participation and earnings outcomes are nearly all concentrated among women without children. Appendix Table A6 shows that the labor market and earnings differentials are larger and significant for women without children, but there are no significant differentials in schooling or learning for them. Mothers with children, on the other hand, have no significant differentials in labor market outcomes but do show a significant differential in education, albeit from a much lower mean.

Third, to provide support for our interpretation that it is mostly the timing of the nutrition shock during adolescent development rather than other program differences (e.g. timing of health education workshops and smaller transfers in the late treatment area) that resulted in the differences in fertility, we compare females in the early and late treatment groups who received the nutrition shock at the same ages. Specifically, we compare females 9–12 years old at the start of the program in 2000 in the early treatment group with females 9–12 years old when the program started in 2003 in the late treatment.[34] Because this is a comparison of former beneficiaries who across treatment groups have different ages when measured in 2010, we consider only completed or uncensored outcomes. Appendix table A7 shows that there are no significant differences for completing grade 4, early sexual activity, or age of menarche when comparing women who became eligible at the same ages in early versus late treatment groups. This suggests the main results for girls are unlikely to be driven by other program differences between early and late treatment groups, and supports the interpretation that the fertility effects for women are related to the timing of the nutrition shock, i.e., the ages at which girls experienced that shock.[35]

For both genders, any gains in off-farm earnings and the role of temporary migration in improving earnings could also reflect other direct or indirect effects of the CCT. One possible additional mechanism for migration and earnings results could be increased strength of local networks (e.g., by children spending more time together in school or beneficiaries attending the regularly scheduled workshops) which in turn led to better information about different opportunities or lower costs. Another possibility is that there were program spillover effects.[36] Variation in pre-program family networks and geographical variation in the density of the early treatment localities suggest that although some of these additional mechanisms may be at play, we lack statistical power to make conclusive inferences (appendix H).[37]

---

[34] This comparison could also reflect possible cohort effects. We use age-adjusted outcome variables to control for linear secular trends.

[35] Similarly for men, no differences in grade 4 attainment or having had sex by age 15 are observed when comparing those who received transfers at similar ages.

[36] The ITT may underestimate absolute treatment effects if there are positive spillovers between treatment arms. Randomization was done at the locality level with some localities sharing borders so that spillovers are possible.

[37] Another possible spillover includes intra-household spillovers in the late treatment group that could lead to an overestimate of the treatment effect if late treatment households reallocated resources away from the 9–12 year-olds towards younger children who would become eligible after the program started in their locality. There are too few children (less than 17 percent) without younger siblings to consider heterogeneity along these dimensions. That said, as the timing of the start of the program in the late treatment was uncertain at the outset and there is evidence of an overall increase in enrollment patterns in the late treatment even prior to them receiving the CCT (a possible anticipation effect), this type of spillover seems unlikely to have played a major role. If anything, the apparent anticipation effect on enrollment in the late treatment group suggests the differential ITT are underestimates.

E. Differential and Absolute Program Impacts for Other Age Groups

In this section we assess impacts on other age groups eligible for the CCT per the original program design. Individuals in the 7–8 year-old cohort at the start of the program were potentially exposed to the full three years of the two program components in both the early and in the late treatment groups. Therefore, a priori, and given the short-term evidence described in section IV.A, we anticipate the two treatment groups would have benefited to a similar degree and differential effects between them would be smaller than for the 9–12 year-old cohort. The long-term differential effects on families of outcomes for both males and females are indeed relatively small, and all are statistically insignificant (appendix Table A8, panels A and B). Despite this, the findings for the 9–12 year-old cohorts reported above are fairly robust to widening the age group to include the 7 and 8 year-olds (appendix Table A8, panels C and D).

The lack of long-term differential effects for the younger 7–8 year-old cohort does not of course imply they did not benefit. First, there is strong evidence that the early treatment group benefitted in the short-term (appendix Table A9, panels A and B). Second, non-experimental double difference impacts using the national census data (appendix Table F1) show large absolute increases in educational outcomes after five years for both males and females that are statistically indistinguishable from effects on the older 9–12 year-old cohorts. Alongside the differential effects in 2010, these findings suggest that the 7–8 year-old late treatment children were able to catch up on education and learning with the early treatment, possibly as a result of program exposure through 2005.

Finally, children that were 13 years old in 2000 were eligible for the schooling transfers for one year in early treatment group. Although we did not intensively track the 13 year-olds, proxy information is available (on schooling, literacy, marriage, and working off-farm) and in many cases these individuals were interviewed while tracking another household member. After including direct and proxy reports, data on those outcomes is available for approximately 94 (95) percent of the males (females) in the 7–13, 9–13, and 13 year-old age cohorts, with percentages of missing information balanced across early and late treatment groups. Results for the combined 9–13 year-old cohort (panel B) are similar in magnitude and significance to the 9–12 year-old cohort (and even a bit larger for education and marriage outcomes of females) and not statistically different for any of the outcomes (panel C).[38] This confirms the non-experimental findings on absolute impacts in 2005 for this same cohort, with Appendix Table F1 demonstrating positive, if more modest, absolute impacts in 2005 on most educational outcomes for 13-year-old males, and on completed grade four for 13-year-old females.

Finally, we show results for the entire 7–13 year-old cohort incorporating proxy information in Table A10 panel D. Although including the younger cohort reduces the magnitudes of the differential impacts relative to estimates without them, the program effects remain statistically significant at the 10 percent level or below for most outcomes, though the impact on marriage is smaller. This is consistent with the 7–8 year-olds having been too young at the time of exposure to the nutritional and reproductive healthcare components of the program to have affected them in the same way as the older cohorts.

VI. Conclusions

This paper provides new experimental evidence that school-age nutrition, health and education interventions can lead to substantive long-term gains. Differential timing and exposure to a CCT in rural Nicaragua led to long-term differences in labor market participation and earnings. Moreover, the evidence suggests that the mechanisms through which the program components generated benefits differ for men and women.

For young men, exposure to the CCT during ages at which they were vulnerable to school dropout led to substantial education and learning gains. As their increased labor market returns reflect employment gains through temporary migration, one plausible interpretation is that with more education and learning the young men developed core competencies that made them better at finding higher paying jobs further away from home. These impacts are observed at the start of their working lives and potentially set the men on a higher earnings trajectory for the future.

For young women, differential timing of exposure to the CCT—and with it, to better nutrition and healthcare—translated into differential impacts on the age of menarche, BMI, and fertility. We find no differential effects on learning, however, and only modest differential effects on educational attainment. Hence, while we argue there were likely absolute

[38] The analysis incorporating the proxy information also provides an additional check on possible attrition bias of the main results for the 9–12 year-old age group, since the overall attrition rates are substantially lower. Results further suggest that overall findings do not appear to be driven by attrition bias.

improvements in female educational outcomes, there was little experimentally driven difference in the effects of the CCT on them making it less likely effects on education explain the labor market differentials. This makes it plausible that it is the differences in reproductive health outcomes that are driving the positive impacts in labor market outcomes. Even with many of the young women only just beginning their transition to work, the results highlight the importance of accounting for nutrition and health components of CCT or related programs when analyzing long-term impacts on labor market outcomes. In addition, the results suggest that both timing and program design are important; transfers and related nutritional shocks during preteen years can affect the age of menarche, indicating that the age at which interventions target girls can have important consequences. For the girls in the late treatment group, these shocks may have partially offset reproductive health effects such as delayed fertility related to higher schooling, while the opposite shock for the early treatment girls possibly reinforced them.

In addition to highlighting the different mechanisms contributing to long-term gains for both men and women, the findings of sustained learning improvements for young men constitute an important contribution to the literature and the policy debate. This is particularly the case as they contrast with evidence showing limited learning effects of CCTs in the literature, which recently led a high-level panel of experts to conclude that CCTs are not cost-effective when the objective is to address learning gaps (World Bank, 2020). One potentially relevant difference of our study setting is that the population is poorer and starting from much lower levels of education than many of the previously studied CCTs. The cohort had only 1.2 grades attained at the start of the program when they were 9–12 years old. As a result, it is possible that the additional time in school induced by the program led to learning gains more effectively in this context than in settings where children may already have been acquiring basic skills prior to the introduction of a CCT. Although the learning and income gains in this context were substantial, the population started with low initial education levels and continues to face many other constraints. It is unsurprising, therefore, that the program did not completely transform the lives of these young men to fundamentally higher levels of welfare, for example through highly skilled employment.

Finally, the positive impacts on earnings suggest the potential for recouping costs of the CCT. A full cost-benefit analysis would require monetary assessment of all resource costs as well as all of the possible benefits of the CCT on each household member, including those not in our study. The clearest evidence of positive returns are for males in the 9–12 year-old cohort.

Even for that group, however, earnings results likely underestimate the positive long-term annual returns for three reasons: 1) they capture only the differential effects; 2) young men in the early treatment group are still more likely to be enrolled in school perhaps leading to even higher earnings differentials when they complete their schooling; and 3) the returns are measured early in the working life cycle. With these caveats, and if we only use the earnings gains for young men, under fairly conservative assumptions regarding cost and benefit flows, and using a discount rate of 10 percent, the net present value (NPV) turns positive within two decades (appendix I).

Overall, our results show that exposure to a CCT in primary school translated into important and sustained benefits a decade after the program began. Despite skepticism in some policy circles, the evidence we present, alongside a few other recent studies, suggest these widespread interventions, which are back high on the agenda as part of the policy response to the pandemic, may well have a role in reducing the intergenerational transmission of poverty.

VII. References

Anderson, M. 2008. “Multiple Inference and Gender Difference in the Effects of Early Interventions: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of American Statistical Association* 103(484):1481-1495.

Araujo, M. C., M. Bosch, and N. Schady. 2018. “Can Cash Transfers Help Households Escape an Intergenerational Poverty Trap?” in C. Barrett, M.R. Carter and JP Chavas (eds.). *The Economics of Poverty Traps*, University of Chicago Press.

Athey, S., and G.W. Imbens. 2017. “The Econometrics of Randomized Experiments.” Banerjee, A. and E. Duflo, (eds.), *Handbook of Economic Field Experiments*. Volume 1. Elsevier.

Baird, S., C. McIntosh, and B. Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126(4): 1709–1753.

Baird, S., C. McIntosh, and B. Özler. 2019. “When the Money Runs Out: Do Cash Transfers have Sustained on Human Capital Accumulation?” *Journal of Development Economics, 140: 169–185*.

Baird, S., J.H. Hicks, M. Kremer, and E. Miguel. 2016. “Worms at Work: Long-run Impacts of a Child Health Investment.” *Quarterly Journal of Economics* 131 (4): 1637–1680.

Bandiera, O. R. Burgess, S. Gulesci, N.Das, I. Rasul, M. Sulaiman. 2017. “Labor Markets and Poverty in Village Economies.” *Quarterly Journal of Economics*, 132(2):811-870.

Banerjee, A., E. Duflo, R. Chattopadhyay, and J. Shapiro. 2016. “The Long Term Impacts of a “Graduation” Program: Evidence from West Bengal.” Unpublished, MIT.

Barham, T. R. Kuhn, and P. Turner. 2018. “No Place Like Home: Long-Run Impacts of Early Child Health and Family Planning on Economic and Migration Outcomes.” CU Boulder.

Barham, T., K. Macours, and J.A. Maluccio. 2013. “Boys’ Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions,” *American Economic Review: Papers & Proceedings* 103(3): 467–471.

Barham, T., K. Macours, and J.A. Maluccio. 2017. “Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings after Ten Years” *CEPR Discussion Paper* 11937.

Barham, T., K. Macours, and J.A. Maluccio. 2018. “Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women’s Fertility and Labor Market Outcomes” *CEPR Discussion Paper* 13165.

Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E. (2018). “Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia.” *American Economic Journal: Applied Economics,* forthcoming.

Bastagli, F., J. Hagen-Zanker, L.Harman, V. Barca, G. Sturge and T. Schmidt. 2016. Cash Transfers: What does the evidence say? Overseas Development Institute, London, UK.

Behrman, J. R., S.W. Parker and P.E. Todd. 2009a. “Medium-Term Impacts of the *Oportunidades* Conditional Cash Transfer Program on Rural Youth in Mexico.” In: S. Klasen and F. Nowak-Lehmann, eds. *Poverty, Inequality, and Policy in Latin America*, 219–270. Cambridge, United States: MIT Press.

Behrman, J. R., S.W. Parker and P E. 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. R., S.W. Parker and P.E. Todd. 2011. “Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of *PROGRESA*/Oportunidades.” *Journal of Human Resources* 46(1): 93–122.

Bettinger, E., M. Kremer, M. Kugler, C. Medina, C. Posso, and J.E. Saavedra. 2016. "Can Educational Voucher Programs Pay for Themselves?" Unpublished, Stanford.
Blum WF, Englaro P, Hanitsch S, A. Juul, N.T. Hertel, J. Muller, N.E. Skakkebaek, M.L. Heiman, M. Birkett, A.M. Attanasio, W. Kiess, and W. Rascher. 1997. "Plasma leptin levels in healthy children and adolescents: dependence on body mass index, body fat mass, gender, pubertal stage, and testosterone." *J Clin Endocrinol Metab*;82(9):2904–2910
Buchman, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen, 2018. "Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial", mimeo, Duke University.
Cahyadi, N., R. Hanna, B.A. Olken et al. 2018. "Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia." NBER Working Paper 24670. Cambridge, MA, United States: National Bureau of Economic Research.
Centro de Investigación y Estudios Rurales y Urbanos de Nicaragua (CIERUNIC). 2009. "Qualitative Findings for the Evaluation of the Long-term Impact of the *Red de Protección Social* in Nicaragua." Unpublished.
Cooper, C., D. Kuh, P. Egger, M. Wadsworth and D. Barker, 1996. "Childhood growth and age at menarche." *Br. J. Obstet. Gynaecol*. 103(8): 814–817.
de Janvry, A., and E. Sadoulet. 2006. "Making Conditional Cash Transfer Programs More Efficient: Designing for Maximum Effect of the Conditionality." *World Bank Economic Review* 20(1): 1–29.
Duflo, E., P. Dupas, and M. Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105(9): 2757–2797.
Dunn, Lloyd M., D.E. Lugo, E.R. Padilla, and Leota M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody*. Circle Pines, Minnesota: American Guidance Service, Inc.
Duque, V., Rosales, M. F., and Sanchez, F., 2018. "How Do Early-Life Shocks Interact with Subsequent Human-Capital Investments? Evidence from Administrative Data." Mimeo, University of Michigan.
Fernald, L., P. Gertler and L. Neufeld. 2009. "10-Year Effect of Oportunidades, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: a Longitudinal Follow-up Study." *Lancet* 371:828–837.
Filmer, D., and N. Schady. 2014. "The Medium-Term Effects of Scholarships in a Low-Income Country." *Journal of Human Resources* 49(3): 663–694.
Fiszbein, A., and N. Schady. 2009. "Conditional Cash Transfers: Reducing Present and Future Poverty." *World Bank Policy Research Report*. Washington, DC: World Bank.
García, A., Romero, O.L., Attanasio, O. and Pellerano, L. 2012. "Impactos de Largo Plazo del Programa Familias en Acción en Municipios de Menos de 100 mil Habitantes en los Aspectos Claves del Desarrollo del Capital Humano." Technical report, Union Temporal Econometria S.A. SEI. con la asesoria del IFS.
Garn, S.M., 1987. "The secular trend in size and maturational timing and its implications for nutritional assessments: a critical review." *J. Nutr.* 117, 817–823.
Gertler, P. J. Heckman, R. Pinto, A. Zanolini, C. Vermeersch, S. Walker, S.M. Chang, and S. Grantham-McGregor. 2014. "Labor Market Returns to an Early Childhood Stimulation Intervention in Jamaica." *Science* 344(6187) 997–1001.
Gluckman, PD and M.A. Hanson 2006, "Evolution, Development and Timing of Puberty" *Trends in endocrinology and metabolism* 17(1):7-12.

Haushofer, J. and J. Shapiro. 2018. "The Long-Term Impact of Unconditional Cash Transfers: Experimental Evidence from Kenya." Unpublished, Princeton.
Hoddinott, J., J.A. Maluccio, J.R. Behrman, R. Flores, and R. Martorell. 2008. "Effect of a Nutrition Intervention During Early Childhood on Economic Productivity in Guatemalan Adults." *Lancet* 371(9610, 2): 411–416.
INSERM (Institut National de la Sante et de la Recherche Medicale), 2007. *Growth and puberty: secular trends, environmental and genetic factors*. A collective expert report.
Kling, J., J. Liebman and L. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83–119.
Laajaj, R. and K. Macours. 2020. "Measuring Skills in Developing Countries." *Journal of Human Resources* (forthcoming).
Levy, S., and N. Schady. 2013. "Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution." *Journal of Economic Perspectives* 27(2): 193–218.
Macours, K. and R. Vakis. 2016. "Sustaining Impacts When Transfers End: Women Leaders, Aspirations, and Investment in Children." *NBER Working Paper* No. 22871.
Macours, K., N. Schady and R. Vakis. 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.
Maluccio, J. A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social.*" *Research Report 141*. Washington, DC, United States: International Food Policy Research Institute.
Markwardt, F. C. 1989. *Peabody Individual Achievement Test-Revised Manual*. Circle Pines, Minnesota, United States: American Guidance Service.
Molina Millán, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." *CEPR Discussion Paper No. DP11962*
Molina Millán, T., T. Barham, K. Macours, J.A. Maluccio, and M. Stampini. 2019. "Long-term Impacts of Conditional Cash Transfers: Review of the Evidence." *World Bank Research Observer*, 3(1): 2019, 34(1): 119–59.
Molina Millán, K. Macours, J.A. Maluccio and L. Tejerina. "Experimental long-term impacts of early childhood and school-age exposure to a Conditional Cash Transfer Program." *Journal of Development Economics* 2020, 143: 1–20.
Mul, D., W. Oostdijk and S.L. Drop, 2002. "Early puberty in adopted children." *Hormone Research in Pediatrics* 57, 1–9.
Parent A-S, Teilmann G, A. Juul, N.E. Skakkebaek, J. Toppari J, and J-P Bourguignon, 2003. "The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration." *Endocr Rev* 24:668–693
Parker, S. W and T. Vogl. 2018. "Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico." *NBER Working Paper* No. 24303.
Raven, J.C., Court, J.H. and Raven, J. 1984. *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: Coloured Progressive Matrices*. London: H. K. Lewis.
World Bank. 2015. *The State of Social Safety Nets 2015*. Washington, DC: World Bank.
World Bank, 2020. *Cost-effective Approaches to Improve Global Learning*. Recommendations of the Global Education Evidence Advisory Panel, Washington, DC: World Bank.
Young, A., 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results", *Quarterly Journal of Economics*, 134(2): 557-598.

TABLE 1: 2002 AND 2004 AVERAGE ENROLLMENT RATES, BY AGE AT START OF PROGRAM AND GENDER

	Males			Females		
	7–8	9–10	11–12	7–8	9–10	11–12
2002 Early Treatment	1.00	0.97	0.92	0.95	1.00	0.96
2004 Late Treatment	0.95	0.89	0.58	0.97	0.86	0.68

Notes: Outcome data source: 2002 and 2004 short-term evaluation surveys. All means are weighted to account for sampling providing population estimates of the study area.

TABLE 2: 2002 AND 2004 SHORT-TERM EVALUATION SURVEY EXPERIMENTAL IMPACTS ON EDUCATION, 9–12 YEAR-OLD COHORTS BY GENDER

	Grades Attained (years)	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1
	(1)	(2)	(3)	(4)	(5)
Panel A: Males 2002					
ITT	0.361***	0.053*	0.182***	0.360***	0.150***
	(0.094)	(0.031)	(0.042)	(0.055)	(0.034)
N	475	475	475	475	475
Mean LT	2.396	0.277	0.733	0.544	0.735
Panel B: Males 2004					
ITT	0.487***	0.086*	-0.049	-0.100	0.124***
	(0.155)	(0.045)	(0.063)	(0.066)	(0.029)
N	458	458	458	458	458
Mean LT	3.585	0.536	0.626	0.564	0.815
Panel C: Females 2002					
ITT	0.266***	0.048*	0.099***	0.177***	0.096***
	(0.048)	(0.025)	(0.019)	(0.027)	(0.029)
N	450	450	450	450	450
Mean LT	2.952	0.343	0.875	0.781	0.845
Panel D: Females 2004					
ITT	0.573***	0.184***	-0.141***	-0.149**	0.032
	(0.117)	(0.036)	(0.051)	(0.060)	(0.022)
N	394	394	394	394	394
Mean LT	4.357	0.665	0.766	0.682	0.930

Notes: Outcome data source: 2002 and 2004 short-term evaluation surveys. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling providing population estimates of the study area. (Results for other age groups shown in Appendix Table A9). The experimental ITT results compare early to late treatment groups. The late treatment group began receiving the program in 2003 so that the ITT captures absolute effects in 2002, and short-term differential effects in 2004. Grades attained reflect the completed school year prior to each survey year. Attended school for more than 95% of the time is measured over the previous month and is zero for those who were not enrolled in school at the time. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 3: 2002 AND 2004 SHORT-TERM EVALUATION SURVEY EXPERIMENTAL IMPACTS ON CONSUMPTION, HOUSEHOLDS OF THE FEMALE 9–12 YEAR-OLD COHORT

	Log Food Consumption per Capita	Share of Food Consumption On:			
		Animal Protein	Fruit and Vegetables	Staples	Other Food
	(1)	(2)	(3)	(4)	(5)
Panel A: 2002					
ITT	0.345***	0.054***	0.042***	-0.074***	-0.023**
	(0.083)	(0.013)	(0.007)	(0.018)	(0.009)
N	475	475	475	475	475
Mean LT	7.343	0.216	0.059	0.508	0.216
Panel B: 2004					
ITT	-0.133**	-0.037**	0.004	0.023	0.015*
	(0.065)	(0.015)	(0.011)	(0.018)	(0.008)
N	459	465	465	465	465
Mean LT	7.827	0.179	0.089	0.530	0.201

Notes: Outcome data source: 2002 and 2004 short-term evaluation surveys. Regressions include households with females ages 9–12 at the start of the program in November 2000, weighted to account for sampling providing population estimates for households with girls this age in the study area. The experimental ITT results compare early to late treatment groups. The late treatment group began receiving the program in 2003 so that the ITT captures absolute effects in 2002, and short-term differential effects in 2004. Food consumption is measured as the total value of purchased and otherwise obtained food using a comprehensive food consumption module (Maluccio and Flores, 2005). Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS ON ALL FAMILIES OF OUTCOMES, 7–12 AND 9–12 YEAR-OLD COHORTS OF MALES AND FEMALES

	Labor Market Participation Family Z-Score	Earnings Family Z-Score Rank	Earnings Family Z-Score Absolute (5 % Trim)	Education Family Z-score	Learning Family Z-Score	Socio-Emotional Family Z-score	Fertility Family Z-Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Males and Females Ages 7–12							
ITT	0.174***	0.140***	0.120***	0.064*	0.061	0.007	-
	(0.046)	(0.045)	(0.044)	(0.034)	(0.050)	(0.023)	
N	2,897	2,897	2,873	2,898	2,711	2,690	
Panel B: Males and Females Ages 9–12							
ITT	0.198***	0.141***	0.140***	0.097***	0.092*	-0003	-
	(0.055)	(0.048)	(0.050)	(0.030)	(0.048)	(0.030)	
N	1,894	1,894	1,875	1,895	1,734	1,721	
Panel C: Males Ages 9–12							
ITT	0.272***	0.194***	0.192***	0.098**	0.183**	0.053	-0.059
	(0.075)	(0.057)	(0.067)	(0.043)	(0.070)	(0.039)	(0.064)
FWER p	[0.005]	[0.005]	[0.011]	[0.023]	[0.014]	[0.067]	[0.116]
RI p	[0.002]	[0.001]	[0.008]	[0.036]	[0.030]	[0.257]	[0.442]
N	1,006	1,006	997	1,007	907	900	907
Panel D: Females Ages 9–12							
ITT	0.169**	0.116*	0.104*	0.096**	-0.005	-0.053	-0.167***
	(0.074)	(0.061)	(0.060)	(0.040)	(0.057)	(0.050)	(0.060)
FWER p	[0.066]	[0.072]	[0.076]	[0.066]	[0.361]	[0.141]	[0.066]
RI p	[0.038]	[0.072]	[0.098]	[0.020]	[0.936]	[0.325]	[0.025]
N	888	888	878	888	827	821	809

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5–6 using the 2010 individual survey instrument, and column 7 using both. Regressions include males and females ages 7–12 or 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family (shown in subsequent tables), with individual z-scores calculated using the mean and standard deviation of the late treatment group. The fertility family is different for males versus females and therefore not combined. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. FWER p shows in brackets p-values adjusted for multiple hypothesis testing using the sharpened familywise error rate following Anderson (2008) and based on the variables included in the table. RI p shows Fisher exact p-values obtained through randomization inference using Young's (2019) randomization-t. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION FAMILY COMPONENTS AND MIGRATION, 9–12 YEAR-OLD COHORTS BY GENDER

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components			
		Worked Off Farm =1 (Last 12 Months)	Migrated for Work =1 (Last 12 months)	Ever Had a Salaried Non-Agricultural Job =1	Ever worked in Urban Area =1
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
ITT	0.272***	0.062***	0.093***	0.084**	0.065*
	(0.075)	(0.022)	(0.032)	(0.036)	(0.034)
N	1,006	1,006	1,006	998	998
Mean LT		0.828	0.312	0.226	0.127
Panel B: Females					
ITT	0.169**	0.069*	0.087***	0.020	0.016
	(0.074)	(0.038)	(0.024)	(0.037)	(0.031)
N	888	888	887	883	883
Mean LT		0.463	0.074	0.312	0.234

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the four individual components with the average calculated even if some components are missing. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY COMPONENTS, 9–12 YEAR-OLD COHORTS BY GENDER

	Earnings Family Z-Score	Earnings Family Components (all unconditional)			
		Earnings Per Month Worked (Last 12 Months)	Annual Earnings	Maximum Monthly Earnings (Last 12 Months)	Maximum Non-agricultural Salary Ever
	(1)	(2)	(3)	(4)	(5)
Panel A: Rank of Earnings, Males					
ITT	0.194***	41.780**	25.568	43.899**	49.313**
	(0.057)	(19.471)	(18.493)	(19.290)	(19.684)
N	1,006	1,006	1,006	998	1,006
Mean LT		497.2	502.5	486.9	498.3
Panel B: Earnings (C$) Five Percent Trim, Males					
ITT	0.192***	201.152***	595.013	211.421***	142.260*
	(0.067)	(63.624)	(619.322)	(69.318)	(71.919)
N	997	956	956	956	955
Mean LT		1436	8222	1619	227.5
Panel C: Rank of Earnings, Females					
ITT	0.116*	33.758**	28.195*	4.148	31.313*
	(0.061)	(15.706)	(14.011)	(15.038)	(15.544)
N	888	888	888	883	888
Mean LT		414.8	417.4	434.9	415.5
Panel D: Earnings (C$) Five Percent Trim, Females					
ITT	0.104*	97.623	95.031	134.973**	16.071
	(0.060)	(62.401)	(328.513)	(56.948)	(43.562)
N	878	848	856	848	839
Mean LT		464.2	2628	492.9	309.4

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panels B and D are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C$) and the exchange rate is approximately 20 C$ per U.S. dollar. The family z-score is calculated by averaging the z-score for the four individual components with the average calculated even if some components are missing. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 7: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY COMPONENTS. LITERACY AND LEARNING, 9–12 YEAR-OLD COHORTS BY GENDER

	Education Family Z-Score	Education Family Components			Read and Write =1	Learning Families			Mixed Cognition and Learning	Cognition (Raven)
		Grades Attained	Completed Grade 4=1	Enrolled =1		Math and Spanish	Math	Spanish		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Males										
ITT	0.098**	0.288*	0.035	0.045**	0.052**	0.183**	0.160**	0.204**	0.113	-0.016
	(0.043)	(0.167)	(0.024)	(0.021)	(0.021)	(0.070)	(0.069)	(0.081)	(0.082)	(0.095)
N	1,007	1,006	1,006	1,005	1,007	907	905	907	906	906
Mean LT		5.498	0.747	0.181	0.870					
Panel B: Females										
ITT	0.096**	0.177	0.066***	0.022	0.001	-0.005	-0.000	-0.010	-0.047	-0.011
	(0.040)	(0.141)	(0.020)	(0.029)	(0.014)	(0.057)	(0.055)	(0.065)	(0.061)	(0.088)
N	888	888	888	885	888	827	827	826	826	826
Mean LT		6.758	0.825	0.296	0.956					

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–5 measured using the 2010 household survey instrument and columns 6–10 using the 2010 individual survey instrument. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components with the average calculated even if some components are missing. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 8: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR THE FERTILITY FAMILY FOR FEMALES, 9–12 YEAR-OLD COHORT

	Fertility Family Z-Score	Fertility Family Components					Attended CCT Workshop on Reproductive Health =1	Knows What a Pap Test is =1
		Had Sex by Age 15=1	Ever Married =1	Any Children =1	Age of Menarche	Body Mass Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Females Age 9–12								
ITT	-0.167***	-0.109**	-0.039	-0.064	0.249**	-0.656***	-0.058**	-0.063*
	(0.060)	(0.041)	(0.043)	(0.047)	(0.119)	(0.236)	(0.028)	(0.037)
N	809	809	809	809	806	765	749	792
Mean LT		0.287	0.612	0.527	13.11	23.64	0.815	0.751
Panel B: Females Age 9–10 and 11–12 Separately								
ITT Age 9–10	-0.219***	-0.104**	-0.111**	-0.092*	0.317**	-0.650*	-0.070	-0.114**
	(0.071)	(0.051)	(0.052)	(0.053)	(0.136)	(0.326)	(0.048)	(0.056)
ITT Age 11–12	-0.112	-0.114*	0.036	-0.035	0.180	-0.662	-0.045	-0.010
	(0.082)	(0.063)	(0.056)	(0.069)	(0.183)	(0.398)	(0.050)	(0.047)
N	809	809	809	809	806	765	749	792
Mean LT								
Age 9–10		0.293	0.564	0.439	13.01	23.26	0.814	0.710
Age 11–12		0.281	0.661	0.617	13.22	24.03	0.816	0.793
Panel C: P-value of Difference between 11–12 and 9–10								
P-value	0.263	0.898	0.033	0.459	0.533	0.983	0.760	0.166

Notes: Outcome data source: 2010 long-term evaluation survey. Variables are measured using the 2010 individual survey instrument with the exception of marital status measured in the household instrument. Regressions include females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Age of menarche is reversed when it is included in the fertility family. Body mass index does not include women who were pregnant at the time of measurement, and therefore has lower N. The family z-scores are calculated by averaging the z-score for the individual components with the average calculated even if some components are missing. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 9: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR SOCIO-EMOTIONAL FAMILY COMPONENTS, 9–12 YEAR-OLD COHORTS BY GENDER

	Family Z-Score	Socio-emotional Family Components			
		Positive Self Evaluation	Optimism	Stress	Negative Self Evaluation
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
ITT	0.053	0.249**	0.287***	0.170**	0.155*
	(0.039)	(0.093)	(0.078)	(0.071)	(0.086)
N	900	900	900	900	900
Panel B: Females					
ITT	-0.053	-0.012	0.022	0.069	0.153*
	(0.050)	(0.092)	(0.064)	(0.083)	(0.084)
N	821	821	821	821	821

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Socio-emotional components are the first four factors resulting from exploratory factor analysis of all socio-emotional questions. The family z-score is calculated by averaging the z-score for the four individual components. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE 10: 2005 ABSOLUTE IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY, 9–12 YEAR-OLD COHORTS BY GENDER

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married = 1	Has had live birth=1
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Males						
Treatment municipality * 2005 (δ3)	0.597*** (0.078)	0.124*** (0.014)	0.037*** (0.014)	0.091*** (0.013)	-0.003 (0.004)	n.a.
N	18,399	18,399	18,421	18,399	18,421	
R^2	0.107	0.088	0.037	0.107	0.282	
Mean comparison group 2005	3.922	0.559	0.456	3.922	0.001	
Panel B: Females						
Treatment municipality * 2005 (δ3)	0.536*** (0.084)	0.105*** (0.015)	0.028* (0.015)	0.066*** (0.013)	-0.022** (0.011)	-0.023** (0.010)
N	17,061	17,061	17,075	17,061	17,075	14,104
R^2	0.093	0.073	0.025	0.044	0.001	0.004
Mean comparison group 2005	4.542	0.642	0.502	0.830	0.162	0.085

Notes: Outcome data source: 1995 and 2005 Nicaraguan national censuses. Regressions include males or females ages 9–12 in November of 1990 and 2000, respectively, who were living in rural areas of the six program or six comparison municipalities. δ3 is the double-difference estimate of the absolute program effect in 2005. Childbirth information (for females only) is unreported for 17 percent of observations. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Online Appendices for

Experimental Evidence from a Conditional Cash Transfer Program: Schooling, Learning, Fertility, and Labor Market Outcomes After 10 Years

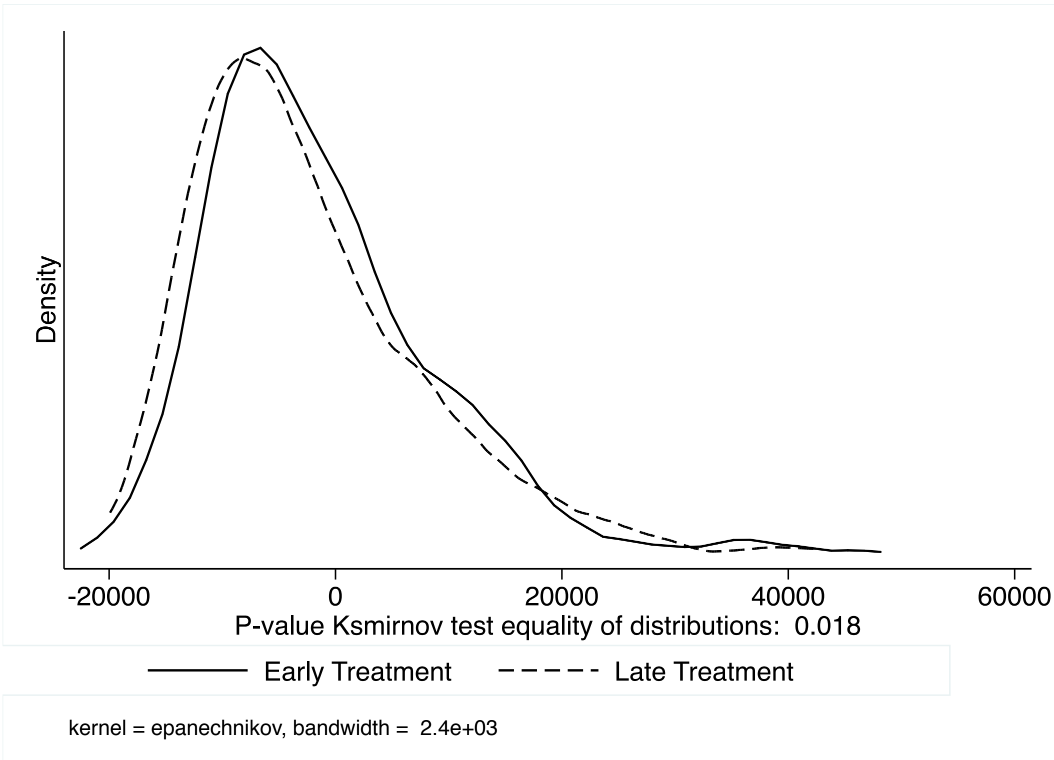
By TANIA BARHAM, KAREN MACOURS, AND JOHN A. MALUCCIO

February 2021

List of Appendices

APPENDIX A: ADDITIONAL TABLES AND FIGURES

FIGURE A1: MONTHLY EARNINGS OFF-FARM, MALES 9–12 IN 2000



Notes: Outcome data source: 2010 long-term evaluation household survey. Sample includes males 9–12 years old at the start of the program in November 2000. Earnings per month worked are demeaned using three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region and trimmed at the top five percent of values.

TABLE A1-A: BASELINE BALANCE FOR 9–12 YEAR-OLD COHORT, MALES

	Early Treatment			Late Treatment			Diff. in Means		Mean/ SD
	Mean	SD	N	Mean	SD	N	Diff.	P-value	
Individual Characteristics									
Age at start of transfers in months	11.0	1.12	516	11.0	0.43	490	-0.02	0.67	-0.06
No grades attained (=1)	0.46	1.31	516	0.43	1.18	490	0.03	0.75	0.02
Highest grade attained	1.21	3.39	516	1.19	2.86	490	0.02	0.88	0.01
Worked last week (=1)	0.17	0.61	516	0.21	0.48	490	-0.04	0.16	-0.09
Mother not living in same household	0.08	0.28	516	0.07	0.33	490	0.01	0.43	0.03
Father not living in same household	0.22	0.66	516	0.18	0.40	490	0.04	0.16	0.11
Child of household head	0.86	0.41	516	0.88	0.35	490	-0.02	0.39	-0.06
Mother no grades attained (=1)	0.45	0.81	516	0.49	0.77	490	-0.04	0.32	-0.05
Mother 3 plus grades attained (=1)	0.37	0.85	516	0.32	0.75	490	0.04	0.33	0.06
Household Head Characteristics									
Age	44.8	14.6	516	44.4	13.2	490	0.40	0.58	0.03
No grades attained (=1)	0.53	0.80	516	0.50	0.55	490	0.04	0.33	0.06
3 plus grades attained (=1)	0.29	0.57	516	0.28	0.47	490	0.02	0.54	0.04
Worked last week (=1)	0.85	0.68	516	0.90	0.39	490	-0.05	0.06	-0.13
Household Characteristics									
Log predicted expenditures (pc)	7.71	0.73	516	7.74	0.68	490	-0.03	0.38	-0.04
Number of household members	8.26	4.57	516	8.22	5.20	490	0.04	0.91	0.01
Number of children aged 0-8	2.10	2.31	516	2.08	2.57	490	0.02	0.92	0.01
Number children aged 9-12	1.76	0.89	516	1.80	1.38	490	-0.04	0.53	-0.03
Log of size of landholdings	7.81	7.38	516	8.10	8.87	490	-0.29	0.57	-0.03
Family network size (individuals)	92.2	201.3	516	68.36	162	490	23.83	0.03	0.15
Own house (=1)	0.81	0.89	516	0.88	0.62	490	-0.08	0.11	-0.12
Some in household work in ag	0.82	0.78	516	0.85	0.86	490	-0.03	0.61	-0.03
Wealth index - housing characteristics	0.10	3.81	516	-0.02	2.90	490	0.13	0.57	0.04
Wealth index - productive assets	-0.13	1.53	516	0.14	2.08	490	-0.27	0.02	-0.13
Wealth index - other assets	-0.04	3.18	516	0.00	3.45	490	-0.03	0.78	-0.01
Number of rooms in house	1.59	1.65	516	1.58	1.34	490	0.01	0.89	0.01
Cement block walls (=1)	0.17	0.90	516	0.15	0.64	490	0.02	0.68	0.04
Zinc roof (=1)	0.54	1.60	516	0.49	1.39	490	0.05	0.62	0.04
Dirt floor (=1)	0.85	0.67	516	0.87	0.58	490	-0.02	0.64	-0.04
Latrine (=1)	0.65	1.25	516	0.60	1.14	490	0.04	0.57	0.04
Electric light (=1)	0.26	1.18	516	0.20	1.01	490	0.06	0.24	0.06
Radio (=1)	0.22	0.59	516	0.22	0.81	490	-0.01	0.86	-0.01
Work animals (=1)	0.11	0.35	516	0.18	0.73	490	-0.07	0.04	-0.09
Fumigation sprayer (=1)	0.31	0.81	516	0.38	1.15	490	-0.07	0.13	-0.06
Distance to nearest school (minutes)	24.9	80.5	516	23.9	66.8	490	1.09	0.84	0.02
Live in Tuma region (=1)	0.52	2.74	516	0.29	2.40	490	0.23	0.13	0.09
Live in Madriz region (=1)	0.19	2.07	516	0.18	2.19	490	0.01	0.89	0.00
Village Population	595	1962	516	335	1431	490	259	0.01	0.18
Characteristics of Nearest School									
Highest grade school offers	4.78	3.34	512	4.85	6.12	484	-0.07	0.80	-0.01
Student-teacher ratio	36.7	28.2	507	36.8	40.7	419	-0.16	0.83	0.00
School under local governance	0.30	2.22	512	0.27	2.09	484	0.03	0.76	0.02

Notes: Outcome data sources: 2000 program census. Results include males ages 9–12 at the start of the program in November 2000 with information in the 2010 household survey. Results are similar using the sample with information in the 2010 individual survey. Means are weighted to account for sampling and attrition. The mean divided by the standard deviation uses the standard deviation for the late treatment group. Standard deviations are clustered at the locality level. The predicted per capita expenditures are calculated from the 2000 program census data using the proxy means method (based on the 1998 Nicaraguan Living Standards Measurement Survey; Maluccio, 2009). The wealth indices are constructed using principal component analysis for the housing characteristics and assets. School characteristics are from program administrative data collected for monitoring conditionalities. Schools under local governance are schools that participated in Nicaraguan's school autonomy reform that provided schools and parents a certain level of autonomy over their own management and operations.

TABLE A1-B: BASELINE BALANCE FOR 9–12 YEAR-OLD COHORT, FEMALES

	Early Treatment			Late Treatment			Diff. in Means		Mean
	Mean	SD	N	Mean	SD	N	Diff.	P-value	/SD
Individual Characteristics									
Age at start of transfers in months	10.98	0.88	448	10.93	1.23	440	0.05	0.51	0.04
No grades attained (=1)	0.41	0.97	448	0.38	1.02	440	0.03	0.69	0.03
Highest grade attained	1.27	2.31	448	1.36	3.24	440	-0.09	0.69	-0.03
Worked last week (=1)	0.02	0.24	448	0.05	0.52	440	-0.03	0.30	-0.05
Mother not living in same household	0.08	0.25	448	0.08	0.22	440	0.01	0.61	0.03
Father not living in same household	0.27	0.79	448	0.28	1.00	440	0.00	0.96	0.00
Child of household head	0.86	0.43	448	0.84	0.49	440	0.02	0.47	0.04
Mother no grades attained (=1)	0.44	0.68	448	0.40	0.68	440	0.04	0.45	0.06
Mother 3 plus grades attained (=1)	0.33	0.73	448	0.39	0.95	440	-0.06	0.30	-0.07
Household Head Characteristics									
Age	45.40	15.87	448	43.43	16.25	440	1.97	0.03	0.12
No grades attained (=1)	0.55	0.53	448	0.48	0.52	440	0.07	0.02	0.14
3 plus grades attained (=1)	0.27	0.47	448	0.31	0.81	440	-0.04	0.35	-0.05
Worked last week (=1)	0.88	0.43	448	0.90	0.38	440	-0.01	0.58	-0.04
Household Characteristics									
Log predicted expenditures (pc)	7.74	0.69	448	7.78	0.54	440	-0.04	0.32	-0.07
Number of household members	8.15	5.08	448	8.27	4.22	440	-0.12	0.74	-0.03
Number of children aged 0-8	2.11	2.76	448	2.18	2.56	440	-0.07	0.66	-0.03
Number children aged 9-12	1.75	0.84	448	1.77	1.19	440	-0.02	0.73	-0.02
Log of size of landholdings	8.20	6.91	448	8.06	8.14	440	0.13	0.84	0.02
Family network size (individuals)	89.84	202.89	448	78.32	198.5	440	11.53	0.34	0.06
Own house (=1)	0.82	0.82	448	0.83	0.83	440	0.00	0.82	0.00
Some in household work in ag.	0.85	0.72	448	0.84	0.84	440	0.01	0.90	0.01
Wealth index - housing characteristics	-0.06	3.32	448	0.23	3.52	440	-0.29	0.23	-0.08
Wealth index - productive assets	-0.05	1.79	448	-0.10	2.53	440	0.06	0.77	0.02
Wealth index - other assets	-0.10	2.82	448	0.05	3.07	440	-0.15	0.42	-0.05
Number of rooms in house	1.58	1.14	448	1.65	1.28	440	-0.07	0.47	-0.05
Cement block walls (=1)	0.19	0.83	448	0.23	0.84	440	-0.05	0.41	-0.06
Zinc roof (=1)	0.57	1.68	448	0.56	1.40	440	0.01	0.92	0.01
Dirt floor (=1)	0.82	0.74	448	0.76	1.25	440	0.05	0.50	0.04
Latrine (=1)	0.59	1.00	448	0.65	1.05	440	-0.06	0.38	-0.06
Electric light (=1)	0.21	0.88	448	0.26	1.24	440	-0.04	0.65	-0.04
Radio (=1)	0.17	0.50	448	0.28	0.96	440	-0.11	0.03	-0.11
Work animals (=1)	0.16	0.55	448	0.16	0.51	440	0.00	0.80	-0.01
Fumigation sprayer (=1)	0.31	0.67	448	0.30	0.86	440	0.01	0.84	0.01
Distance to nearest school (minutes)	31.01	116.82	448	22.72	71.40	440	8.29	0.24	0.12
Live in Tuma region (=1)	0.52	2.51	448	0.36	2.58	440	0.17	0.30	0.06
Live in Madriz region (=1)	0.18	1.84	448	0.19	2.07	440	0.00	0.94	0.00
Village population	612	2552	448	357	1146	440	255	0.02	0.22
Characteristics of Nearest School									
Highest grade school offers	4.71	3.63	441	4.69	5.93	434	0.02	0.88	0.00
Student-teacher ratio	37.38	31.15	440	37.98	35.90	376	-0.60	0.57	-0.02
School under local governance	0.31	2.03	441	0.31	2.12	434	0.00	0.89	0.00

Notes: Results include females ages 9–12 at the start of the program in November 2000 with information in the 2010 household survey. Otherwise, refer to notes to Table A1-A.

TABLE A2: 2010 DIFFERENTIAL QUANTILE REGRESSIONS ON EARNINGS FAMILY (Z-SCORE), MALE 9–12 YEAR-OLD COHORT

	Percentile of Earnings Family								
	10	20	30	40	50	60	70	80	90
ITT	0.219	0.130	0.133	0.150*	0.172*	0.230*	0.223	0.300**	0.154
	(0.153)	(0.134)	(0.087)	(0.082)	(0.095)	(0.132)	(0.141)	(0.137)	(0.216)
N	997	997	997	997	997	997	997	997	997

Notes: Outcome data source: 2010 long-term household evaluation survey. Regressions include males ages 7–12 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010, estimated at the different quantiles. Family z-scores are calculated by averaging the z-score for the individual components in the family (shown in Table 6), with individual z-scores calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION BY TEST (Z-SCORES), 9–12 YEAR-OLD COHORT

	Learning Family Components					Mixed Cognition and Learning Family Components	
	Math Fluency	Math Problems	Reading Fluency	Spelling	Word Identification	Receptive Vocabulary	Memory Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Males							
ITT	0.183**	0.179**	0.137**	0.252***	0.206**	0.128	0.094
	(0.070)	(0.082)	(0.067)	(0.078)	(0.087)	(0.102)	(0.074)
N	907	904	904	898	905	906	902
R^2	0.448	0.349	0.346	0.425	0.358	0.353	0.291
Panel B: Females							
ITT	-0.035	0.036	-0.005	0.036	-0.043	-0.047	-0.042
	(0.064)	(0.059)	(0.059)	(0.075)	(0.079)	(0.075)	(0.069)
N	827	823	824	824	823	825	825
R^2	0.389	0.322	0.464	0.292	0.340	0.326	0.199

Notes: Outcome data source: 2010 long-term evaluation individual survey. Regressions include males or females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR FERTILITY FAMILY COMPONENTS, MALES 9–12 YEAR-OLD COHORT

	Fertility Family Z-Score	Fertility Family Components			Attended CCT Workshop on Reproductive Health =1
		First Had Sex by Age 15=1	Ever Married =1	Any Children =1	
	(1)	(2)	(3)	(4)	(5)
ITT	-0.059	0.080**	-0.094**	-0.048	-0.192***
	(0.064)	(0.035)	(0.042)	(0.038)	(0.032)
N	907	875	907	875	825
Mean LT		0.269	0.311	0.225	0.863

Notes: Outcome data source: 2010 long-term evaluation survey. Variables are measured using the 2010 individual survey instrument with the exception of marital status measured in the household instrument. Regressions include males ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components with the average calculated even if some components are missing. Individual component z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR OTHER FAMILIES, BY AGES 9–10 AND 11–12, FEMALES

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-Score	Socio-Emotional Family Z-Score
		Rank	Absolute (5 % Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: ITT Results By Age							
Age 9–10	0.108	0.127	0.155*	0.090	0.056	-0.219***	-0.004
	(0.093)	(0.089)	(0.090)	(0.286)	(0.074)	(0.071)	(0.067)
Age 11–12	0.235*	0.104	0.049	0.272	-0.060	-0.112	-0.100
	(0.117)	(0.105)	(0.105)	(0.238)	(0.071)	(0.082)	(0.064)
N	888	888	878	888	809	809	809
Panel B: P-value of Difference between 11–12 and 9–10							
P-value	0.398	0.876	0.494	0.687	0.217	0.263	0.269

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5 and 7 using the 2010 individual survey instrument, and column 6 using both as described in Table 8. Regressions include females of the indicated ages at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR AGE 9–12 FOR ALL FAMILIES BY MOTHERHOOD STATUS, FEMALES

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Socio-Emotional Family Z-Score
		Rank	Absolute (5 % Trim)			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Woman Has No Children						
ITT	0.294***	0.261***	0.243***	-0.006	-0.022	0.097
	(0.103)	(0.090)	(0.081)	(0.068)	(0.073)	(0.060)
N	447	447	442	447	412	407
Panel B: Woman Has At least One Child						
ITT	0.117	0.017	0.016	0.180***	0.016	-0.204***
	(0.091)	(0.095)	(0.085)	(0.060)	(0.064)	(0.074)
N	440	440	435	440	415	414

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5–6 using the 2010 individual survey instrument. Regressions include females 9–12 years old at the start of the program in November 2000 (by motherhood status), weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A7: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS INDIVIDUALS EXPOSED TO TREATMENTS AT THE SAME AGES, 9–12 YEARS-OLD WHEN FIRST EXPOSED

	Completed Grade 4 (=1)	Had Sex by age 15 (=1)	Age of Menarche
	(1)	(2)	(3)
Panel A: Males			
ITT	0.0107	0.0147	-
	(0.023)	(0.028)	
N	1022	922	
Panel B: Females			
ITT	0.0095	-0.0321	0.122
	(0.022)	(0.033)	(0.10)
N	934	857	853

Notes: Outcome data source: 2010 long-term evaluation survey. Schooling is measured using the 2010 household survey instrument and the other variables using the 2010 individual survey instrument. Samples include early treatment males and females 9–12 years old at the start of the program for them in November 2000 and late treatment males and females 9–12 years old at the start of the program for them in January 2003. (For this reason, sample sizes differ from other tables.) Regressions are weighted to account for sampling and attrition providing population estimates of the study area. Differential ITT results compare early to late treatment groups in 2010. Controls include indicator variables for stratification and region. The outcome variables are age-adjusted, using the residual of the regression of raw outcome on age estimated on the full sample of early and late treatment. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A8: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR ALL FAMILIES, BY VARIOUS AGE COHORTS

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-score	Socio-Emotional Family Z-Score
		Rank	Absolute (5 % Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Differential Effects Age 7–8, Males							
ITT	0.131	0.160	0.097	-0.008	-0.041	-0.117	-0.013
	(0.091)	(0.096)	(0.091)	(0.079)	(0.080)	(0.153)	(0.045)
N	498	498	496	498	499	548	492
Panel B: Differential Effects Age 7–8, Females							
ITT	0.097	0.076	0.060	0.046	0.079	-0.001	0.079
	(0.101)	(0.097)	(0.102)	(0.073)	(0.096)	(0.079)	(0.062)
N	505	505	502	505	478	473	477
Panel C: Differential Effects Age 7–12, Males							
ITT	0.213***	0.172***	0.150**	0.060	0.107*	-0.096	0.035
	(0.057)	(0.059)	(0.059)	(0.051)	(0.061)	(0.084)	(0.033)
N	1,504	1,504	1,493	1,505	1,406	1,455	1,392
Panel D: Differential Effects Age 7–12, Females							
ITT	0.136**	0.099	0.083	0.070**	0.016	-0.104**	-0.011
	(0.066)	(0.059)	(0.060)	(0.033)	(0.063)	(0.049)	(0.042)
N	1,393	1,393	1,380	1,393	1,305	1,282	1,298

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5 and 7 using the 2010 individual survey instrument, and column 6 using both as described in Tables 8 and A4. Regressions include males or females of the indicated ages at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group. The fertility family is different for males versus females. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE A9-A: 2002 (ABSOLUTE) AND 2004 (DIFFERENTIAL) SHORT-TERM EXPERIMENTAL IMPACTS BY AGE GROUPS, MALES

	Grades Attained (years)	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1
	(1)	(2)	(3)	(4)	(5)
Panel A: 7–8, 2002					
ITT	0.442***	0.000	0.221***	0.230***	0.230***
	(0.135)	(0.000)	(0.064)	(0.064)	(0.081)
N	133	133	133	133	133
Mean LT	0.786	0.000	0.765	0.721	0.469
Panel B: 7–8, 2004					
ITT	0.706***	0.112*	-0.000	-0.028	0.127
	(0.227)	(0.057)	(0.033)	(0.056)	(0.098)
N	130	130	130	130	130
Mean LT	1.992	0.148	0.949	0.862	0.815
Panel C: 7–12, 2002					
ITT	0.412***	0.036*	0.187***	0.311***	0.184***
	(0.079)	(0.020)	(0.034)	(0.043)	(0.034)
N	735	735	735	735	735
Mean LT	1.934	0.186	0.762	0.623	0.666
Panel D: 7–12, 2004					
ITT	0.610***	0.116***	-0.045	-0.076	0.114***
	(0.129)	(0.040)	(0.042)	(0.052)	(0.037)
N	703	703	703	703	703
Mean LT	3.107	0.425	0.742	0.666	0.827

Notes: Outcome data source: 2002 and 2004 short-term evaluation surveys. Regressions include males of indicated ages at the start of the program in November 2000, weighted to account for sampling providing population estimates of the study area. The experimental ITT results compare early to late treatment groups. The late treatment group began receiving the program in 2003 so that the ITT captures absolute effects in 2002, and short-term differential effects in 2004. Grades attained reflect the completed school year prior to each survey year. Attended school for more than 95% of the time is measured over the previous month and is zero for those who were not enrolled in school at the time. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE A9-B: 2002 (ABSOLUTE) AND 2004 (DIFFERENTIAL) SHORT-TERM EXPERIMENTAL IMPACTS BY AGE GROUPS, FEMALES

	Grades Attained (years)	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1
	(1)	(2)	(3)	(4)	(5)
Panel A: 7–8, 2002					
ITT	0.134	0.000	0.182**	0.316***	0.136
	(0.121)	(0.000)	(0.074)	(0.082)	(0.100)
N	131	131	131	131	131
Mean LT	0.769	0.0123	0.767	0.593	0.522
Panel B: 7–8, 2004					
ITT	0.631**	0.017	-0.030	-0.137	0.117
	(0.268)	(0.054)	(0.034)	(0.097)	(0.095)
N	123	123	123	123	123
Mean LT	1.829	0.240	0.973	0.861	0.728
Panel C: 7–12, 2002					
ITT	0.323***	0.031*	0.130***	0.219***	0.131***
	(0.055)	(0.016)	(0.018)	(0.028)	(0.029)
N	713	713	713	713	713
Mean LT	2.161	0.218	0.855	0.751	0.753
Panel D: 7–12, 2004					
ITT	0.655***	0.147***	-0.101***	-0.125**	0.077***
	(0.116)	(0.029)	(0.033)	(0.050)	(0.024)
N	641	641	641	641	641
Mean LT	3.508	0.511	0.849	0.751	0.877

Notes: Regressions include females with indicated ages measured at the start of the program in November 2000. Otherwise, refer to notes on Table A9-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A10-A: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, WITH PROXY VARIABLES BY VARIOUS AGE GROUPS, MALES

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1	Worked Off-Farm =1 (last 12 months)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Differential Effect Age 9–12 - Population Weighted Including Proxy Reports						
ITT	0.353**	0.053**	0.028	0.049**	-0.055*	0.062**
	(0.157)	(0.021)	(0.018)	(0.018)	(0.032)	(0.025)
N	1,072	1,072	1,069	1,073	1,073	1,071
Mean LT	5.391	0.733	0.182	0.864	0.276	0.828
Panel B: Differential Effect Age 9–13 - Population Weighted Including Proxy Reports						
ITT	0.383**	0.059***	0.027	0.042***	-0.061**	0.049*
	(0.143)	(0.016)	(0.019)	(0.015)	(0.029)	(0.025)
N	1,293	1,293	1,288	1,294	1,296	1,291
Mean LT	5.352	0.724	0.172	0.857	0.314	0.825
Panel C: Difference in ITT effect between Age 13 and 9–12						
P-value	0.233	0.404	0.924	0.326	0.269	0.430
Panel D: Differential Effect Age 7–13 - Population Weighted Including Proxy Reports						
ITT	0.287**	0.046***	0.011	0.027*	-0.047**	0.034
	(0.140)	(0.016)	(0.028)	(0.014)	(0.021)	(0.022)
N	1,841	1,841	1,836	1,842	1,844	1,839
Mean LT	5.463	0.742	0.249	0.873	0.240	0.788

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females with indicated ages at the start of the program in November 2000, weighted to account for sampling but not attrition providing population estimates of the study area. Because 13-year-olds were not intensively tracked (appendix E), it was not possible to construct attrition weights similar to those done for the 7–12 cohort. The differential experimental ITT results compare early to late treatment groups in 2010. Proxy reports included for all variables except enrolled because they were not collected. If respondent no longer member of their original household nor member of a split-off household, parental proxy reports obtained in the original household are included. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE A10-B: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS WITH PROXY VARIABLES BY VARIOUS AGE GROUPS, FEMALES

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1	Worked Off-Farm =1 (last 12 months)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Differential Effect Age 9–12 – Using Sampling Weights and Including Proxy Reports						
ITT	0.225	0.054***	0.015	0.016	-0.048	0.043
	(0.140)	(0.018)	(0.023)	(0.013)	(0.033)	(0.030)
N	998	998	993	999	999	996
Mean LT	6.592	0.726	0.275	0.937	0.592	0.462
Panel B: Differential Effect Age 9–13 Using Sampling Weights and Including Proxy Reports						
ITT	0.292**	0.058***	0.010	0.021	-0.056*	0.024
	(0.109)	(0.016)	(0.020)	(0.013)	(0.028)	(0.024)
N	1,227	1,227	1,219	1,229	1,229	1,222
Mean LT	6.466	0.806	0.252	0.926	0.625	0.453
Panel C: Difference in ITT effect between Age 13 and 9–12						
P-value	0.270	0.760	0.949	0.391	0.296	0.190
Panel D: Differential Effect Age 7–13 - Using Sampling Weights and Including Proxy Reports						
ITT	0.234**	0.034**	0.009	0.018**	-0.032	0.001
	(0.087)	(0.014)	(0.020)	(0.009)	(0.028)	(0.022)
N	1,756	1,756	1,748	1,759	1,759	1,751
Mean LT	6.479	0.826	0.331	0.933	0.522	0.438

Notes: Regressions include females with indicated ages measured at the start of the program in November 2000. Otherwise, refer to notes on Table A10-A. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX B: ROBUSTNESS—ALTERNATIVE SPECIFICATIONS AND SAMPLES

In Table B1-A and B, we examine the robustness of the 2010 differential results to a number of alternative specifications and samples for males and females separately. Results from the main specification reported in the paper are reproduced in panel A for comparison.

First, we estimate effects controlling only for strata as per the stratified randomized design of the program.[1] As expected, the standard errors are larger. The results in panel B are consistent with the main results with the exception of the education outcomes for males which, while the point estimates are slightly larger, are no longer statistically significant, and the education family for females, which is smaller and no longer statically significant.

Second, in panel C we included an extended set of controls to test sensitivity to the factors that were not balanced at baseline in Table A1. For males, this includes an indicator for whether the child worked at baseline, a binary variables for the following variables being above the mean: productive assets component of the wealth index, family network size and village population. For females, controls include indicator variables for the household head not having any education, household head age being above the mean, and the population size of the village being above the mean for the population. Results are similar to the main findings for both males and females, although for females the earnings rank family is only marginally significant (p-value=0.11).

Third, we exclude children from households that were oversampled in the 2010 survey, and only keep children who were in the original baseline evaluation survey sample (panel D). This reduces the sample size by about half. Point estimates are broadly similar, for males and females and statistical significance the same for males despite the smaller sample sizes. However, for females, given the reduction in statistical power, results are no longer significant, with the exception of the labor market participation and fertility families. Point estimates for the other families, however, are similar to the main findings. Overall, these estimates confirm that oversampling in the long-term follow-up was important to obtain sufficient statistical power to estimate differential treatment effects.

Fourth, we explore the sensitivity of the analyses to the attrition correction weights (see appendix E for more details). The main results in the paper are weighted to account for attrition, by multiplying the attrition corrected weights and sampling weights. Because attrition rates and baseline characteristics of the interviewed samples are balanced across treatment groups, we re-estimate the main results without any attrition corrections, but with sampling weights in panel E. Tables B2-B6 repeat the more detailed analysis of the variables in each family (labor market participation, earnings, education, and fertility) without attrition weights but with the sampling weights. Alternatively, we also calculate a different set of attrition weights using individuals found during both the regular and the intensive tracking phases, and show inverse probability weighted regressions with these alternative attrition weights multiplied by the sampling weights in panel F. All findings are robust to only using sampling weights (panel E and Tables B2-B6). For the alternative attrition weights (panel F), results are qualitatively similar with the exception of grades attained for females where the point estimate is slightly higher and statistically significant at the 10 percent level.

[1] The main model includes dummies for baseline education levels, region, and 3-month age groups.

Fifth, in panel G we re-estimate effects for the outcome variables derived from the household instrument for the subsample for whom there is also information from the individual-level instrument, and use the weights designed to account for attrition and sampling in the individual survey. Results are robust to this sample restriction for females, and qualitatively similar for males even if the earnings estimates are slightly lower and no longer significant. Variables to construct the fertility family are from the individual-level instrument and therefore not repeated here.

Sixth, to account for correlation between outcomes in the same family in panel H we present results for an alternative index using inverse covariance weighting (Anderson, 2008). Significant effect sizes for males are the same or larger. For females, earnings families are no longer marginally significant but effects on labor market participation and fertility families are even larger.

Finally, panel I shows Lee (2009) bounds, where the bounds assume the tracked sample is either entirely negatively or entirely positively selected (the monotonicity assumption). The Lee bounds are estimated based on ITT estimations without any covariates (for example, without controls for strata and therefore are not directly comparable to the other panels in the table). Given limited sample size, we are unable to trim the samples using covariates to obtain tighter bounds. For males, omitting covariates and strata controls increases point estimates for the grades attained and learning results, with both upper and lower bounds significantly different than zero. Upper and lower bounds for earnings results are also significant. On the other hand, omitting covariates and strata controls reduces the size of the point estimates for the labor market participation and education families, and the lower bounds become insignificant. For females, omitting covariates and strata controls increases point estimates for the labor market and earnings results, with both upper and lower bounds significantly different than zero. On the other hand, omitting covariates and strata controls reduces the size of the point estimates for the fertility family, and the upper bound becomes insignificant. (Lee bounds for the age-of-menarche variable suggest it is bounded between 0.17 and 0.38, not shown). For education, both the upper and lower bounds are insignificant.

TABLE B1-A: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ALTERNATIVE SPECIFICATIONS, MALES 9–12 COHORT

	Labor Market Participation Family Z-Score	Earnings Family Z-Score Rank	Earnings Family Z-Score Absolute (5% Trim)	Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-Score	Socio-emotional Family Z-Score
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Age 9–12 Main results							
ITT	0.272***	0.194***	0.192***	0.098**	0.183**	-0.059	0.053
	(0.075)	(0.057)	(0.067)	(0.043)	(0.070)	(0.064)	(0.039)
N	1,006	1,006	997	1,007	907	907	900
Panel B: Strata Controls Only							
ITT	0.165	0.133**	0.165**	0.109	0.326**	-0.029	0.128*
	(0.099)	(0.059)	(0.063)	(0.078)	(0.141)	(0.077)	(0.070)
N	1,006	1,006	997	1,007	907	907	900
Panel C: Extended Controls							
ITT	0.260***	0.202***	0.214***	0.102**	0.180***	-0.053	0.061
	(0.078)	(0.064)	(0.070)	(0.047)	(0.063)	(0.067)	(0.039)
N	1,006	1,006	997	1,007	907	907	900
Panel D: Excluding Over-Sample Children							
ITT	0.243**	0.178**	0.200**	0.148***	0.232***	-0.002	0.088
	(0.110)	(0.078)	(0.097)	(0.053)	(0.076)	(0.071)	(0.054)
N	537	537	529	538	479	479	476
Panel E: Sampling Weights (No attrition correction)							
ITT	0.266***	0.195***	0.188***	0.105**	0.175***	-0.009	0.033
	(0.070)	(0.057)	(0.069)	(0.040)	(0.063)	(0.055)	(0.041)
N	1,006	1,006	997	1,007	907	1,073	900
Panel F: Attrition Correction using Observations From Intensive & Regular Tracking							
ITT	0.282***	0.195***	0.181**	0.0891*	0.164**	-0.0132	0.0450
	(0.070)	(0.056)	(0.067)	(0.047)	(0.078)	(0.049)	(0.040)
N	1,006	1,006	997	1,007	907	907	900
Panel G: Restrict Household Survey Variables to Individual Survey Sample and Weight							
ITT	0.245***	0.100	0.121	0.138***			
	(0.080)	(0.071)	(0.077)	(0.048)			
N	907	907	898	907			
Panel H: Construction of Families Using Correlations Matrices							
ITT	*0.282****	0.217***	0.243***	0.197***	0.195**	-0.042	0.113
	(0.072)	(0.064)	(0.076)	(0.068)	(0.073)	(0.101)	(0.068)
N	1,006	1,006	997	1,007	907	907	900
Panel I: Lee bounds (no controls)							
Lower Bound	-0.084	0.152**	0.149**	0.018	0.237***	-0.097	0.043
	(0.070)	(0.072)	(0.072)	(0.060)	(0.077)	(0.077)	(0.053)
Upper Bound	0.184**	0.200**	0.239***	0.180***	0.310***	0.049	0.090
	(0.082)	(0.081)	(0.091)	(0.062)	(0.090)	(0.054)	(0.059)

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5 and 7 using the 2010 individual survey instrument, and column 6 using both. Regressions include males ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area unless otherwise indicated. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group unless otherwise indicated. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region unless otherwise indicated. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B1-B: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ALTERNATIVE SPECIFICATIONS, FEMALES 9–12 COHORT

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-Score	Socio-emotional Family Z-Score
		Rank	Absolute (5% Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Age 9–12 Main results							
ITT	0.169**	0.116*	0.104*	0.096**	-0.005	-0.167***	-0.053
	(0.074)	(0.061)	(0.060)	(0.040)	(0.057)	(0.060)	(0.050)
N	888	888	878	888	827	809	821
Panel B: Strata Controls Only							
ITT	0.193***	0.171**	0.148**	0.055	-0.012	-0.141*	-0.039
	(0.070)	(0.070)	(0.071)	(0.069)	(0.069)	(0.074)	(0.063)
N	888	888	878	888	827	809	821
Panel C: Extended Controls							
ITT	0.155*	0.100	0.107*	0.116***	0.010	-0.169***	-0.025
	(0.079)	(0.060)	(0.057)	(0.037)	(0.055)	(0.059)	(0.047)
N	888	888	878	888	827	809	821
Panel D: Excluding Over-Sample Children							
ITT	0.214**	0.137	0.139	0.066	-0.042	-0.136**	-0.077
	(0.098)	(0.103)	(0.095)	(0.049)	(0.054)	(0.063)	(0.063)
N	472	472	467	472	442	431	437
Panel E: Sampling Weights (No attrition correction)							
ITT	0.186***	0.107*	0.103*	0.078**	0.037	-0.117*	-0.072*
	(0.067)	(0.053)	(0.053)	(0.034)	(0.057)	(0.061)	(0.041)
N	888	888	878	888	827	809	821
Panel F: Attrition Correction using Observations From Intensive & Regular Tracking							
ITT	0.204***	0.113**	0.108*	0.0851**	0.0199	-0.106*	-0.0672
	(0.0637)	(0.0559)	(0.0600)	(0.0322)	(0.0557)	(0.0581)	(0.0489)
N	888	888	878	888	827	809	821
Panel G: Restrict Household Survey Variables to Individual survey Sample and Weight							
ITT	0.181**	0.125**	0.124**	0.095*			
	(0.070)	(0.059)	(0.057)	(0.049)			
N	826	826	818	826			
Panel H: *Construction of Families Using Correlations Matrices*							
ITT	0.218**	0.097	0.090	0.094*	0.004	-0.299***	-0.088
	(0.085)	(0.065)	(0.076)	(0.051)	(0.062)	(0.097)	(0.082)
N	888	888	878	888	827	809	821
Panel I: Lee bounds (no controls)							
Lower	0.216***	0.176**	0.164**	0.005	-0.057	-0.114**	-0.123**
Bound	(0.074)	(0.072)	(0.071)	(0.072)	(0.086)	(0.057)	(0.060)
Upper	0.399***	0.208**	0.200*	0.050	0.075	-0.031	-0.029
Bound	(0.097)	(0.090)	(0.111)	(0.068)	(0.085)	(0.062)	(0.058)

Notes: Results are for females, otherwise refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION AND MIGRATION, SAMPLING WEIGHTS 9–12 COHORT

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components			
		Worked Off-Farm =1 (last 12 months)	Migrated for Work =1 (last 12 months)	Ever Had a Salaried Non-Agricultural Job =1	Ever Worked in Urban Area =1
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
ITT	0.266***	0.060***	0.086***	0.083**	0.066**
	(0.070)	(0.022)	(0.027)	(0.034)	(0.032)
N	1,006	1,006	1,006	998	998
Mean LT		0.828	0.309	0.221	0.118
Panel B: Females					
ITT	0.186***	0.033	0.093***	0.040	0.044
	(0.067)	(0.031)	(0.026)	(0.032)	(0.028)
N	888	888	887	883	883
Mean LT		0.485	0.087	0.298	0.213

Notes: Outcome data source: 2010 long-term household evaluation survey. Otherwise, refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS, SAMPLING WEIGHTS 9–12 COHORT

	Family Z-Score	Earnings Family Components – all unconditional			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Monthly Earnings (last 12 months)	Maximum Non-agricultural Salary Ever
	(1)	(2)	(3)	(4)	(5)
Panel A: Rank of Earning, Males					
ITT	0.195***	43.002**	24.689	47.126**	45.817**
	(0.057)	(19.421)	(18.515)	(19.653)	(18.666)
N	1,006	1,006	1,006	1,006	998
Mean LT		493.3	492	494.5	482.6
Panel B: Earnings (C$) — Five Percent Trim, Males					
ITT	0.188***	203.325***	555.966	196.323**	157.602**
	(0.069)	(68.572)	(630.163)	(72.834)	(72.671)
N	997	956	956	956	955
Mean LT		1437	7996	1626	219
Panel C: Rank of Earnings, Females					
ITT	0.107*	27.841*	23.562*	12.873	25.052*
	(0.053)	(14.874)	(12.443)	(12.630)	(14.559)
N	888	888	888	888	883
Mean LT		419.7	421.3	427.8	420.8
Panel D: Earnings (C$) — Five Percent Trim, Females					
ITT	0.103*	82.735	61.230	119.109**	26.449
	(0.053)	(61.254)	(278.319)	(55.517)	(39.218)
N	878	848	856	848	839
Mean LT		469.1	2597	499.1	296.1

Notes: Outcome data source: 2010 long-term household evaluation survey. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Córdobas (C$) and the exchange rate was approximately 20. Otherwise, refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY, SAMPLING WEIGHTS 9–12 COHORT

	Education Family Z-Score	Education Family Components			Read and Write =1
		Grades Attained	Completed Grade 4 =1	Enrolled =1	
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
ITT	0.105**	0.351**	0.048**	0.033	0.053***
	(0.040)	(0.153)	(0.021)	(0.020)	(0.019)
N	1,007	1,006	1,006	1,005	1,007
Mean LT		5.452	0.743	0.185	0.865
Panel B: Females					
ITT	0.078**	0.194	0.050***	0.016	0.005
	(0.034)	(0.134)	(0.016)	(0.025)	(0.015)
N	888	888	888	885	888
Mean LT	6.682	0.837	0.294	0.947	

Notes: Outcome data source: 2010 long-term evaluation individual survey. Otherwise, refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES), SAMPLING WEIGHTS 9–12 COHORT

	Learning			Mixed Cognition and Learning	Cognition (Raven)
	Math and Spanish	Math	Spanish		
	(1)	(2)	(3)	(4)	(5)
Panel A: Males					
ITT	0.175***	0.135**	0.205***	0.109*	0.033
	(0.063)	(0.064)	(0.069)	(0.061)	(0.077)
N	907	905	907	906	906
Panel B: Females					
ITT	0.040	0.053	0.030	-0.042	0.060
	(0.057)	(0.053)	(0.067)	(0.061)	(0.075)
N	826	826	825	825	825

Notes: Outcome data source: 2010 long-term evaluation individual survey. Otherwise, refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR FERTILITY FAMILY OUTCOMES AND MECHANISMS, SAMPLING WEIGHTS, FEMALES 9–12 COHORT

	Fertility Family Z-Score	Fertility Family Components					Attended CCT Workshop on Reproductive Health =1	Knows What a Pap Test is =1
		Age First Had Sex <=15 (=1)	Ever Married =1	Any Children =1	Age of Menarche	Body Mass Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Males								
	-0.008	0.054*	-0.064*	0.001			-0.214***	
	(0.052)	(0.028)	(0.035)	(0.029)			(0.033)	
	875	875	875	875			825	
Panel B: Females								
ITT	-0.117*	-0.043	-0.054	-0.044	0.248*	-0.424**	-0.067***	-0.063
	(0.061)	(0.036)	(0.042)	(0.044)	(0.128)	(0.208)	(0.024)	(0.040)
N	809	809	809	809	806	766	749	792
R^2	0.062	0.091	0.079	0.108	0.053	0.060	0.101	0.078
Mean LT		0.229	0.573	0.488	13.13	23.40	0.816	0.719

Notes: Outcome data source: 2010 long-term evaluation survey. Variables are measured using the 2010 individual survey instrument with the exception of marital status measured in the household instrument. Age of menarche is reversed for fertility family. Body mass index does not include women who were pregnant at the time of measurement. Otherwise, refer to Table B1-A. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX C: TARGETING AND DESIGN OF THE NICARAGUAN *RPS* CCT PROGRAM [2]

The Nicaraguan CCT, *Red de Protección Social* (RPS), was designed to alleviate both current and future poverty through cash transfers targeted to poor and extremely poor households in rural Nicaragua. Its specific stated objectives included: 1) supplementing household income for up to three years to increase expenditures on food; 2) reducing dropout rates during the first four years of primary school; and 3) increasing the nutritional status and healthcare of children under five. A randomized evaluation was incorporated into the initial program design and was carried out by the International Food Policy Research Institute (IFPRI). The program began in 2000 and was implemented by the Government of Nicaragua with technical assistance and financial support from the Inter-American Development Bank (IDB), over two budgetary phases covering six years. The first phase lasted three years with a budget of $11 million. In late 2002, based in part on the positive findings of various program impact evaluations, the Government of Nicaragua and the IDB agreed to a continuation and expansion for a second phase until 2006, with an additional budget of $22 million.

Program Targeting—The CCT first targeted the rural areas in six municipalities from three regions in central and northern Nicaragua, on the basis of poverty as well as on local capacity to implement the program. The focus on rural areas reflected the distribution of poverty in Nicaragua—of the 48 percent of Nicaraguans identified as poor in 1998, 75 percent resided in rural areas (World Bank 2001). While the six targeted municipalities were not the poorest in the country, the proportion of impoverished people living in those areas was still well above the national average (World Bank 2003). At the same time, the selected municipalities had relatively good communication and access (for example, less than one day's drive to Managua, where the program's central administrative office was located), relatively strong institutional capacity and local coordination, and generally adequate access to primary schools.

In the next stage of (pre-program) targeting, a marginality index was constructed for all 59 of the rural census *comarcas* (hereafter localities)[3] within the six selected municipalities. The index was the weighted average of a set of locality-level indicators (including average family size, lack of access to potable water and latrines, and illiteracy rates)—such that localities with higher marginality index scores were more impoverished.[4] The 42 localities with the highest scores were selected for inclusion in the randomized program evaluation (divided into the early and late treatment groups). This is the group of localities used in this paper. Finally, while the initial program design called for all households in these 42 targeted localities to be eligible for the CCT, prior to the start of the program the government excluded approximately three percent of them determined to have substantial resources using information from the registration census, in particular those who owned a vehicle or had large landholdings. These households are excluded from the analyses in the paper.

---

[2] This appendix draws on IFPRI (2005), Maluccio and Flores (2005), and Maluccio (2009).

[3] Census *comarcas* are administrative areas within municipalities based on the 1995 Nicaraguan national census that included as many as 10 small communities for a total of approximately 250 households.

[4] More specifically, the marginality index for each locality included average family size (10 percent), percent without piped water in the home or yard (50 percent), percent without a latrine (10 percent), and percent of persons over age five who were illiterate (30 percent)—all calculated from the 1995 Nicaraguan *National Population and Housing Census*.

While not statistically representative of rural Nicaragua, the 42 localities comprising the randomized evaluation area were nevertheless similar to other rural areas in the program regions and elsewhere in the country. For example, three-quarters of the approximately 1,000 rural localities in the country had marginality index scores in the same range as the program areas. Moreover, poverty rates in the targeted localities were about 10 percentage points higher than rural national rates: 80 versus 69 percent poor and 42 versus 29 percent extremely poor.

During operation, the CCT expanded to the remaining 59 – 42 = 17 rural localities in the six municipalities not initially targeted and not in the randomized evaluation. In those 17 localities, which were less poor according to the marginality index, the CCT began in late 2001 and was offered to approximately 80 percent of the population based on a household-level proxy means targeting model. Consequently, over the course of years 2000–2005 the three-year program had been implemented (at different times and to modestly different degrees) in all (59) rural localities in the six municipalities, and more than 90 percent of the population had been eligible.

Program Components and Conditionalities—The CCT had two core components: 1) food security, nutrition, and health; and 2) education. Corresponding to these, transfers were conditional on household health and education behaviors, with conditionalities monitored by specially contracted healthcare providers and teachers. Conditionalities and benefits were first explained to eligible families in the early treatment group during registration assemblies in September and October 2000 and transfers began in November 2000. The designated household representative (referred to in Spanish as the *titular*) received the cash transfers and, where possible, the program appointed the mother or other female caregiver to this role. As a result, more than 95 percent of the household representatives were women. The CCT also worked with local volunteer coordinators (beneficiary women chosen by the community and referred to in Spanish as the *promotora*) to help implement the program. The coordinators organized and informed their group of household representatives regarding upcoming program activities, upcoming transfer payments, and failure to fulfill the conditions. Conditions were monitored for compliance and when reported by the school or healthcare provider as not having been met, relevant transfers were withheld. The program also had a strong social marketing message that the money was intended to be used for food and education investments and beneficiaries were required to sign an agreement to that effect, although explicit expenditures were not formally monitored.

Food security, nutrition, and health component. Each eligible household received a bimonthly cash transfer known as the nutrition and health transfer that was a fixed amount per household, regardless of household size and regardless of whether a household had children subject to the component's associated conditionalities. The transfer was contingent upon the household representative attending bimonthly health education workshops and bringing children under age five for scheduled preventive healthcare appointments. The workshops were held within the communities and initially covered household sanitation and hygiene, nutrition, reproductive health, and related topics. The required preventive healthcare appointments were scheduled monthly for children under age two and bimonthly for those age two to five. Health services at the scheduled visits included growth monitoring, vaccination, and provision of ferrous sulfate and anti-parasite medicine.

The program supplemented the supply of specialized healthcare services in the areas to assure that increased demand could be met without reducing quality. Specifically, the CCT contracted and trained private healthcare providers to deliver the program-related services free of charge (Regalia and Castro 2007), and beneficiaries were required to use those providers for fulfillment of the conditions. Providers visited program areas on scheduled dates and delivered services in existing health facilities, community centers, or private homes. Parallel to what was done in schools, they completed specially designed scan forms recording the services delivered that were regularly submitted to the central office to verify conditions, assess compliance, and determine transfer amounts.

In 2003, as the early treatment group was phasing out and the late treatment group phasing in, a number of additional services and corresponding conditions were added. Chief among them were vaccination for school-age children, family planning services for women of childbearing age, prenatal care consultations, and an additional set of health education workshops for adolescents. Segregated by age, adolescents were required to attend the newly introduced workshops which focused on healthy living and reproductive health, including contraception. At the same time, modern contraceptive methods were made available to beneficiaries through the healthcare providers. These additional services were designed to be implemented in both the early and the late treatment groups after 2003, but attendance was a conditionality for transfer payments only for the late treatment group as household transfers to the early treatment group ended in 2003. Thus all children 7–12 in 2000 were eligible for these sessions, and attendance for those in the late treatment group was a program condition. In practice, the services were only partially implemented in the early treatment group because there were fewer synergies with other ongoing program components. Consequently, attendance of adolescents at the health education workshops was lower in the early compared to late treatment groups.

Education component. Each eligible household received a bimonthly (every two months) cash transfer known as the school attendance transfer, contingent on enrollment and regular school attendance of children aged 7–13 years who had not yet completed the fourth grade of primary school. For each eligible child, the household also received an annual cash transfer at the start of the school year (which begins in January) intended for supplies (including uniforms and shoes) known as the school supplies transfer, which was contingent on enrollment. Unlike the school attendance transfer, a fixed amount per household regardless of the number of eligible children, the school supplies transfer was a per child transfer.

To provide incentives to the teachers and increase resources available to the schools, there was also a small cash transfer, known as the teacher transfer. In rural Nicaragua, school parents associations often requested small monthly contributions from parents to support the teacher and the school; the teacher transfer was, in part, intended to substitute for this type of fee. This transfer was per child as well, and delivered directly to the household, which was then required to pass the funds along to the teacher. The program guidelines were that the teacher keep half with the other half earmarked for the school. Although delivery of the funds by the household to the teacher was a program condition that was monitored, the ultimate use of those funds after delivery was not.[5] Teachers and school administrators completed specially designed scan forms that were regularly submitted to the program central office to verify conditions, assess compliance, and determine transfer amounts.

While there was no explicit supply-side intervention for education such as a school building program (having targeted the program to areas with adequate schooling infrastructure), the centrally administered CCT had a multisectoral approach promoting inter-institutional cooperation through specially formed committees at the national, municipal, and local levels.

[5] The small teacher transfer was continued to the end of the overall program in all areas, even in those starting in 2000 in the first budgetary cycle where the other household transfers ended after three years, in 2003.

This coordination proved useful in some areas for directing ad hoc supply-side responses to increased demand, including the placement of additional government teachers.[6]

Transfer Sizes—The initial annual transfer amounts in U.S. dollars (using the September 2000 average exchange rate of C$ 12.85 Nicaraguan Córdobas to US$ 1.00) were as follows: the nutrition and health transfer was $224 a year; the school attendance transfer $112; the school supply transfer $21; and the teacher transfer $5. On its own, the potential nutrition and health transfer represented about 13 percent of total annual household expenditures in beneficiary households before the program. A household with one child benefiting from the education component would have received additional transfers of about 8 percent, yielding an average total potential transfer of 21 percent of total annual household expenditures. On average, transfers made were 18 percent of pre-program expenditures in the first phase. The nominal transfer amounts remained constant during the first budgetary phase, with the consequence that the real value of the transfers declined by about eight percent due to local inflation. The total nominal transfer amount was reduced for households entering in 2003, the late treatment group. For that group, the nutrition and health transfer started at $168 for the first year of program participation and then declined to $145 and $126 in the second and third years. The school attendance transfer also declined slightly, to $90 per year, but the school supplies and teacher transfer amounts increased to $25 and $8 per eligible child. These represent potential transfer amounts, i.e., the transfer amount received upon complying fully with all associated conditions. Table C1 outlines the ages at baseline and 2010 follow-up and potential transfers in the two treatment groups.

Examination of administrative transfer data for all households of males and females 9–12 at the start of transfers demonstrates that take-up and average transfers received per household were lower in late treatment. (Results are nearly identical for households of 7–12-year-olds and for girls versus boys.) Even if take-up were complete (100 percent), because nominal transfers were lower in the late-treatment group we would expect average transfers to be lower. For example, based on the final schedule of transfers, a household with one child eligible for the education transfer in the late treatment would have received 88 percent of the Córdoba transfers provided to a household with identical structure and compliance in the early treatment. Average transfers received for households of 9–12-year-olds, which reflect take-up, actual household structure, eligibility and compliance over the program years were approximately $950 in early treatment and $690 in late treatment. On average, therefore, late treatment households received 73 percent of the nominal transfers received by early treatment households, lower than the full take-up potential.[7] Consistent with this pattern, as well as with the aging-out of children in the late treatment group by the time the program reached their localities, average take-up based on receipt of at least partial transfers over the three program years was 92 percent in early treatment versus 84 percent in late treatment for the food security, nutrition and health component and 87 percent versus 66 percent for the education component.

To enforce compliance with program requirements, beneficiaries did not receive the nutrition and health transfer or separately, the education transfer, in a given transfer period when they failed to carry out all of the relevant conditions described above for that component transfer. Compliance was measured via the reporting from schools and the private healthcare providers. Repeated violations, including two consecutive periods of non-compliance, led to households losing their overall program eligibility.

---

[6] In the second phase the CCT also offered limited teacher training.

[7] Adjustments for inflation further reduce (to about 65 percent) the percentage received by late treatment.

TABLE C1. POTENTIAL TRANSFER RECEIPT BY AGE AND TREATMENT GROUP

		Early treatment (2000–2003)		Late treatment (2003–2005)	
Age at baseline	Age at 2010 follow-up	Age at baseline of eligibility of transfer			
		Food security, nutrition and health transfer	Education transfer	Food security, nutrition and health transfer	Education transfer
7	17	7, 8, 9	7, 8, 9	10, 11, 12	10, 11, 12
8	18	8, 9 10	8, 9 10	11, 12, 13	11, 12, 13
9	19	9, 10, 11	9, 10, 11	12, 13, 14	12, 13
10	20	10, 11, 12	10, 11, 12	13, 14, 15	13
11	21	11, 12, 13	11, 12, 13	14, 15, 16	None[1]
12	22	12, 13, 14	12, 13	15, 16, 17	None[1]
13	23	13, 14, 15	13	16, 17, 18	None[1]

Notes: 1. The household in which the individual resided potentially received education transfers for other children.

APPENDIX D: RPS SURVEY AND LEARNING, COGNITIVE, AND SOCIO-EMOTIONAL OUTCOMES

The RPS survey had sample frame had a total of 1,330 households from the early and 1,379 households from the late treatment groups. We attempted to interview all 2,709 households, as well as all split-off households with at least one original household member under 22 years old in 2010. For the 2010 survey, we expanded the short-term evaluation household survey instrument, and added a separate, individual-level instrument. The household instrument included questions on educational attainment and current schooling, and a detailed module to measure participation and earnings in all economic activities each household member had engaged in over the last 12 months, distinguishing between activities conducted while residing in their current home community versus those performed during periods of temporary migration. For all young adults 15–22 years-old in 2010, an additional module also collected their full labor market history, to capture all off-farm activities since the young adults had entered the labor market. Given the seasonality and temporary nature of many of their economic activities, this comprehensive approach is key to accurately reflecting labor market returns for this population. All questions in the household instrument were answered by the best-informed person available for the interview. Hence, responses were obtained from the young adults themselves when they could by located at home, or if not from the household head or spouse.

The 2010 individual survey instrument was conducted through direct in-person interviews with the young adults in their homes and designed to measure individual learning and socio-emotional outcomes, as well as marriage market and reproductive health outcomes for each individual born after January 1, 1988. To measure learning, we administered three Spanish language and two math tests to all individuals 15–22 years old in 2010.

All standardized tests included in the 2010 individual survey instrument were extensively piloted and minor adjustments made for the local context as necessary, such as rephrasing questions for maximum understanding. Similar tests have been applied in other populations in Latin America, including in the evaluations of CCT programs in Ecuador and Mexico, and a different CCT program in Nicaragua (Behrman, Parker, and Todd 2009a; Fernald, Gertler, and Neufeld 2009; Paxson and Schady 2010; Macours, Schady, and Vakis 2012).

Tests were conducted in the respondents' homes by specially trained female test administrators. Therefore, the results were obtained independent of whether the respondent was in school, avoiding associated potential selection concerns.

Test administrators were selected for their background (trained as psychologists, social workers, or in other similar fields) and for their ability to quickly establish a strong rapport with children and young adults. They were trained to motivate the respondents to participate in the tests, keeping final non-response to a minimum. Tests were administered inside the home (or compound) and the privacy of the test-taker and the confidentiality of the results were assured throughout the process. During the test administrators' training, emphasis was placed both on gaining the confidence of the respondents before starting and on the standardized application of each of the tests. The quality and standardized application of the tests was monitored closely in the field, and given the long survey period, several re-standardization trainings carried out.

Data collection and test administration was organized to ensure individual test administrators maintained an approximate balance between the number of children from early and late treatment localities they tested. Fieldwork visits to early and late treatment localities also were balanced over time to avoid possible seasonal differences in measurement between the two experimental groups. Consistent with these field protocols, results are robust to controls for the identity of the test administrator (not shown).

All tests were scored using standard practice. For the digit span, we used the combined score of the forward and backward components of the test. When estimating the differential ITT on each component separately, estimates of the forward digit span (measuring math memory) were in line with the TVIP and significant, while the backward digit span (often considered a measure of executive functioning) was not significant and similar to the results for the Raven.

Exploratory Factor Analysis for Socio-Emotional Outcomes—We applied two standardized instruments to measure socio-emotional outcomes. The first was the Strength and Difficulties test (SDQ), a self-reported behavioral screening test consisting of 25 questions aimed at measuring a set of positive and negative behaviors. The second was the Center for Epidemiologic Studies Depression Scale or CESD Scale (Radloff 1977), a commonly used mental health scale, developed as a screening test for depression and depressive disorder, consisting of 22 questions asking for the frequency of both positive and negative self-perceptions. Both tests were available in Spanish.

Males—We examined the internal consistency of the different scales for the sample of males 9–12 years old at baseline using exploratory factor analysis. The overall Cronbach's alpha of the 25 items of the SDQ together (0.70) indicated that the scale as a whole was internally consistent. The alphas were much lower, however, when considering the standard five subdomains (emotional symptoms, conduct problems, hyperactivity, peer relationships, and pro-social behavior), and varied from 0.26 to 0.51, hence much lower than the usual thresholds for statistical validity. Exploratory factor analysis on the 25 items suggested there were only two meaningful factors that could be retained (i.e., two factors with eigenvalues above one and for which the scree plot led to a similar conclusion). Moreover, when imposing the five-factor structure, the items did not group along the standard five subscales, with the first factor having high factor loads on items from three of the five subscales.[8] For the CESD, the Cronbach's alpha for internal consistency of the 20 items also was high (0.83), but the factor analysis only pointed to one or two factors, and did not allow further differentiation.

Females—As described for the males above, we examined the internal consistency of the different scales for the sample of females 9–12 years old at baseline using exploratory factor analysis. The overall Cronbach alpha of the 25 items of the SDQ together (0.72) indicated that the scale as a whole is internally consistent. The alphas were much lower, however, when considering the standard five subdomains, and varied from 0.21 to 0.54, hence much lower than the usual thresholds for statistical validity. Exploratory factor analysis on the 25 items suggested there are only two meaningful factors that could be retained (i.e., two factors with eigenvalues above one and for which the scree plot led to a similar conclusion). Moreover, when imposing the five-factor structure, the items do not group along the standard five subscales, with the first factor having high factor loads on items from three of the five subscales. When we consider the CESD, the Cronbach alpha for internal consistency of the 20 items is high (0.86) but the factor analysis only pointed to one or two factors, and did not allow further differentiation.

[8] Laajaj and Macours (2020) report similar findings for other measures of socio-emotional outcomes when scales originally designed for developed country settings were used in developing country settings.

Because the results from the exploratory factor analyses suggested that responses should not be pooled together based on the standard subcategories of the SDQ for males or for females, we constructed new indices based on both tests to capture the relevant latent traits. Specifically, we pooled together all questions from both the SDQ and the CESD scales and identified the latent socio-emotional traits in our sample, following among others, Cunha, Heckman, and Schennach (2010) and Attanasio et al. (2015). Based on both the eigenvalues and the scree plots, and using an oblique quartimin rotation to allow the different factors to be correlated with one another, we retained four factors for each gender, two with high factor loads on items from the SDQ scale, and two with high loads on items from the CESD scale. Notably, questions referring to positive, and respectively negative, attitudes or behavior loaded onto similar factors in each of the scales. Therefore, considering the factor loadings of the different items together points to a plausible interpretation of these factors as capturing stress, positive self-evaluation, negative self-evaluation, and optimism.

APPENDIX E: TRACKING PROTOCOLS AND ATTRITION CORRECTIONS

In the 2010 survey, we placed special emphasis on tracking all temporary and permanent migrants and otherwise difficult to interview individuals. In the first phase of the survey, lasting about six months, we interviewed individuals and households residing in or near their original localities. We refer to this period as the "regular" tracking phase. This was followed by an "intensive" tracking phase, lasting approximately 1.5 years, during which we made exhaustive efforts to find all individuals not found during regular tracking, through repeat visits to original locations and tracking to any known location in Nicaragua or Costa Rica. Repeat visits to original localities were important for interviewing temporary migrants who had returned in a context where temporary migration is common. For individuals who could not be located, however, some information on selected individual variables is available, collected through proxy reports when interviewing the original household.

We therefore distinguish between three sets of outcomes based on their source—from the household-level survey instrument, the individual-level instrument, or by proxy from the household-level instrument. Outcomes on individuals collected in the household instrument include educational attainment, labor market and earnings outcomes, and marital status, all self-reported by the individual or in cases when she was resident but temporarily absent by another informed household member. Outcomes in the individual-level instrument include achievement and cognitive tests, socio-emotional outcomes, fertility, reproductive behaviors and for females BMI and age at menarche. Outcomes for which proxy information was collected in the original household include highest grade attained, marital status, migration and labor market status.

Attrition is highest for the outcomes that required direct in-person interactions with the respondents during the individual-level instrument. For the main sample of 9–12 cohort of males examined, 19 percent could not be tracked for the individual instrument, 10 percent for the household instrument, and for 4 percent we are also missing a proxy report.[9] For females, 22 percent could not be tracked for the individual instrument, 16 percent for the household instrument, and for 6 percent we are also missing a proxy report.[10] In this appendix we provide details for the 9–12 year-old cohorts; the same methodology was used for 7–8 year-olds.

Given the relatively small sample sizes, we intensively tracked all migrants in the long-term 2010 follow-up. Tracking rates in the intensive phase are comparable to intensive tracking rates obtained in other studies for random subsamples, resulting in a final tracking rate of 90 percent for males (84 percent for females) for the household instrument, comparable to other long-term follow-ups of RCTs with both regular and intensive tracking phases.[11] Attrition rates in our study

---

[9] Fifteen males in the 9–12 age cohort had died by 2010. These individuals are not used to predict the probability of attrition as selection for them is most likely driven by other factors. Including deceased individuals, final attrition rates for males are 10 percent for the household and 19 percent for the individual instruments.

[10] Seven females in the 9–12 age cohort had died by 2010. As done for the males, these individuals are not used to predict the probability of attrition. There are also 18 observations for whom the reproductive health module in the individual survey is missing, leading to lower numbers of observations for the fertility family. Including deceased individuals, final attrition rates for females are 16 percent for household and 22 percent for individual instruments.

[11] They are, for instance, comparable to Blattman, Fiala, and Martinez (2014) who report an 82 percent effective tracking rate for a 4-year follow-up survey of young adults in Uganda; the 10-year follow-up of the Kenya Longitudinal Panel Survey with an effective tracking rate of 84 percent (Baird et al., 2016); and the 88 percent effective tracking rate for children after 5-7 years in the Moving to Opportunity evaluation in the United States (Orr et al., 2003). Our tracking efforts, however, were less successful than the eight-year follow-up of a scholarship

are similar to or lower than related studies analyzing the long-term experimental impacts of CCTs.[12]

Table E1 shows that attrition rates are well balanced between the early and late treatment groups for both genders with the coefficients on the ITT indicator on the probability of having been found, i.e., interviewed, smaller than |1.5| percentage points for both the household and individual survey instruments and p-values > 0.6.[13] Table A1 confirms that this resulted in a sample that is balanced on baseline characteristics, with at most five out of the 52 baseline variables (from 2000 program census) significantly different at 10 percent level for each gender.

To explore balance further, in Table E2 using the males to illustrate we confirm that in contrast to those found during regular tracking, intense tracking resulted in final samples that maintained balance on baseline characteristics, finding four (respectively six) out of the 42 baseline variables examined significantly different at the 10 percent level for the final samples with the individual (columns 5 and 6) and household (columns 9 and 10) instruments. These are the same differences observed for the full sample (columns 1 and 2), and hence do not appear to be driven by selective attrition.

To further assess the possibility of selective attrition, for each baseline variable we estimate $Prob(found) = F(X, T, XT)$. As seen in Table E3, none of the estimated coefficients on T are significant and only a few on XT are, confirming that overall, both the rates of attrition and the observable characteristics of individuals who attrited are similar between the two treatment groups. At the same time, the coefficients on X make clear that those who attrited are different than those found along a number of dimensions. This implies that, even if internal validity is not jeopardized, the differential ITT estimates may not be representative of the effects for the full target population, i.e., external validity is a potential concern. This is particularly relevant if treatment heterogeneity is correlated with attrition selection. Because marriage and labor market opportunities are the two main reasons for migration (and consequently for attrition) of young men and women in this context, and also principal outcome variables in this paper, this is a potential concern for analyses.

In our preferred specifications, therefore, we specifically account for such selection using inverse probability weights (IPW) constructed as described below and estimated separately for each gender. We also present a number of alternative estimations to examine the sensitivity of the findings to different assumptions about attrition. First, because attrition rates and baseline characteristics of the interviewed samples are balanced across treatment groups, we re-estimate the main results without any attrition corrections and with an alternative IPW approach (appendix B, Tables B1 panels E and F). Results from that exercise demonstrate that the main findings are not driven by the attrition correction and point in the same direction for all the

---

program in Ghana (Duflo, Dupas, and Kremer, 2017), who like us use intensive tracking for the entire sample of those not found during regular tracking, but where special protocols to track respondents were incorporated into the RCT design from baseline including maintaining regular contact with respondents throughout the duration of the study, resulting in a 98 percent tracking rate.

[12] Behrman, Parker, and Todd (2011) and Adhvaryu et al. (2018) use the 6-year follow up of Mexico's PROGRESA evaluation sample, with attrition of 40 percent for individual-level information on comparable age groups. The 10-year PROGRESA follow-up (used for instance in Kugler and Rojas, 2018) has more than 60 percent attrition. Attrition rates in Baird, McIntosh, and Özler (2019) are 13–16 percent after five years for young women, and the 10-year follow-up of a much younger cohort in Ecuador has 19 percent attrition (Araujo, Bosch, and Schady, 2018).

[13] Attrition rates from the oversample households are not significantly different than for the rest of the sample and are also equally balanced (p-values of the differences 0.5 or higher).

families of outcomes. Second, we estimate Lee bounds to test the sensitivity of results for corrections of the small, observed differences in attrition rates (Table B1 panel H). These rely on the monotonicity assumption, implying that assignment to the treatment group can affect attrition in only one direction. Because attrition may be correlated with marriage, labor market, and migration outcomes, and each of those could be affected by early and late treatment exposure in various ways, the monotonicity assumption may be violated in this context where we are estimating differential treatment effects (but we nevertheless present them for completeness). Third, for selected available outcomes we present results using proxy measures, which leads to much lower rates of missing information, but also likely introduces measurement error in the outcome variables (Table A8-A and -B).

Attrition selection correction with inverse probability weights[14]

Because several baseline characteristics are correlated with the probability of having been found, at least some of the potential attrition selection is likely related to observables. For example, Table E2 shows for males that prior to the intensive tracking phase, there were more baseline characteristics that were not balanced (columns 3 and 7) than would be expected by chance, but that baseline balance was restored through the intensive tracking (columns 5 and 9).[15] This holds for both household- and individual-level data. Intensive tracking, therefore, proved important for internal validity. Those found during the intensive tracking phase differ from those found during the regular tracking phase in terms of observed baseline characteristics. Moreover, individuals that were difficult to find, but were ultimately found, are more similar on baseline observable characteristics to those that were ultimately not found. Among other differences, administrative data indicate that those that were more difficult to find had lower compliance rates with the CCT, suggesting that attrition correction may be important to account for heterogeneity in the ITT effects. Patterns are qualitatively similar for females.

To account for potential selection due to attrition, we calculate attrition-corrected weights. In our preferred specifications, we use a modified version of the standard inverse probability weighting adjustment that more fully exploits information obtained during the intensive tracking phase and allows putting higher weight on individuals who were more difficult to find. The key assumption underlying this strategy is that the probability of being found during the intensive tracking phase is explained by observable characteristics. Overweighting individuals whose observed characteristics predict they were more difficult to find corrects for the sample selection. We assign a weight of 1 to individuals found during the regular tracking phase, and only estimate the weights for those found during the intensive tracking phase. To determine the weights, we estimate the probability of being found for those found during the intensive tracking phase.

We calculate attrition-correction weights separately for each survey instrument and gender. A large set of socio-economic variables observed in or calculated from the program census was considered for predicting attrition, informed in particular by the nature of migration from the regions. These include all baseline variables shown in Tables E3-A and E3-B, capturing individual-, household-, and locality-level characteristics. As connectedness to the locality could be a good predictor of tracking success, we included two variables to capture the social network of the individual (village size and family network size), some more detailed household structure

[14] This section draws from Molina Millán and Macours (2017), which contains a more detailed explanation of the approach taken and further rationale for the selection correction.

[15] The few variables that are not balanced in the final 2010 sample (column 5 and 9) are similarly not balanced on the full baseline target sample (column 1).

variables, and a set of proxy variables meant to capture the possible temporary nature of residency in the village at baseline for some households.[16] We similarly consider locality-level characteristics that could be push or pull factors for migration: remoteness (measured using distance to night light and altitude), location in a coffee producing area, and having been affected by hurricane Mitch, a severe storm in the area in 1998. Finally, as further proxy measures for locations with a concentration of more temporary residents, we introduce two variables capturing the level of attrition between the program census and the first baseline survey (i.e., between May and August of 2000): 1) the share of individuals in the locality that attrited; and 2) whether any individual in the target age group attrited.[17] Individuals from such localities not only were more likely to attrit, but also could be more difficult to trace, as contacts with the community of origin could be limited.

Because of the large number of potential variables to consider and because there are relatively few individuals not found after intensive tracking, we follow the approach of Doyle et al. (2017) to select a reduced set of predictors. Separate estimates are carried out to model attrition for the household and individual instruments. First, we estimate bivariate regressions in which each potential predictor is examined to determine whether a significant difference exists between the means for those found and not found during intensive tracking. All estimates use the survey sample weights and standard errors are clustered at the locality level. This testing is conducted separately for the early and late treatment groups. The correlates of having been found during intensive tracking differ between treatment groups and also between the survey instruments.

We retain as potential predictors all indicators found to be statistically significantly different for the early or late treatment group. We then carry out a first estimate of the probability of having been found during intensive tracking on this set of baseline predictor variables for each treatment group, using the sample of those not found during regular tracking. To account for collinearity between measures, the baseline predictor set is restricted further by conducting stepwise selection of variables with backward elimination and using the adjusted R^2 as the information criterion. Strata and regional fixed effects, as well as 6-month age dummies are included as fixed predictors in all models.

In the final step, we estimate the probability of having been found during the intensive tracking phase for both early and late treatment group together, keeping only the predictors as indicated by the stepwise procedure, as well as the strata and regional fixed effects and the 6-month age dummies, all interacted with the treatment variable. Tables E4-A and E4-B present the linear probability model estimates for each survey. The resulting regressions have good predictive power for both males and females.

The probabilities of having been found during intensive tracking (conditional on not having been found during regular tracking) are then estimated via probits for the specifications shown in Tables E4-A and E4-B, and then used to determine the weights for the attrition selection. All observations found during regular tracking are assigned a weight of 1, while those found during the intensive tracking are assigned a weight 1/Prob (found during intensive tracking | not having been found during regular tracking). Finally, these weights are then multiplied with the sample weights. Final attrition-correction weights for males vary between 1 and 35 for the household

[16] In particular, we include a set of indicators as proxies for whether the household comprised temporary workers on one of the large coffee plantations known as *haciendas*. These households were captured by the program census but not likely to have been permanent residents of the areas.

[17] The baseline survey was conducted shortly after the public lottery and before the start of the transfers.

instrument, and 1 and 86 for the individual instrument, with corresponding averages of 3.4 and 3.9. For females the final attrition-correction weights vary between 1 and 32 for the household instrument, and 1 and 89 for the individual instrument, both averaging about 4.[18]

We overweight individuals interviewed during the intensive tracking phase, as these individuals were, by definition, more difficult to find, and therefore more likely to be similar to those not found at all. Empirically, observable characteristics are also better predictors for the subsample that was intensively tracked than for the full sample, indicating that selection on observables for this subsample is a more plausible assumption than for the full sample. Nevertheless, we also present results with the more common inverse probability weighting (with weights estimated using the probability of being found in the entire sample) in Table B1 panel H, for comparison. We followed a similar process for covariate selection in the estimation of weights for that version of the IPW.

Alternative treatments of attrition relying on different assumptions are presented in appendix B and include no correction, Lee bounds, and proxy measures.

[18] With the exception of a few outliers, the distribution of weights is not highly skewed with 97 percent of household and 95 percent of individual weights less than two for males and less than nine for females. Only two observations for each gender have individual weights higher than 32 and omitting these two observations from the analysis does not alter any of the findings. With the exception of those two outliers, the distribution of weights is similar when using conventional IPW estimates.

TABLE E1. ATTRITION AND TRACKING IN THE EARLY AND LATE TREATMENT GROUPS, MALES AND FEMALES 9–12 YEAR-OLD COHORT

Panel A	Probability of having been interviewed: Household instrument			
	(1)	(2)	(3)	(4)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)	Found (incorporating proxy information)
<u>Males</u>				
ITT	-0.014	-0.023	-0.020	-0.025
	(0.027)	(0.037)	(0.075)	(0.021)
Mean late treatment	0.905	0.751	0.62	0.969
N	1138	1138	297	1138
Females				
ITT	-0.010	-0.005	-0.019	-0.021
	(0.038)	(0.038)	(0.073)	(0.022)
Mean late treatment	0.849	0.584	0.636	0.958
N	1062	1062	444	1062
	1062	1062	444	1062

Panel B	Probability of having been interviewed: Individual instrument		
	(1)	(2)	(3)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)
<u>Males</u>			
ITT	0.008	0.017	0.004
	(0.037)	(0.044)	(0.071)
Mean late treatment	0.793	0.455	0.62
N	1138	1138	611
<u>Females</u>			
ITT	-0.014	0.011	-0.034
	(0.041)	(0.039)	(0.067)
Mean late treatment	0.786	0.444	0.615
N	1062	1062	585

Notes: Outcome data source: 2010 long-term evaluation survey. Males or females ages 9–12 at the start of the program in November 2000. Standard errors clustered at the locality level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE E2. COMPARISON OF BASELINE BALANCE BY TRACKING STATUS AND 2010 SURVEY INSTRUMENT FOR MALES, 9–12 YEAR-OLD COHORT

	Baseline Sample: Cohort 9–12 Males		Household Instrument Outcomes				Individual Instrument Outcomes			
			Found During Regular Tracking		Final Sample		Found During Regular Tracking		Final Sample	
	N=1138		N=826		N=1006		N=527		N=907	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Individual Characteristics										
Age at start of transfer in months	-0.046*	(0.026)	-0.086*	(0.047)	-0.068**	(0.032)	-0.129*	(0.068)	-0.052	(0.041)
No grades attained (=1)	-0.008	(0.064)	-0.025	(0.065)	-0.014	(0.063)	-0.018	(0.073)	-0.016	(0.065)
Highest grade attained	0.071	(0.16)	0.136	(0.16)	0.094	(0.15)	0.173	(0.18)	0.111	(0.16)
Worked in last week (=1)	-0.054*	(0.027)	-0.064**	(0.029)	-0.064**	(0.030)	-0.039	(0.038)	-0.060**	(0.029)
Participated in some economic activity (=1)	-0.009	(0.035)	-0.016	(0.037)	-0.014	(0.035)	0.009	(0.042)	-0.008	(0.037)
Household Characteristics: Education										
Distance to nearest school (minutes)	0.744	(4.76)	2.547	(4.56)	0.594	(4.54)	4.308	(5.91)	1.153	(4.31)
Household head no grades attained (=1)	0.005	(0.031)	0.005	(0.040)	0.020	(0.035)	0.024	(0.034)	0.036	(0.036)
Household head 3 plus grades attained (=1)	0.025	(0.029)	0.046	(0.031)	0.032	(0.027)	0.060*	(0.033)	0.033	(0.026)
Mother no grades attained (=1)	-0.046	(0.037)	-0.097**	(0.042)	-0.053	(0.043)	-0.100*	(0.053)	-0.040	(0.045)
Mother 3 plus grades attained (=1)	0.072	(0.044)	0.111***	(0.038)	0.068	(0.041)	0.171***	(0.052)	0.057	(0.042)
Household Characteristics: Demographics										
Father not living in same household (=1)	0.017	(0.031)	0.041	(0.029)	0.013	(0.027)	0.092**	(0.040)	0.011	(0.025)
Mother not living in same household (=1)	0.011	(0.017)	0.027	(0.017)	0.013	(0.017)	0.022	(0.018)	0.015	(0.019)
Child of household head (=1)	-0.016	(0.024)	-0.061**	(0.024)	-0.020	(0.023)	-0.074**	(0.031)	-0.023	(0.024)
Number of children of household head	-0.241	(0.22)	-0.511**	(0.23)	-0.326	(0.23)	-0.487*	(0.29)	-0.311	(0.23)
Female household head (=1)	0.030	(0.020)	0.035*	(0.018)	0.018	(0.018)	0.053*	(0.026)	0.019	(0.018)
Age of household head	0.339	(0.87)	0.556	(0.87)	0.473	(0.79)	0.290	(1.11)	0.843	(0.85)
Number of household members	-0.047	(0.19)	-0.205	(0.24)	-0.159	(0.22)	0.028	(0.27)	-0.089	(0.22)
Nuclear household (=1)	-0.017	(0.040)	-0.023	(0.044)	-0.019	(0.043)	-0.071	(0.052)	-0.038	(0.044)
Multigenerational household (=1)	-0.037	(0.035)	-0.005	(0.034)	-0.020	(0.037)	0.023	(0.039)	-0.009	(0.038)

Other household structure (=1)	0.054*	(0.028)	0.028	(0.031)	0.038	(0.028)	0.047	(0.038)	0.047	(0.032)
Number of children aged 0-8	0.021	(0.11)	-0.067	(0.12)	-0.097	(0.11)	0.095	(0.14)	-0.087	(0.12)
Number of children age 9 to 12	-0.036	(0.058)	-0.078	(0.070)	-0.072	(0.062)	-0.087	(0.076)	-0.066	(0.068)
Household Characteristics: Economic Activities & Assets										
Household head main occupation is agric. (=1)	-0.021	(0.038)	-0.003	(0.046)	-0.018	(0.041)	0.004	(0.049)	-0.018	(0.042)
Size of landholdings (1000 square meters)	-1.987	(1.85)	-2.822	(2.02)	-2.500	(1.80)	-3.730	(2.28)	-1.656	(1.82)
Log of size of landholdings	-0.270	(0.41)	-0.049	(0.43)	-0.087	(0.42)	0.144	(0.60)	0.013	(0.44)
Number of parcels of land	-0.036	(0.073)	-0.018	(0.076)	-0.008	(0.074)	0.010	(0.094)	0.017	(0.073)
Log predicted expenditures (per capita)	-0.009	(0.022)	0.029	(0.026)	0.003	(0.025)	0.009	(0.034)	0.001	(0.025)
Wealth index - housing characteristics	0.200	(0.16)	0.328**	(0.16)	0.223	(0.15)	0.226	(0.17)	0.159	(0.15)
Wealth index - productive assets	-0.258**	(0.11)	-0.209*	(0.12)	-0.263**	(0.11)	-0.188	(0.15)	-0.201*	(0.11)
Wealth index - other assets	-0.040	(0.18)	-0.127	(0.19)	-0.059	(0.18)	-0.123	(0.21)	-0.067	(0.18)
Village Characteristics										
Village affected by hurricane Mitch (=1)	-0.062	(0.050)	-0.062	(0.051)	-0.059	(0.053)	-0.016	(0.057)	-0.039	(0.049)
Altitude of village	-21.10	(34.6)	-25.69	(35.9)	-22.51	(35.9)	-26.79	(37.0)	-24.49	(36.9)
Village in coffee producing area (=1)	-0.018	(0.065)	-0.021	(0.066)	-0.010	(0.066)	-0.008	(0.065)	-0.008	(0.066)
Distance to night light (meters)	2276	(2642)	1291	(2619)	1880.0	(2607)	2730	(2758)	1980	(2645)
Live in Tuma region (=1)	0.193	(0.15)	0.155	(0.14)	0.194	(0.14)	0.162	(0.15)	0.187	(0.14)
Live in Madriz region (=1)	0.017	(0.12)	0.044	(0.12)	0.027	(0.13)	0.021	(0.12)	0.017	(0.13)
Social Capital										
Family network size (individuals)	23.52**	(10.3)	27.65**	(12.1)	25.94**	(10.9)	19.20	(14.1)	26.17**	(11.7)
Population size village	257.6***	(88.6)	249.0**	(94.6)	257.2***	(88.7)	250.1**	(101)	242.1***	(86.9)
Proxy's of Permanent Residence in Village										
Own house (=1)	-0.052	(0.043)	-0.021	(0.046)	-0.064	(0.043)	-0.008	(0.041)	-0.048	(0.041)
House obtained for service/labor (=1)	0.019	(0.039)	-0.016	(0.032)	0.023	(0.036)	-0.014	(0.033)	0.011	(0.037)
Address in hacienda (=1)	-0.022	(0.058)	-0.023	(0.058)	-0.004	(0.056)	-0.028	(0.059)	-0.007	(0.054)
Address in hacienda & house rented (=1)	0.032	(0.033)	0.018	(0.031)	0.047	(0.031)	0.017	(0.022)	0.038	(0.028)

Notes: Outcome data source: 2010 long-term evaluation survey. Regressions include males ages 9–12 at the start of the program in November 2000, weighted to account for sampling. Controls include strata fixed effects. Distance to night light (meters) is linear distance from household to an area with stable night light detected by a satellite (DMSP-OLS Nighttime Lights). House received in exchange for services is an indicator variable for households who received the house in exchange for labor services. Address in hacienda is an indicator for households whose address refers to a location on a large plantation (hacienda). Standard errors clustered at the locality level are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE E3-A. RELATIONSHIP BETWEEN THE PROBABILITY OF BEING FOUND, BASELINE COVARIATES AND TREATMENT, MALES 9–12 YEAR-OLD COHORT

	X		*X*T*		*T*	
Individual Characteristics						
Age at start of transfer in months	-0.024**	(0.011)	-0.017	(0.018)	0.170	(0.20)
No grades attained (=1)	-0.012	(0.029)	-0.056	(0.060)	0.010	(0.037)
Highest grade attained	-0.010	(0.012)	0.019	(0.026)	-0.038	(0.044)
Worked in last week (=1)	0.005	(0.035)	-0.051	(0.057)	-0.007	(0.030)
Participated in some economic activity (=1)	-0.007	(0.024)	-0.018	(0.047)	-0.011	(0.032)
Household Characteristics: Education						
Distance to nearest school (minutes)	0.000	(0.000)	0.000	(0.000)	-0.018	(0.033)
Household head no grades attained (=1)	-0.039	(0.031)	0.041	(0.054)	-0.037	(0.040)
Household head 3 plus grades attained (=1)	-0.006	(0.035)	0.036	(0.051)	-0.026	(0.033)
Mother no grades attained (=1)	0.100**	(0.040)	-0.031	(0.064)	0.002	(0.052)
Mother 3 plus grades attained (=1)	-0.008	(0.027)	-0.016	(0.069)	-0.009	(0.035)
Household Characteristics: Demographics						
Father not living in same household (=1)	-0.099*	(0.050)	-0.017	(0.064)	-0.011	(0.027)
Mother not living in same household (=1)	-0.12	(0.088)	0.059	(0.13)	-0.020	(0.031)
Child of household head (=1)	0.057	(0.057)	-0.012	(0.088)	-0.004	(0.083)
Number of children of household head	0.019***	(0.005)	-0.014	(0.011)	0.055	(0.068)
Female household head (=1)	-0.046	(0.051)	-0.065	(0.071)	-0.007	(0.027)
Age of household head	0.002	(0.002)	0.000	(0.003)	-0.027	(0.130)
Number of household members	0.011	(0.007)	-0.016	(0.011)	0.120	(0.099)
Nuclear household (=1)	0.038	(0.031)	-0.010	(0.051)	-0.009	(0.043)
Multigenerational household (=1)	-0.029	(0.033)	0.106*	(0.053)	-0.042	(0.038)
Other household structure (=1)	-0.035	(0.063)	-0.123	(0.095)	0.006	(0.026)
Number of children aged 0-8	0.018	(0.012)	-0.058**	(0.022)	0.108*	(0.055)
Number of children age 9 to 12	0.023	(0.015)	-0.071*	(0.039)	0.111	(0.068)
Household Characteristics: Economic Activities & Assets						
Household head main occupation is ag. (=1)	0.063	(0.047)	-0.008	(0.074)	-0.008	(0.075)
Size of landholdings ('000 sq meters)	0.000	(0.000)	-0.001	(0.002)	0.006	(0.044)
Log of size of landholdings	0.004	(0.005)	0.011	(0.010)	-0.100	(0.091)
Number of parcels of land	0.030	(0.020)	0.063	(0.043)	-0.072	(0.059)
Log predicted expenditures (pc)	-0.121*	(0.066)	0.134	(0.089)	-1.055	(0.680)
Wealth index - housing characteristics	-0.024*	(0.014)	0.017	(0.020)	-0.014	(0.029)
Wealth index - productive assets	0.034**	(0.015)	0.001	(0.023)	-0.007	(0.028)
Wealth index - other assets	-0.011	(0.014)	-0.027	(0.019)	-0.016	(0.026)
Village Characteristics						
Village affected by hurricane Mitch (=1)	-0.096***	(0.024)	0.0935*	(0.048)	-0.101**	(0.038)

Altitude of village (‘000 meters)	0.027	(0.14)	-0.098	(0.20)	0.046	(0.13)
Village in coffee producing area (=1)	-0.004	(0.036)	0.003	(0.060)	-0.018	(0.058)
Distance to night light (‘000 meters)	-0.003	(0.002)	0.000	(0.002)	0.062*	(0.036)
Live in Tuma region (=1)	-0.123***	(0.032)	-0.013	(0.042)	0.017	(0.018)
Live in Madriz region (=1)	0.0375	(0.028)	0.091**	(0.037)	-0.033	(0.032)
Social Capital						
Family network size (individuals) ‘000	0.678***	(0.22)	0.094	(0.27)	-0.040	(0.042)
Population size village ‘000	-0.013	(0.027)	0.003	(0.042)	-0.014	(0.042)
Proxy's of Permanent Residence in Village						
Own house (=1)	0.146**	(0.064)	-0.080	(0.091)	0.057	(0.082)
House is obtained in exchange for service/labor (=1)	-0.154**	(0.071)	0.034	(0.13)	-0.015	(0.029)
Address in hacienda (=1)	-0.111	(0.072)	0.076	(0.099)	-0.028	(0.030)
Address in hacienda & house rented (=1)	-0.289**	(0.14)	0.231	(0.16)	-0.026	(0.029)

Notes: Outcome data source: 2010 long-term evaluation survey. Regressions include males ages 9–12 at the start of the program in November 2000, weighted to account for sampling. For each baseline variable X, the table presents coefficient estimates of the linear probability model: $Prob(found) = F(X, T, XT)$. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE E3-B. RELATIONSHIP BETWEEN THE PROBABILITY OF BEING FOUND, BASELINE COVARIATES, AND TREATMENT, FEMALES 9–12 YEAR-OLD COHORT

	X		*X*T*		*T*	
Individual Characteristics						
Age at start of transfer in months	-0.0516**	(0.021)	0.0394	(0.027)	-0.437	(0.30)
No grades attained (=1)	-0.0459	(0.054)	0.0545	(0.070)	-0.0268	(0.045)
Highest grade attained	-0.0050	(0.023)	0.0074	(0.028)	-0.0152	(0.056)
Worked in last week (=1)	-0.0171	(0.10)	-0.0987	(0.22)	-0.0024	(0.038)
Participated in some economic activity (=1)	-0.0065	(0.069)	0.0093	(0.11)	-0.0061	(0.038)
Household Characteristics: Education						
Distance to nearest school (minutes)	-0.0013*	(0.001)	0.0015*	(0.001)	-0.0423	(0.044)
Household head no grades attained (=1)	0.114***	(0.031)	-0.0869	(0.055)	0.0331	(0.053)
Household head 3 plus grades attained (=1)	-0.0166	(0.047)	-0.0204	(0.074)	-0.0001	(0.047)
Mother no grades attained (=1)	0.0610	(0.050)	0.0566	(0.067)	-0.0287	(0.058)
Mother 3 plus grades attained (=1)	0.0006	(0.061)	-0.102	(0.093)	0.0312	(0.048)
Household Characteristics: Demographics						
Father not living in same household (=1)	-0.0062	(0.034)	-0.100	(0.066)	0.0222	(0.036)
Mother not living in same household (=1)	-0.103	(0.064)	-0.0156	(0.11)	-0.0031	(0.038)
Child of household head (=1)	-0.0040	(0.034)	0.122	(0.096)	-0.108	(0.10)
Number of children of household head	0.0020	(0.007)	0.0134	(0.0097)	-0.0711	(0.071)
Female household head (=1)	-0.0086	(0.050)	-0.0441	(0.074)	0.002	(0.043)
Age of household head	0.0029**	(0.001)	-0.0007	(0.003)	0.0228	(0.13)
Number of household members	0.0055	(0.0055)	-0.0047	(0.009)	0.0337	(0.084)
Nuclear household (=1)	-0.0389	(0.032)	0.0607	(0.066)	-0.0415	(0.054)
Multigenerational household (=1)	0.0455	(0.040)	0.0053	(0.083)	-0.0059	(0.055)
Other household structure (=1)	0.0011	(0.053)	-0.130	(0.10)	0.0123	(0.038)
Number of children aged 0-8	-0.0076	(0.012)	-0.00673	(0.020)	0.0087	(0.057)
Number of children age 9 to 12	-0.0367	(0.034)	0.00138	(0.061)	-0.0082	(0.10)
Household Characteristics: Economic Activities & Assets						
Household head main occupation is agriculture (=1)	0.0068	(0.043)	0.0004	(0.061)	-0.0054	(0.072)
Size of landholdings ('000 sq meters)	-0.0004	(0.0013)	0.00145	(0.0016)	-0.0316	(0.058)
Log of size of landholdings	0.0098	(0.0073)	0.0179	(0.011)	-0.149	(0.11)
Number of parcels of land	0.0472	(0.035)	0.124*	(0.065)	-0.120	(0.087)
Log predicted expenditures (pc)	0.0591	(0.057)	-0.0643	(0.073)	0.493	(0.57)
Wealth index - housing characteristics	0.0039	(0.017)	-0.0255	(0.027)	-0.005	(0.040)
Wealth index - productive assets	0.0160	(0.012)	0.0161	(0.029)	-0.004	(0.039)
Wealth index - other assets	0.0302*	(0.018)	-0.0246	(0.025)	-0.0034	(0.039)
Village Characteristics						
Village affected by hurricane Mitch (=1)	0.103	(0.091)	-0.0915	(0.12)	0.0792	(0.10)
Altitude of village ('000 meters)	0.265***	(0.082)	-0.229	(0.18)	0.140	(0.12)
Village in coffee producing area (=1)	-0.0567*	(0.032)	-0.0542	(0.050)	0.0368	(0.045)
Distance to night light ('000 meters)	-0.0058*	(0.0029)	-0.0020	(0.003)	0.0491	(0.058)

Live in Tuma region (=1)	-0.134***	(0.041)	-0.0301	(0.063)	0.0374	(0.029)
Live in Madriz region (=1)	0.160***	(0.027)	-0.0213	(0.046)	-0.0004	(0.043)
Social Capital						
Family network size (individuals) ‘000	0.682***	(0.21)	0.00606	(0.30)	-0.0130	(0.059)
Population size village ‘000	0.0414	(0.072)	-0.0336	(0.091)	0.0045	(0.061)
Proxy's of Permanent Residence in Village						
Own house (=1)	0.175**	(0.074)	-0.0368	(0.089)	0.0294	(0.093)
House is obtained in exchange for service/labor (=1)	-0.287***	(0.078)	0.0970	(0.12)	-0.0130	(0.036)
Address in hacienda (=1)	-0.185***	(0.059)	0.130	(0.091)	-0.0282	(0.042)
Address in hacienda & house rented (=1)	-0.334***	(0.087)	0.247**	(0.12)	-0.0233	(0.040)

Notes: Outcome data source: 2010 long-term evaluation survey. Regressions include females ages 9–12 at the start of the program in November 2000, weighted to account for sampling. For each baseline variable X, the table presents coefficient estimates of the linear probability model: $Prob(found) = F(X, T, XT)$. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE E4-A. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF BEING FOUND DURING INTENSIVE TRACKING PHASE, MALES

	Household Instrument		Individual Instrument	
	Coef. (SE)	Coef. ET Interaction (SE)	Coef (SE)	Coef. ET Interaction (SE)
Early treatment (ET)=1		-0.581		-0.347
		(0.38)		(0.30)
No grades attained (=1)			-0.097	-0.014
			(0.073)	(0.12)
Distance to nearest school (minutes)			0.000	-0.001
			(0.001)	(0.001)
Mother no grades attained (=1)	0.159*	0.022		
	(0.094)	(0.13)		
Mother 3 plus grades attained (=1)			-0.032	-0.040
			(0.061)	(0.078)
Household head no grades attained (=1)			-0.103*	0.140*
			(0.054)	(0.082)
Father not living in same household (=1)	-0.144	0.134		
	(0.17)	(0.19)		
Child of household head (=1)	0.0501	0.209		
	(0.18)	(0.22)		
Female household head (=1)	0.374***	-0.506***	0.039	-0.196*
	(0.12)	(0.18)	(0.088)	(0.10)
Other household structure (=1)	0.426***	-0.434***		
	(0.12)	(0.15)		
Number of children aged 0-8	0.072***	-0.131***	0.008	-0.058**
	(0.022)	(0.036)	(0.017)	(0.022)
Number of parcels of land	-0.168**	0.303***		
	(0.066)	(0.092)		
Size of landholdings ('000 sq meters)			-0.002	0.003
			(0.002)	(0.002)
Wealth index - housing characteristics			-0.035	0.007
			(0.023)	(0.033)
Wealth index - productive assets	0.058	-0.033	-0.005	0.046
	(0.035)	(0.060)	(0.033)	(0.056)
Village affected by hurricane Mitch (=1)	-0.278**	0.163	-0.104*	0.208*
	(0.12)	(0.15)	(0.053)	(0.11)
Distance to night light (km)			0.007*	-0.009
			(0.004)	(0.007)
Live in Tuma region (=1)	-0.283***	0.379***	-0.276***	0.246

	(0.082)	(0.11)		(0.079)	(0.15)
Live in Madriz region (=1)	0.179	-0.080		0.162**	-0.257**
	(0.13)	(0.19)		(0.070)	(0.10)
Population size village	0.001	0.001		0.000	0.001*
	(0.001)	(0.001)		(0.000)	(0.001)
Own house (=1)	0.371**	-0.331*			
	(0.15)	(0.18)			
House is obtained in exchange for service/labor (=1)	0.407***	-0.356*			
	(0.12)	(0.21)			
Address in hacienda & house rented (=1)	-0.231	0.427		-0.204	0.306*
	(0.23)	(0.27)		(0.16)	(0.18)
Probability of attrition prior to program start in comarca				-0.133	-0.475
				(0.31)	(0.97)
Nobody of target sample attrited before program start	0.120	-0.075		0.043	0.006
	(0.11)	(0.15)		(0.11)	(0.16)
Age fixed effects	YES	YES		YES	YES
Strata fixed effects	YES	YES		YES	YES
Supervisor fixed effects	YES			YES	
Observations	297			611	
R-squared	0.51			0.37	

Notes: Outcome data source: 2010 long-term evaluation survey. Regressions include males ages 9–12 at the start of the program in November 2000, weighted to account for sampling. First column shows coefficient of variable alone, second column coefficient of the variable interacted with the early treatment dummy. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE E4-B. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF BEING FOUND DURING INTENSIVE TRACKING PHASE, FEMALES

	Household instrument		Individual instrument	
	(a) Coef (s.e.)	(b) Coef ET interaction (s.e.)	(a) Coef (s.e.)	(b) Coef ET interaction (s.e.)
Early treatment (ET)=1	0.249		-0.0501	
	(0.30)		(0.36)	
Household head no grades attained (=1)	0.128**	-0.112	0.117**	-0.000
	(0.062)	(0.082)	(0.049)	(0.082)
Mother no grades attained (=1)	0.157	-0.021	0.102	-0.009
	(0.10)	(0.13)	(0.065)	(0.088)
Mother not living in same household (=1)			-0.0735	-0.0900
			(0.13)	(0.18)
Father not living in same household (=1)			0.195	-0.0881
			(0.16)	(0.17)
Child of household head (=1)			0.0861	0.0841
			(0.13)	(0.19)
Female household head (=1)			-0.279**	0.245
			(0.12)	(0.15)
Age of household head	-0.0002	-0.005	-0.004	0.001
	(0.003)	(0.004)	(0.003)	(0.004)
Multigenerational household (=1)			0.0978*	0.0121
			(0.053)	(0.085)
Other household structure (=1)	0.145*	-0.294***		
	(0.078)	(0.11)		
Log of size of landholdings	0.0233	-0.0106	0.001	0.012
	(0.017)	(0.021)	(0.008)	(0.011)
Number of parcels of land	-0.106	0.185		
	(0.088)	(0.12)		
Wealth index - housing characteristics			0.0493**	-0.059**
			(0.021)	(0.027)
Wealth index - other assets			0.0200	-0.051
			(0.039)	(0.049)
Village affected by hurricane Mitch (=1)	0.124	-0.148		
	(0.095)	(0.12)		
Altitude of village	0.0006**	-0.0003	0.0005***	-0.0005**
	(0.0003)	(0.0003)	(0.0001)	(0.0002)

Village in coffee producing area (=1)	-0.225	-0.0873	-0.235***	0.0470
	(0.14)	(0.16)	(0.082)	(0.10)
Live in Tuma region (=1)	-0.130	0.0761	-0.121**	-0.00271
	(0.095)	(0.14)	(0.059)	(0.099)
Live in Madriz region (=1)	-0.0254	0.0456	0.120	-0.0302
	(0.15)	(0.18)	(0.073)	(0.11)
Population size village	0.0009	0.0004	0.0010*	0.0001
	(0.0006)	(0.0007)	(0.0005)	(0.0005)
Own house (=1)	0.0404	0.120		
	(0.12)	(0.15)		
House is obtained in exchange for service/labor (=1)	-0.0615	0.219	-0.223**	0.172
	(0.15)	(0.18)	(0.10)	(0.14)
Address in hacienda & house rented (=1)	-0.193	0.261		
	(0.18)	(0.24)		
No individual in targeted age cohort had attrited before program start			-0.113**	0.144**
			(0.052)	(0.070)
Age fixed effects	YES	YES	YES	YES
Strata fixed effects	YES	YES	YES	YES
Observations	444		584	
R^2	0.36		0.34	

Notes: Outcome data source: 2010 long-term evaluation survey. Regressions include females ages 9–12 at the start of the program in November 2000, weighted to account for sampling. First column shows coefficient of variable alone, second column coefficient of the variable interacted with the early treatment dummy. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX F: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —DOUBLE DIFFERENCE

Over the course of years 2000–2005, the three-year CCT operated in all rural areas of the six municipalities where the evaluation took place. This included the 42 rural localities randomized into early and late treatment as well as the 17 rural localities not initially targeted because they were less poor according to the marginality index used for geographic targeting. In these additional 17 localities, the program began in 2001 and was offered to approximately 80 percent of the population based on a household-level proxy means targeting model. Consequently, by 2005 the three-year program had been implemented (at different times and to modestly different degrees) in all 59 rural localities in the six municipalities, and more than 90 percent of the population had been eligible. (See appendix C for further details.) Given this high coverage, it is possible to use national census data to provide evidence on absolute program effects, albeit for a somewhat limited set of educational and demographic outcomes available in the census.

Specifically, we use the 1995 and 2005 censuses and a non-experimental double difference approach to estimate absolute program effects five years after the CCT began in the early treatment group. Together, the two national censuses provide repeated cross sections at the individual level, in 1995 before the start of the program and in 2005, the year the program ended.[19] The censuses provide information on current municipality of residence (and whether rural or urban), as well as municipality of residence five years prior to the census administration date and, where relevant, municipality of residence at birth.[20] As such, they allow us to account for selection due to domestic migration and therefore provide important complementary evidence that is possibly less vulnerable to selection bias due to migration-related attrition. We assigned all individuals to the municipality where they lived five years prior to each census and assume the type of prior residence (rural or urban) is the same as the current residence.[21] We calculated double difference impacts using the various relevant age cohorts (calculating ages on November 1 as done for the main analyses using the experimental data, in 1990 and in 2000) and comparing changes in educational and other outcomes in rural areas of the six program municipalities to changes in rural areas of the six adjacent municipalities that never received the program but had localities with similar marginality index scores. We estimate

$$Y_{imt} = \delta_0 + \delta_1 T_{m,t-5} + \delta_2 C_t + \delta_3\, T_{m,t-5} * C_t + \varepsilon_{imt} \quad (1)$$

where Y_{imt} is the educational outcome for individual *i* in municipality *m* measured in census year *t*, $T_{m,t-5}$ is an indicator for whether they resided in a treatment municipality five years prior to the census year, and C_t an indicator for the 2005 census. δ_3 yields the double difference estimate of the effect of the program on *Y*, five years after the program began.[22] Outcomes examined include

---

[19] It is not possible to link individuals across the two census rounds.

[20] While each census round also includes more detailed location information, changes in the definitions and boundaries of census areas between 1995 and 2005 make it impossible to match them across time. Municipality boundaries, however, remained constant.

[21] For the age cohorts under consideration, mobility over this period was limited; for example, only about three percent of the 9–12 year-olds moved during the five years previous to each census.

[22] δ_3 is the average impact for the three different groups—early treatment, late treatment, and the other 17 localities—each of which by 2005 had been eligible for the three-year program. It is not possible to isolate the three distinct treatment group areas within the census data (see previous footnote).

grades attained, enrollment, literacy, whether ever married, and for women, whether she has had a live birth. Standard errors are robust to heteroskedasticity. We examine alternative age cohorts as done for the experimental estimates and assess common trends as well as alternative comparison groups.

Because the CCT did not operate in urban areas, we limit the sample to individuals living in rural areas. The main double difference estimation equation takes a first difference between outcomes measured in 2005 for those living in program and non-program comparison municipalities in 2000, and a second difference between outcomes measured in 1995 for those living in program and non-program municipalities in 1990, as indicated by municipality of residence five years before the census administration date. Apart from the CCT there were no other major education-related programs in these municipalities during the period. The main results are presented in Table 10 in the text and reproduced below.

We present the double difference impacts after five years for three age cohorts: 7–8, 9–12, and 13-year-olds for males (Table F1-A) and females (Table F1-B) separately. With the exception of ever having been married effects for the younger cohort are similar in magnitude and significance to the 9–12 year-olds for both males and females, consistent with the program having had a positive absolute effect on these individuals. Less than 1 percent of the 7–8 year-old cohort at the start of the program were married by 2005 given their relatively young ages; there was, however, a small negative effect (2.2 percentage point reduction) in the probability of having been married for females in the 9–12 year-old cohort. This effect is mirrored in the reduction for the same cohort of having had a live birth. Evidence of effects for the 13-year-old males is similar to the younger cohort, but for 13-year-old females somewhat weaker though not statistically significantly different.

We next explored the sensitivity of the double difference findings for the 9–12 year-old cohort to different sets of comparison municipalities and definitions for treatment status than those presented in Table F1. First, we expanded the comparison municipalities to include rural areas in all non-program municipalities in the central regions of Nicaragua where the program was located. Second, we examined whether results differ when we instead use current residence or, separately, municipality of residence at birth to determine program eligibility. While there were some differences in point estimates, the findings are robust to these alternatives and the different approaches all point to significant positive absolute impacts of the program on educational outcomes for both genders after five years, and negative absolute impacts on marriage and fertility for females.

Finally, we provide evidence in support of the identifying assumption by estimating program "effects" on outcomes for two different cohorts of individuals unlikely to have been affected by the intervention—household heads in households of each boy or girl in the 9–12 year-old cohort (using one observation per household) and an older cohort of men or women 30–34 at the start of the program. The same empirical specification suggests there are no effects on educational outcomes or marriage for either group (Tables F2 and F3). Moreover, point estimates are all close to zero, providing support for common trends.

TABLE F1-A: 2005 ABSOLUTE IMPACTS ON EDUCATION AND CIVIL STATUS, BY AGE COHORT, MALES

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1
	(1)	(2)	(3)	(4)	(5)
δ3 for 7–8 age cohort	0.479***	0.104***	0.032*	0.084***	-0.001
	(0.076)	(0.018)	(0.018)	(0.017)	(0.001)
δ3 for 9–12 age cohort	0.597***	0.124***	0.037***	0.091***	-0.003
(*as in Table 11*)	(0.078)	(0.014)	(0.014)	(0.013)	(0.004)
δ3 for 13 age cohort	0.494**	0.078**	0.046*	0.036	-0.024
	(0.198)	(0.031)	(0.025)	(0.029)	(0.019)
N for 7–8 age cohort	11,056	11,056	11,068	11,045	11,067
N for 9–12 age cohort	18,399	18,399	18,421	18,403	18,421
N for 13 age cohort	4,144	4,144	4,148	4,145	4,148
N Total	33,599	33,599	33,637	33,593	33,636
Mean comparison group 7–8	2.965	0.416	0.709	0.765	0.002
Mean comparison group 9–12	3.922	0.559	0.456	0.779	0.017
Mean comparison group 13	4.321	0.580	0.271	0.746	0.091
P-value (7–8 vs 9–12)	0.282	0.376	0.833	0.523	0.053
P-value (9–12 vs 13)	0.623	0.179	0.764	0.627	0.191

Notes: Data source: 1995 and 2005 Nicaraguan national censuses. Regressions include males ages as shown in November of 1990 and 2000, respectively, who were living in rural areas of the six program or six comparison municipalities. δ3 is the double-difference estimate of the absolute program effect in 2005. The mean of the comparison group is for the six comparison group municipalities in 2005. Heteroskedasticity-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE F1-B: 2005 ABSOLUTE IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY, BY AGE COHORT, FEMALES

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever married =1	Has had live birth=1
	(1)	(2)	(3)	(4)	(5)	(6)
δ3 for 7–8 age cohort	0.418***	0.099***	0.028	0.084***	0.000	-
	(0.081)	(0.019)	(0.018)	(0.017)	(0.002)	
δ3 for 9–12 age cohort	0.536***	0.105***	0.028*	0.066***	-0.022**	-0.023**
(*as in Table 6*)	(0.084)	(0.015)	(0.015)	(0.013)	(0.011)	(0.010)
δ3 for 13 age cohort	0.320	0.079**	-0.020	0.036	0.023	0.023
	(0.212)	(0.032)	(0.026)	(0.029)	(0.033)	(0.033)
N for 7–8 age cohort	10,260	10,260	10,271	10,260	10,271	-
N for 9–12 age cohort	17,061	17,061	17,075	17,061	17,075	14,104
N for 13 age cohort	3,728	3,728	3,736	3,728	3,736	3,408
N Total	31,049	31,049	31,082	31,049	31,082	17,512
Mean comparison group 7–8	3.462	0.525	0.776	0.825	0.005	-
Mean comparison group 9–12	4.542	0.642	0.502	0.830	0.162	0.085
Mean comparison group 13	4.754	0.620	0.260	0.785	0.426	0.320
P-value (7–8 vs 9–12)	0.312	0.788	0.996	0.387	0.053	-
P-value (9–12 vs 13)	0.343	0.450	0.114	0.349	0.191	0.190

Notes: Data source: 1995 and 2005 Nicaraguan national censuses. Regressions include females ages as shown in November of 1990 and 2000, respectively, who were living in rural areas of the six program or six comparison municipalities. Birth information unavailable for 7–8 year-old cohort (too young at time of census) and unreported (missing) for 7–17 percent of observations in other age cohorts. Heteroskedasticity-robust standard errors are given in parentheses. δ3 is the double-difference estimate of the absolute program effect in 2005. The mean of the comparison group is for the six comparison group municipalities in 2005. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE F2: 2005 : ASSESSMENT OF COMMON TRENDS WITH PLACEBO OUTCOMES: IMPACTS ON HOUSEHOLD HEAD CHARACTERISTICS

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Female Head =1	Age of Head in Years
	(1)	(2)	(3)	(4)	(5)
Panel A: Heads of Households with 9–12 Year-Old Male					
Treatment municipality * 2005 (δ_3)	-0.020	0.014	0.005	0.007	0.342
	(0.079)	(0.013)	(0.016)	(0.013)	(0.417)
N	15,192	15,192	15,286	15,292	15,292
Mean comparison group 2005	1.673	0.226	0.507	0.221	46.901
Panel B: Heads of Households with 9–12 Year-Old Female					
Treatment municipality * 2005 (δ_3)	0.009	0.003	0.018	0.018	0.420
	(0.084)	(0.014)	(0.017)	(0.013)	(0.452)
N	14,438	14,438	14,512	14,524	14,524
Mean comparison group 2005	1.814	0.251	0.515	0.208	45.290

Notes: Data source: 1995 and 2005 Nicaraguan national censuses. Regressions include heads of households with a male or female ages 9–12 in November of 1990 and 2000, respectively, who were living in rural areas of the six program or six comparison municipalities. The mean of the comparison group is for the six comparison group municipalities. δ3 is the double-difference estimate of the placebo absolute program effect in 2005 on the head outcome. Heteroskedasticity-robust standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$.

TABLE F3: 2005 : ASSESSMENT OF COMMON TRENDS WITH PLACEBO OUTCOMES: IMPACTS ON OLDER (30–34) AGE COHORTS

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Ever Married =1
	(1)	(2)	(3)	(4)
Pa*nel A: Absolute Impacts on Outcomes, Males*				
Treatment municipality * 2005 (δ_3)	0.145	0.034	0.008	0.003
	(0.127)	(0.021)	(0.022)	(0.016)
N	8,256	8,256	8,274	8,279
Mean comparison group 2005	2.861	0.341	0.676	0.843
Pa*nel B: Absolute Impacts on Outcomes, Females*				
Treatment municipality * 2005 (δ_3)	-0.054	-0.005	0.004	0.009
	(0.124)	(0.019)	(0.021)	(0.013)
N	8,772	8,772	8,783	8,790
Mean comparison group 2005	2.914	0.354	0.610	0.884

Notes: Data source: 1995 and 2005 Nicaraguan national censuses. Regressions include males or females 30–34 years old in November of 1990 and 2000, respectively, who were living in rural areas of the six program or six comparison municipalities. The mean of the comparison group is for the six comparison group municipalities. δ3 is the double-difference estimate of the placebo absolute program effect in 2005 on the adult outcome. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX G: INSIGHTS FROM MEDICAL AND NUTRITIONAL LITERATURE ON AGE OF MENARCHE

The medical and nutrition literatures provide the scientific basis for understanding the role of nutritional status and nutritional shocks for the onset of puberty. A series of studies show that poor childhood nutrition is associated with delayed puberty for females and that better childhood nutrition and health are associated with earlier menarche (Garn, 1987; Cooper et al., 1996).[23] Originally hypothesized to be directly related to attaining a critical weight (Frisch and Revelle, 1971), more recent work indicates menarche is related to a minimum body fat mass (Blum et al., 1997). Nutritional status in childhood is believed to affect menarche through the leptin hormone (which helps regulate the body's energy balance[24]), with some uncertainty as to whether leptin plays a permissive versus a triggering role (Shalitlin and Philip, 2003; INSERM, 2007). Leptin levels fluctuate with nutritional intakes, pointing to the mechanisms through which nutritional shocks can translate into relatively sudden advances or delays in the age of menarche.[25]

Nutrition also may affect the onset of puberty differently at different stages of childhood. The well-documented phenomenon for migrant and adopted children of early onset puberty (Mul et al., 2002; Parent et al., 2003) is often explained by an interaction of prenatal undernutrition with a subsequent enriched nutritional context during childhood (Gluckman and Hanson, 2006). Koziel and Jankowska (2002) show that females with high BMI at age 14 were more likely to have had early menarche when they also had low birth weight, but not otherwise. Sloboda et al. (2007) find that combined low birth weight and high BMI at age eight predict early age at menarche. To the best of our knowledge, however, the medical literature does not provide firm conclusions on the specific ages at which positive or negative nutritional shocks matter most. A review by Kaplowitz (2008), for example, points to the need for further study of the relationship between nutrition and puberty from ages as early as 6–7 all the way up to 13–14. Both cross-sectional and longitudinal evidence indicate rises in leptin concentration between ages 7 and 15, that are paralleled by increases in body fat during female puberty—as well as strong correlations between leptin, body fat mass, and age of menarche (Blum et al., 1997; Garcia-Mayor et al., 1997; Ahmed et al., 1999).

Overall, numerous studies document the close correlation between body fat mass and the onset of puberty, even if it is not entirely clear whether excess weight induces early sexual maturity, or whether early sexual maturity triggers excess weight (INSERM, 2007).[26] The links between body fat and the female reproductive system females have been interpreted as mechanisms for ensuring that pregnancy does not occur unless there are adequate fat stores to sustain both mother and fetus—studies with rodents indicate those mammalian females are able to turn off their reproductive systems when food intake is inadequate (Kaplowitz, 2008).

---

[23] Age of menarche declined from 17 to 14 in the United States and several Western European countries between the mid 19th and 20th centuries, a trend believed to be directly related to improvements in nutrition and health (INSERM, 2007).

[24] When fat stores are low, decreased leptin leads to increased appetite, helping to restore body fat and body weight.

[25] Various lab experiments with rodents and monkeys show that injection of leptin led to sudden changes in the age of menarche and reproductive outcomes (Chehab et al., 1996, 1997; Cheung et al., 2001; Wilson et al., 2003).

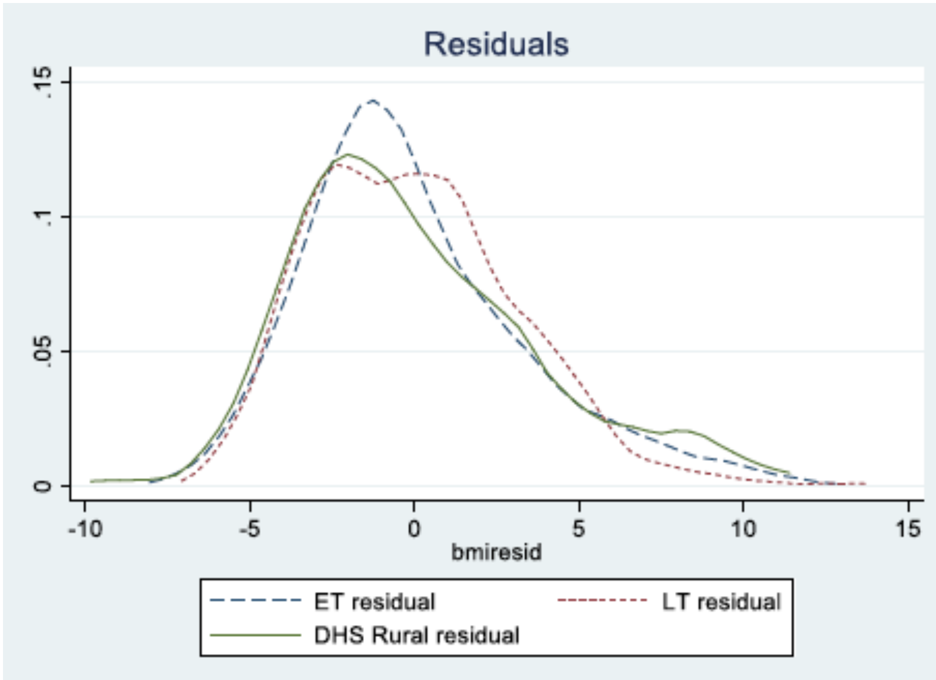
[26] More generally, compensatory postnatal growth during childhood combined with poor prenatal growth is often argued to lead to adverse health outcomes (Hales et al, 2003).

Based on this evidence from the medical and nutrition literatures, we hypothesize that the differential timing of nutritional shocks in early versus late treatment groups could have led to differences in the age of menarche, as well as adult female BMI.

To explore this further, we compare 2010 BMI distributions for the early- and late-treatment group 9–12 year-old cohorts with same-age girls from the Nicaraguan Demographic and Health Survey (DHS) carried out in in 2011–2012. For a plausibly comparable sample we restrict the DHS to young women 19–22 years old living in rural areas of the two departments of Nicaragua where the program operated (Madriz and Matagalpa), excluding municipalities in which the program had been implemented. Compared to women exposed to the program, women in the DHS subsample were 10+ percentage points more likely to have had their first sexual encounter by age 15, 4+ percentage points more likely to be married, and 10+ percentage points more likely to have had a child. They also had slightly less schooling than those exposed to the program, averaging 6.28 grades attained. While not impact estimates, these patterns are consistent with the possibility that the CCT led to absolute improvements on these key indicators for both early- and late-treatment groups.

On the other hand, average BMI (for non-pregnant) women in the DHS was higher, 24.3 compared to 23.6 in late- and 23.0 in early-treatment (before adjustments). Because prior pregnancy influences BMI, we condition on whether the woman has ever been pregnant (and age) and then compare the distributions of the residuals (Figure G1). Kolmogorov-Smirnov tests indicate that the early- and late-treatment group distributions are significantly different ($p=0.075$). Both the early- and late-treatment group distributions appear to match the DHS distribution closely in the left tail, but have visually different shapes going toward the right. This suggests the possibility that BMI was influenced in both groups, possibly more so in the late-treatment, and probably not to the detriment of the CCT groups at the lowest levels of nourishment. Comparisons (early-treatment versus DHS and late-treatment versus-DHS) are not significant (with p-values > 0.5) with these samples, possibly due to the relatively small sample size from the DHS [N=140].

FIGURE G1: CONDITIONAL COMPARISONS OF BMI, EARLY-TREATMENT, LATE-TREATMENT AND DHS RURAL MADRIZ AND MATAGALPA



Notes: Data sources: 2010 long-term evaluation individual survey instrument survey and 2011–12 Nicaraguan Demographic and Health Survey. Estimated residuals from model of BMI controlling for age (single-year indicator variables) and whether ever pregnant in each subsample. ET = Early Treatment; LT = Late Treatment.

APPENDIX H: SPILLOVERS AND FAMILY KINSHIP NETWORKS

In this appendix we consider the possibility of heterogenous differential program effects due to spillovers and separately, pre-program family kinship networks based on surnames.

Spillovers—Experimental differential estimates may differ from absolute treatment effects if there are any spillovers between the early and late treatment groups. As described in the paper (Section II.B), a set of 42 rural localities in six municipalities were randomized into early or late treatment. Many of these shared geographic borders (with no buffer zone), including some with the opposite treatment group, opening up the possibility of spillover effects across areas. In this appendix, we describe the methodology we used to test for spillover effects using the GPS locations of sample households before the program at baseline.

In contrast to a randomized design maintaining a control group, for which one might expect spillovers only in one direction (whether positive or negative), the phase-in design of early and late treatment implies that the direction of spillovers could have been different at different times. For example, if there were positive spillovers from treatment to non-treatment areas during the CCT operations, then there could have been spillovers from early to late treatment areas in the first years of the program when only early treatment households received transfers, followed by spillovers from late to early treatment areas in the latter years when only late treatment households received transfers. Even if all spillover effects were positive, the net effect of spillovers on the differential (measured after 10 years), would be ambiguous.

To explore the potential influence of spillover effects on the main differential findings, we use GPS location data for sample households and analyze whether differential program effects vary depending on whether households were living in close proximity to a high concentration of others receiving the program during early treatment.[27] To do so, we use the variation in the fraction, or density of households that are early-treatment households in a three-kilometer radius around the household of each individual, variation that results from the locality-level randomization and also reflects the feature that locality borders (based on 1995 national census segments) do not necessarily correspond to village borders in all locations. The density within the three-kilometer radius circle itself may capture other unobserved characteristics of the local environment, but the interaction of density and early treatment provides an estimate that accounts for spillovers, as well as for any potential multiplier effects among early treatment households. The possibility of the latter, in areas where treatment density was high, is consistent with evidence from contemporaneous qualitative work. For example there was strong community enthusiasm for improved schooling during the CCT, evidenced in some areas by local mobilization to hire additional teachers so that students could fulfill the conditions (Adato and Roopnaraine, 2004).

Results for the primary 9–12 year-old cohorts in Tables H1 suggest that if anything differential treatment effects for both males and females appear to be larger when early treatment density is higher, but only a small number of the interactions are significant. Noting we use fractional representations of density, the effect sizes are for the most part modest. This pattern is consistent with the possibility of a positive multiplier effect within the early treatment group.

---

[27] Initial baseline GPS locations are available only for households targeted for 2010 follow-up and not for all households in the 2000 program census. We use the observed program density in the sample as a proxy for density in the population, with confidence that the high sampling proportion makes it a good estimate.

Family Kinship Networks—Family networks have been shown to be important for anti-poverty programs in a variety of contexts. Moreover, heterogeneous treatment effects related to such networks can shed light on the mechanisms underlying impacts. Building on the literature examining the importance of family and broader kinship networks for impacts of CCTs in the short run (Angelucci and Di Giorgio, 2009; Angelucci et al., 2010), we explored to what extent program effects differ depending on the strength of pre-existing family networks.

The program census carried out in May 2000 included individual rosters for all households in treatment localities. In addition to the demographic and schooling information, the census also recorded and digitized first and last names of household members for program administration. In many parts of Latin America including Nicaragua, individuals have two surnames and naming conventions commonly follow the pattern in which a child takes the first surname of their father as their first surname, and the first surname of their mother as their second surname. For example, if the paternal surname is Godoy Sandino (F1, f1) and the maternal surname is Darío Martí (F2, f2), then their child's surname would be Godoy Darío (F1, F2).

Similar to the approach taken for constructing family networks in analyses of the Mexican CCT *PROGRESA* (Angelucci and De Giorgi, 2009; Angelucci et al., 2010; Angelucci, De Giorgi, and Rasul, 2018), we use this naming convention to construct different measures of the potential family network in each village based on surname matches. Specifically, we use the program census data to determine, for each index individual in the sample, the total number of other individuals living in the same village[28] who have at least one surname in common with the index individual, regardless of its position (i.e., first or second surname). Alternatives to this definition, for example requiring the same position (first or second) or using the name of the household head instead, lead to quantitatively smaller networks that are highly correlated with this broader definition. While we cannot verify that all individuals within the family network as defined here would have a direct familial relationship, many likely do (including siblings and half-siblings, cousins, grandfathers, and aunts/uncles). Median family network size in these rural communities is nearly 60 persons (average 83) and the median family network as a share of the total number of individuals in the community is nearly 20 percent (average 26 percent). We transform the family network size into a binary variable indicating above and below the median network size, in part to mitigate concerns about measurement error for those with the same name who do not have direct family relationships (Angelucci et al., 2010).

We use these constructed variables to analyze heterogeneous treatment effects by initial family network size. Results for the primary 9–12 year-old cohorts are presented in Tables H2. With the additional variables, the interpretation on the treatment variable is now the ITT differential impact for individuals with below median family network size at baseline and the interaction captures differences in that impact for those with initially high (above median) family network size. There is little evidence of heterogeneous treatment effects along these lines for males. For females, the results are suggestive for some outcomes (and significant for the learning family) that effects were larger for females with higher than median sized networks. Assuming the lack of more significant differential effects is not due only to low power,[29] however, we conclude that the long-term program effects were not systematically concentrated among those with initially stronger or weaker family networks as measured here.

---

[28] We use village, as defined in the 1995 National Census, as the relevant geographic area rather than the larger locality area used for randomization and clustering in the evaluation, as the former constitutes a more concentrated settlement and is arguably a more natural geographic area in which to consider network effects.
[29] For comparison, Angelucci et al. (2010) have a male sample four times as large.

TABLE H1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY EARLY TREATMENT DENSITY 9-12 COHORT

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-Score	Socio-emotional Family Z-Score
		Rank	Absolute (5% Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Males							
Early treatment density	0.093	0.085	0.233	-0.258**	-0.223	0.485	-0.318**
	(0.203)	(0.227)	(0.214)	(0.123)	(0.203)	(0.361)	(0.155)
ITT * Early treatment density	0.238	0.481	0.434	0.124	0.350	-0.372	0.399*
	(0.327)	(0.289)	(0.290)	(0.194)	(0.333)	(0.430)	(0.215)
N	1,006	1,006	997	1,007	907	907	900
Panel B: Females							
Early treatment density	0.103	0.117	0.116	-0.245	-0.149	0.267	0.035
	(0.275)	(0.287)	(0.304)	(0.177)	(0.200)	(0.171)	(0.158)
ITT * Early treatment density	-0.015	-0.061	0.059	0.625***	0.290	-0.437	0.463*
	(0.356)	(0.374)	(0.376)	(0.213)	(0.257)	(0.275)	(0.261)
N	883	883	873	883	823	805	817

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5 and 7 using the 2010 individual survey instrument, and column 6 using both. Regressions include males and females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group. Early treatment density is defined as the fraction of early treatment sample households over all sample households within a three-kilometer radius. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE H2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY FAMILY NETWORK SIZE 9–12 COHORT

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education Family Z-Score	Learning Family Z-Score	Fertility Family Z-Score	Socio-emotional Family Z-Score
		Rank	Absolute (5% Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Males							
ITT	0.260**	0.185*	0.177**	0.127*	0.147	-0.083	0.096
	(0.099)	(0.094)	(0.085)	(0.069)	(0.089)	(0.075)	(0.063)
Family network (>median =1)	0.137	0.001	-0.033	-0.029	0.110	-0.114	0.120*
	(0.105)	(0.143)	(0.127)	(0.065)	(0.081)	(0.086)	(0.066)
ITT * Family network (>median =1)	-0.026	0.015	0.038	-0.040	0.025	0.080	-0.116
	(0.133)	(0.168)	(0.164)	(0.089)	(0.105)	(0.105)	(0.095)
N	1,006	1,006	997	1,007	907	907	900
Panel B: Females							
ITT	0.070	0.043	0.040	0.084	-0.098	-0.186**	0.008
	(0.117)	(0.111)	(0.111)	(0.074)	(0.075)	(0.074)	(0.093)
Family network (>median =1)	0.094	0.076	0.053	0.027	-0.136*	-0.079	0.052
	(0.113)	(0.103)	(0.092)	(0.063)	(0.077)	(0.069)	(0.086)
ITT * Family network (>median =1)	0.174	0.128	0.114	0.019	0.210**	0.048	-0.128
	(0.180)	(0.163)	(0.156)	(0.102)	(0.102)	(0.090)	(0.124)
N	888	888	878	888	826	808	820

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument and columns 5 and 7 using the 2010 individual survey instrument, and column 6 using both. Regressions include males and females ages 9–12 at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. Family z-scores are calculated by averaging the z-score for the individual components in each family, with individual z-scores calculated using the mean and standard deviation of the late treatment group. Family network size is the number of individuals with common surnames in the village. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX I: COST-BENEFIT CALCULATIONS

A comprehensive cost-benefit analysis of the Nicaraguan CCT program would require monetary assessment of all of the resource costs as well as all of the potential (positive or negative) effects of the CCT. In addition, such an assessment is complicated by of the particular experimental research design in which we estimate differential rather than absolute effects. In this appendix, we outline the approach taken for approximating the net present value (NPV) of the program focusing on males in the main 9–12 year-old cohort, and arguing this is likely a conservative estimate for the entire combined cohort of 7–13 year-old males and females given the design of the program, its costs, and the pattern of experimental differential impacts.

Following the approach outlined in Dhaliwal et al. (2012), all values are first translated into U.S. dollars using market exchange rates after which they are deflated to 2000, the first year of the program, using the U.S. CPI. As recommended for cost-benefit analysis of transfer programs, we include all direct program costs for running the program (such as administrative, management, targeting, monitoring, delivery of transfers, and conditionality costs) but not the transfers themselves or evaluation costs (reported in Caldés and Maluccio 2005). We allocate 50 percent of the average per household program cost as an estimate of costs related to the components of the program for the 7–13 year-old cohorts in the early treatment group, in nominal terms $60 a year for three years. (This assumes that the other 50 percent were costs related to the nutrition and health components of the program aimed at the younger children, and the health education workshops.) Benefits are calculated based on estimates for males in the 9–12 year-old cohort as the estimated average increase in monthly earnings ($10) starting in 2009 (when we first measure them), multiplied by average number of months worked (3.5) and by the average number of males in the cohort in each household at baseline in 2000 (1.37). All these values are first deflated to year 2000 constant U.S. dollars and then discounted at 5 or 10 percent discount rates. Using a 10 percent discount rate, the NPV turns positive in 2027, approximately 20 years after the program ended; at 5 percent, the NPV turns positive in 2016.

The above approach does not directly account for other important costs related to the potential deadweight losses associated with taxation necessary to make the transfers or the possibility that the CCT crowded out other governmental expenditures with higher returns (Harberger, 1997). Assuming households in the early treatment received full education transfers for three years and that the cost of raising those funds was 5 percent, the NPV with a 10 percent discount rate would take an additional eight years to turn positive, in 2035.

While strictly speaking it is impossible to be certain that there are no substantive negative absolute impacts of the program for males or for females[30] that should be treated as additional costs, at each discount rate arguably the estimates are a lower bound on NPV for several reasons. First, we attribute all of the costs of the education component to a subset of the beneficiaries even though eligible males 7–8 and 13, as well as females 7–13 were covered by those costs. Second, we only consider benefits of one type—increased off-farm earnings. For example, this only includes benefits from the schooling, learning, or other gains to the extent that they operate directly through the labor market. And, while in theory off-farm earnings gains might have been

[30] Even for females, where the differential estimates indicate modest negative effects associated with the mechanisms, for example on BMI, since both early and late treatment females received program components (albeit at different ages) it is plausible to assume the absolute impact on each group was non-negative.

offset by reduced earnings on the family farm, the estimates indicate that the CCT did not decrease work on the family farm at the extensive margin so that a reduction in farm earnings is unlikely. Third, we do not incorporate any potential gains for the younger (7–8) and older (13) age cohorts of males, as well as for all females (7–13). Analyses reported in the paper for males of different ages and for females point to positive, or at least non-negative, effects on labor market outcomes for these other groups as well. Because they are not all statistically significant and because for females there were modest negative differential effects on a few other outcomes although the evidence points to non-negative absolute effects for both treatment groups, however, we do not directly incorporate them into the cost-benefit calculations. Fourth, we only consider benefits starting in 2009 but for some the benefits likely began earlier. Last, we ignore program costs associated with the late treatment group, which even when discounted would further offset the costs of the program in early treatment in the differential framework.

Overall, under relatively conservative assumptions regarding benefit flows (for a subset of the beneficiaries) and costs related to the CCT, these coarse estimates suggest that the program related activities for the 7–13 year-old cohorts could achieve positive NPV within a few decades.

Appendix J: Additional Appendix References

Below we provide additional references not cited in main text.

Adato, M. and T. Roopnaraine. 2004. "A social analysis of the Red de Protección Social. Report submitted to the *Red de Protección Social*." International Food Policy Research Institute. Washington, DC.

Adhvaryu, A., Molina,T., Nyshadham, A., and Tamayo, J. 2018. "Helping Children Catch Up: Early Life Shocks and the *PROGRESA* Experiment." Mimeo, University of Michigan.

Ahmed ML, Ong KK, Morrell DJ, et al. 1999. "Longitudinal study of leptin concentrations during puberty: sex differences and relationship to changes in body composition." *J Clin Endocrinol Metab.* 84(3):899–905

Angelucci, M. and G. De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review* 99(1): 486–508.

Angelucci, M., G. De Giorgi, and I. Rasul. 2018. "Consumption and Investment in Resource Pooling Family Networks." *Economic Journal*, 128(November): 2613–51.

Angelucci, M., G. De Giorgi, M.A. Rangel, and I. Rasul. 2010. "Family Networks and School Enrolment: Evidence from a Randomized Social Experiment." *Journal of Public Economics* 94: 197–221.

Attanasio, O., S. Cattan, E. Fitzsimon, C. Meghir, and M. Codina. 2020. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." *American Economic Review*, 101(1): 48-85.

Baird, S., J.H. Hicks, M. Kremer, and E. Miguel. 2016. "Worms at Work: Long-run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131 (4): 1637–80.

Blattman, C., N. Fiala, S. Martinez. 2014. "Generating Skilled Self-employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics*: 697–752.

Caldés, N., and J.A. Maluccio. 2005. "The Cost of Conditional Cash Transfers." *Journal of International Development* 17(2): 151–68.

Chehab FF, Lim ME, Lu R. 1996. "Correction of the sterility defect in homozygous obese female mice by treatment with human recombinant leptin." *Nat Genet.* 12(3):318–20

Chehab FF, Mounzih K, Lu R, Lim ME. 1997. "Early onset of reproductive function in normal female mice treated with leptin." *Science.* 275(5296):88–90

Cheung CC, Thornton JE, Nurani SD, Clifton DK, Steiner RA. "A reassessment of leptin's role in triggering the onset of puberty in the rat and mouse." *Neuroendocrinology.* 2001;74(1): 12–21

Cunha, F., J.J. Heckman, and S.M. Schennach. 2010. "Estimating the Technology of Cognitive and Non-Cognitive Skill Formation." *Econometrica* 78(3): 883–931.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch. 2013. "Comparing Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education." in P. Glewwe, Ed. *Education Policy in Developing Countries*, Oxford UK: Oxford University Press.

Doyle, O., C. Harmon, J.J. Heckman, C. Logue, and S.H. Moon. 2017. "Early skill formation and the efficiency of parental investment: A randomized controlled trial of home visiting." *Labour Economics*, 45:40–58.

Duflo, E. Dupas, P. and Kremer, M. 2017. "The Impact of Free Secondary School Education: Experimental Evidence from Ghana." Mimeo, Stanford University.

Frisch RE and Revelle R. 1971. "Height and Weight at Menarche and a Hypothesis of Menarche." *Arch Dis Child,* 46(249):695–701

Garcia-Mayor RV, Andrade A, Rios M, Lage M, Dieguez C, Casanueva FF. “Serum Leptin Levels in Normal Children: Relationship to age, gender, body mass index, pituitary-gonadal hormones, and pubertal stage.” *J Clin Endocrinol Metab.* 1997; 82(9): 2849–55.

Hales C.N. and S.E. Ozanne. 2003 "The Dangerous Road of Catch-up Growth." *J Physiol* (Lond) 547: 5–10.

Harberger, A.C. 1997. “New Frontiers in Project Evaluation? A Comment on Devararjan, Squire, and Suthiwart-Narueput.” *World Bank Research Observer*, 12(February): 73–9.

IFPRI. 2005. *Sistema de Evaluación de la Red de Protección Social (RPS) – Mi Familia, Nicaragua: Evaluación de Impacto 2000–04*, Report submitted to the *Red de Protección Social*. International Food Policy Research Institute, Washington, DC. Photocopy.

Lee, D.S. 2009. “Training, Wagers, and Sample Selection: Estimating Sharp Bounds on Treatment Effect.” *The Review of Economic* Studies 76(3):1071–1102.

Kaplowitz, P.B, 2008. “Link Between Body Fat and the Timing of Puberty”, *Pediatrics*: 121; S208.

Koziel S, and EA. Jankowska, 2002. “Effect of low versus normal birthweight on menarche in 14-year-old Polish females.” *J Paediatr Child Health* 38: 268–27.

Kugler, A.D., and Rojas, I. 2018. “Do CCTs Improve Employment and Earnings in the Very Long-Term? Evidence from Mexico”. NBER Working Paper No. 24248.

Maluccio, J. A. 2009. “Household Targeting in Practice: The Nicaraguan Red de Protección Social.” *Journal of International Development* 21(1): 1–23.

Orr, L. J. Feins, R. Jacob, E. Beecroft, L. Sanbonmatsu, L.F. Katz, J.B. Liebman, and J.R. Kling. 2003. Moving to Opportunity: Interim Impacts Evaluation. Unpublished.

Paxson, C., and N. Schady. 2010. “Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador.” *Economic Development and Cultural Change* 59(1): 187–230.

Radloff, Lenore S. 1977. “The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.” *Applied Psychological Measurement* 1 (3): 385–401.

Regalia, F., and L. Castro. 2007. “Performance-Based Incentives for Health: Demand and Supply-Side Incentives in the Nicaraguan *Red de Protección Social.*” *Center for Global Development Working Paper* No. 119. Washington, DC, United States: Center for Global Development.

Shalitin S, Phillip M. Role of obesity and leptin in the pubertal process and pubertal growth: a review. *Int J Obes (Lond).* 2003; 27(8): 869–74.

Sloboda, DM, Hart R, Doherty DA, Pennell CE, Hickey M. 2007. “Age at menarche: influences of prenatal and postnatal growth”, *Journal of Clinical Endocrinology and Metabolism* 92(1): 46–50.

Wilson ME, Fisher J, Chikazawa K, Yoda R., Legendre A., Mook D. Gould KG. 2003. "Leptin administration increases nocturnal concentrations of luteinizing hormone and growth hormone in juvenile female rhesus monkeys." *J Clin Endocrinol Metab.* 88(10): 4874–83.

World Bank. 2001. “Nicaragua Poverty Assessment: Challenges and Opportunities for Poverty Reduction,” Report No. 20488-NI, The World Bank, Washington, D.C.

World Bank. 2003. “Nicaragua Poverty assessment: Raising Welfare and Reducing Vulnerability,” Report No. 26128-NI. Washington DC: The World Bank.