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]

August 2023

Abstract: Conditional cash transfer 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 3-year CCT program began, earlier exposure during primary school ages when children were at risk of dropout led to higher labor market participation for young men and women and higher earnings for men. Results highlight the roles of the different program components with variation in timing of access to nutrition, health and education investments translating into substantial differential effects 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, Nicaragua

[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 CEPR 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 2010 survey. Finally, we are grateful for comments received from Guillermo Cruces, Pierre Dubois, Esther Duflo, Berk Özler, Norbert Schady, 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 the possibility of breaking the intergenerational transmission of poverty. Theory suggests that investments in child human capital can improve future economic outcomes, for example through higher incomes in adulthood. There is substantial evidence that a variety of interventions can increase human capital in low- and middle-income countries in the short term, but less is known about whether they live up to their promise in the longer term.[2] There is also little evidence on the potential importance of specific timing of interventions in late childhood.

A prominent example in this class of interventions is the conditional cash transfer (CCT) program. Started in 1997 in Mexico and Brazil, CCTs spread to 60+ countries worldwide and covered 25% of Latin America by 2013 (World Bank 2015). Numerous rigorous evaluations demonstrate short-term effects including improvements in nutrition and health for young children, increases in schooling for older children and reductions in poverty for households (Fiszbein and Schady 2009; Bastagli et al. 2016). Evidence on longer-term effects is more mixed (Molina-Millán et al. 2019). Using the randomized phase-in of a CCT, we provide new evidence on how experimental differences in the specific timing of CCT exposure during primary school ages can have long-term differential effects on education, health and labor market outcomes.

More specifically, we exploit the randomized phase-in and eligibility rules of a CCT in rural Nicaragua to estimate differential intent-to-treat (ITT) effects 10 years after the start of the program. All households in the early treatment group were eligible for transfers for three years from late 2000 to late 2003; those in the late treatment group were also eligible for transfers for three years, but from 2003 to 2005. In both treatments, all households were eligible for conditional transfers for nutrition and health, and households with children 7–13 years old who had not completed fourth grade were additionally eligible for conditional transfers for education. The program also increased exposure to reproductive health information and services. Together, the conditional transfers provided incentives and resources for the children to remain in school, as well as means to improve food availability and nutrition for them.[3]

---

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

[3] Short-term evaluation results confirm that the CCT led to relatively large increases in enrollment, grade completion and household food expenditures (Maluccio and Flores 2005). The impacts parallel those of many other first

To examine the long-term effects, we collected data in 2010 on individuals and households first interviewed at baseline in 2000, prior to randomization. Consequently, the follow-up survey measured outcomes approximately 10 years after the program start, which was seven years after households in the early treatment group had stopped receiving transfers. At that point individuals 7–13 years old at baseline were transitioning into early adulthood (17–23 years old) with the majority in their initial years in the labor market. Evidence during this labor market entry period is valuable for understanding the mechanisms underlying the long-term effects of CCTs, particularly if those initial experiences can set individuals on different labor market trajectories.

Given the potential relationship between CCT program exposure and migration for the young adults, we tracked all of them throughout Nicaragua and to Costa Rica, the main international migration destination. International tracking is rare in these types of studies. Final attrition is between 4–22% (depending on the outcome and sample), is balanced between early and late treatment, and yields analysis samples balanced on baseline observables. We consider various estimation strategies to gauge the extent of possible remaining selectivity.

To help understand how the CCT-related investments influenced education, health and labor market outcomes we posited a dynamic human capital production function model (Cunha and Heckman 2007) with three periods corresponding to the two program implementation phases described above and to a third period ending with the follow-up survey in 2010. The possibility of self-productivity, dynamic complementarities and, in particular, sensitive periods for education and health investments during primary school ages helps us interpret how the experimental variation in timing led to longer-term differentials in outcomes.

We find that earlier exposure to the CCT starting at primary school ages when children were at risk of dropout (9–12 years old) increased labor market participation and earnings a decade later when the beneficiaries were young adults. To understand what underlies the gains, we examine several intermediate human capital outcomes along the causal pathway. Much of the CCT literature examining effects on primary school age children focuses on education. But there are other mechanisms through which CCTs can affect beneficiaries and all program components—rather than just the education component—potentially influence later outcomes. In

generation CCT programs that, like the Nicaraguan program we study, promoted nutrition, health and education, and targeted children over a relatively wide age range. Recent CCT programs, in contrast, often focus more narrowly on one specific objective (e.g., education or health) or age group.

addition to education, this paper spotlights the role of improved food availability and nutrition during the preteen years which can influence health and, for girls, the timing of menarche and subsequent sexual and reproductive health outcomes. This is particularly relevant for understanding the effects of differential timing of exposure for girls since their education, fertility and labor market outcomes are closely linked. We therefore examine results by sex.

For both men and women separately, earlier exposure at primary school ages (9–12 years old) increased labor market participation. Differential ITT effects on intermediate outcomes by sex, however, reveal salient differences in mechanisms resulting from the timing of the human capital investments and possibly different sensitive periods, interaction with labor market opportunities and the role of fertility for labor market outcomes of young adults in Nicaragua.

By the 10-year follow-up most young men had completed their schooling and virtually all were working. The differential ITT estimates indicate that earlier exposure to the CCT at ages when they were likely to drop out of school led to sustained schooling and learning gains of nearly 0.2 standard deviations (SD). In addition, they were more likely to have wage work and migrate temporarily for higher paid jobs, earning ~15% more per month worked. The pattern of results for young men is in line with standard human capital theory in which investment during sensitive periods leads to more schooling, more learning and subsequent changes in labor market outcomes including higher earnings. In this regard, results for males provide relatively straightforward evidence on the long-term returns to CCTs and on the potential for CCTs to reduce the intergenerational transmission of poverty through improved education.

For the young women we also estimate positive and significant differential ITT effects on schooling; however, despite short-term evidence of gains in early treatment there are no differential effects on learning. This can be explained in part by prior enrollment patterns. Girls in rural Nicaragua typically remain in school longer than boys during the early teenage years, and this may have enabled late treatment girls to catch up later when their households were benefitting from the CCT, even though they themselves were not directly exposed to the education component. Results also indicate that relative to late treatment, average age of menarche increased for early treatment girls and they became sexually active later. These differences can be linked to the random variation in the timing of CCT transfer-related nutrition shocks occurring during ages important for physical development. Consistent with their older age of menarche, early treatment girls started childbearing later and had lower body mass compared to those in the late treatment. Taken together, the pattern of results for females suggests the long-term effects on their labor market outcomes reflect CCT induced changes in nutrition and related reproductive health outcomes, in addition to changes in schooling.

The findings directly relate to the literature on long-term impacts of exposure to CCTs during school-going ages, including Behrman et al. (2009a, 2011), Barrera-Osorio et al. (2019), Baird et al. (2019), Cahyadi et al. (2020) and Molina-Millán et al. (2020).[4] Prior research largely finds that CCTs lead to higher schooling, but is less conclusive for other outcomes including learning and, with the exception of Parker and Vogl (2018) and Araujo and Macours (2021) for Mexico, there is little evidence on labor market or income gains. There is also no evidence on whether and how the specific timing of the interventions changes program impacts.

Our paper contributes to this literature in several ways. We organized extensive tracking of migrants (resulting in relatively low attrition rates) and collected rich information on the diverse portfolio of economic activities in which young Nicaraguan adults engage. This enabled a comprehensive examination of longer-term impacts on the labor market, which identified temporary migration for work as a path to increasing earnings. Measurement of a wide set of intermediate outcomes allowed us to explore effects along the causal pathway to young adult labor market and earning outcomes. That analysis provides a detailed picture of the ways in which CCT exposure during primary-school age years can lead to longer-term outcomes including the importance of timing of investments and how patterns can differ by sex. Finally, we also found that differential exposure to a CCT affected child learning in this low-income context, in line with some results in the literature (Baird et al. 2011; Duque et al. 2018), but in contrast to other results (Filmer and Schady 2014; Baird et al. 2019).

The findings also relate to literature examining the links between schooling, fertility, marriage and the labor market for women. Research on age of menarche and schooling in India and Bangladesh demonstrates that later menarche is associated with higher schooling and delayed marriage, due at least in part to cultural norms (Field and Ambrus 2008; Khanna 2020). Research on the longer-term impacts of scholarships and educational subsidies reveals that effects on fertility-related outcomes are often directly linked to the educational incentives;

[4] The findings also relate to work examining longer-term impacts on individuals of CCT exposure earlier in childhood (Behrman et al. 2009b; Fernald et al. 2009; Barham et al. 2013; Araujo et al. 2018), as well as impacts on household economic well-being of exposure to CCT or other cash transfer programs (Gertler et al. 2012; Macours and Vakis 2016; Banerjee et al. 2016; Bandiera et al. 2017; Handa et al. 2018; Haushofer and Shapiro 2018).

inducing girls to remain in school longer can postpone childbearing (Baird et al. 2011; Filmer and Schady 2014; Duflo et al. 2015, 2021). Financial incentives for delaying marriage given to teenage girls (initially) not in school in Bangladesh actually increased schooling while reducing marriage and teen childbearing, although empowerment training for them only resulted in higher self-employment (Buchmann et al. 2018). Reproductive health and vocational trainings for adolescent girls in Uganda similarly led to increases in self-employment but also to reductions in teen pregnancy and marriage (Bandiera et al. 2020).

Our paper contributes to this literature by emphasizing the nutrition and health related component of transfer programs as an additional channel through which female fertility outcomes—important for labor market outcomes—can be influenced.[5] We document that timing of exposure to the nutritional and reproductive health program components can affect the age of menarche and subsequent fertility outcomes. One reason this is an important evidence gap to fill is that the medical and nutritional literatures indicate exposure to transfers (and their likely effects on household spending and food availability) at different ages can have implications for the onset of puberty and therefore affect later fertility and other outcomes. Furthermore, research indicates that biological sexual maturation may be particularly sensitive to nutritional shocks for young adolescent girls who experienced poor nutritional status earlier in life (INSERM 2007), as was almost certainly the case for our study population.[6] Consequently, a fuller understanding of the potential differential effects of timing of exposure to the nutrition components, which are common to most large CCT programs but have received relatively little attention in comparison to exposure to the education components, is critical for optimal policy design.

II. The Nicaraguan CCT Program and Experimental Design

The *Red de Protección Social* was a CCT implemented by the Nicaraguan government to address both current and future poverty among rural households. Transfers were made every two months to female caregivers and averaged 18% of total pre-program household expenditures. The transfers came with formal enforced conditionalities alongside a strong social marketing message that the money was meant for food, health and education expenditures. A household

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

[6] See Appendix H for a review of the relevant medical and nutrition literature.

representative signed an agreement they would comply with the conditions and spend the money as intended, although actual expenditures were not monitored or part of the formal conditions.

Like most first-generation CCTs, there were two core components: health and education.

The first component aimed to improve food security, nutrition and health via a universal cash transfer to all households, regardless of size or composition. This nutrition and health transfer was paid bimonthly and conditional on preventive healthcare visits for any children under five years old and on attendance by the caregiver at monthly health education workshops. Local supply of the required health services was increased in parallel with the program.

The second component aimed to improve education. All households with a child 7–13 years old who had not completed fourth grade were eligible. The 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 attendance of all eligible children. For each eligible child, the household also received a small annual cash transfer intended for school supplies. We refer to the combined school attendance and school supplies transfers as the education transfer.

Although modeled largely after Mexico's *PROGRESA*, there were two important differences in the Nicaraguan CCT. First, in Nicaragua beneficiary households were only eligible to receive the program for a fixed three-year period, after which it was not possible to renew. Second, the Nicaragua CCT was focused on early dropout and limited eligibility for education transfers to the first four grades of primary. Household decisions on timing of child schooling and investments in human capital (and any consequent effects), therefore, might differ compared to decisions made under a program like *PROGRESA* with continuing eligibility and without grade limits.

A 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 extreme poverty and relatively low health and educational outcomes. In these municipalities, 42 of 59 rural localities were selected based on a locality-level marginality index.[7] A program registry census of all households in the 42 localities was done in May 2000 prior to randomization. Localities were ordered by the marginality index and divided into seven strata of six localities each, and randomization by strata done publicly in July 2000 with 21 localities randomized to the early treatment group and 21 to late treatment. Households in early treatment

[7] Localities were defined as census *comarcas* from the 1995 Nicaraguan national census. These were administrative areas including as many as 10 small villages and averaging approximately 250 households per locality.

localities received their first transfers in November 2000 and were eligible for up to three years of transfers, with the last delivered in late 2003. After that they received no further transfers, nor were they subject to any conditions. After randomization, households in late treatment were informed that the program would start later in their localities; they were incorporated in January 2003 and likewise eligible to receive up to three years of transfers. Prior notification introduces the possibility of anticipation effects (Section III). In December 2005, the program ended.[8]

Previous short-term analysis demonstrates the randomized sample was balanced at baseline (Maluccio and Flores 2005). Appendix Table A1 demonstrates balance for the samples in this paper. Compliance with the experimental design was high; ~85% of households took up at least one component of the program and there was negligible contamination of the late treatment.

Apart from timing, there were some differences between the early and late treatment program potentially important for our study. First, scheduled transfer amounts were modestly lower for late treatment households, approximately 88% of the early treatment amount. New services were added, however, arguably offsetting this reduction in transfer amounts. Starting in 2003, when the late treatment group became eligible for the program, modern contraception and prenatal consultations were made available to all beneficiaries through the healthcare providers. In addition, as part of the conditions for the nutrition and health transfer, adolescents were now required to attend workshops covering among other things sexual and reproductive health.[9] The workshops were available to both groups, but after 2003 attendance was a condition only for the late treatment. Correspondingly, these services were less intensively implemented and taken up in early treatment localities. Apart from these differences, the eligibility criteria and principal services and conditions were the same across the early and late treatment groups (Appendix C).

III. Conceptual Framework and Identification

A. Conceptual Framework

To guide interpretation of the long-term effects of the differential timing of the CCT we outline a multi-period model of human capital formation beginning in late childhood building on

[8] Nationally, all Nicaraguan government CCT programs ended in 2006 having benefitted ~30,000 households from 2000–6. As such, general equilibrium effects on labor or other markets are likely to have been limited.

[9] Comparable topics were covered with adult caregivers in the health workshops required throughout the program in both treatment groups. Most first generation CCTs included similar workshops (Fiszbein and Schady 2009).

Cunha and Heckman (2007) and Attanasio et al. (2020).[10] Reflecting the main program components, we posit two dimensions of human capital important for labor market outcomes: education and health. The model has three periods. The first two correspond to program exposures in the early (2000–3) and late (2003–5) experimental treatment groups, and the third corresponds to the years after the program had ended until our final measurement in the 2010 long-term evaluation survey (2006–10). The evolution of education and health human capital are characterized using dynamic production functions:

$$\theta_{t+1}^{e} = f_t^{e}(\theta_t^{e}, \theta_t^{h}, I_t, Z_1, a_1) \quad \text{(1a)}$$

$$\theta_{t+1}^{h} = f_t^{h}(\theta_t^{e}, \theta_t^{h}, I_t, Z_1, a_1) \quad \text{(1b) with } t = 1, 2, 3$$

Individuals are first observed in period 1 at age a_1 with previously accumulated education (θ_1^e) and health (θ_1^h) human capital levels and background characteristics (Z_1) such as sex. Public and private investments in each period *t* are represented by $I_t = (I_t^{pub}, I_t^{priv})$.

Under conditions outlined in Cunha and Heckman (2007), the dynamic framework points to the potential for self-productivity, dynamic complementarity and sensitive periods in human capital production. Self-productivity refers to the possibility that for a given level of investment higher levels of human capital in one period give rise to higher human capital in the next, both within and across types, i.e., $\partial\theta_{t+1}^{j}/\partial\theta_t^{k} \equiv \partial f_t^{j}(\theta_t^{e}, \theta_t^{h}, I_t, Z_1, a_1)/\partial\theta_t^{k} > 0$ for $j, k \in \{e, h\}$. Dynamic complementarity refers to the possibility that the marginal productivity of investment increases with human capital levels in period *t*, i.e., $\partial^2\theta_{t+1}^{j}/\partial\theta_t^{k}\partial I_t \equiv \partial^2 f_t^{j}(\theta_t^{e}, \theta_t^{h}, I_t, Z_1, a_1)/\partial\theta_t^{k}\partial I_t > 0$. Last, sensitive periods refer to the possibility that a given level of investment has a higher return in some periods. $t*$ is a sensitive period if $\partial\theta_{t+1}^{j}/\partial I_{t*} > \partial\theta_{t+1}^{j}/\partial I_s$ holds for periods $s \neq t*$. The relevance and strength of all three phenomena likely differ by human capital type, period and sex.

The CCT operates in early treatment localities in period 1, late treatment localities in period 2 and no longer operates in period 3. Therefore, it experimentally alters the timing of program-related public investment (I_t^{pub}) from period 1 to period 2. Changes in public investment directly affect human capital but can also lead to changes in private investment (I_t^{priv}) as households comply with both hard and soft program conditions and optimize over time. For example, early

[10] Although there are some important exceptions (including Cunha et al. 2010 and Carneiro et al. 2021), the late childhood period we begin with is less well studied in the dynamic human capital production function literature.

treatment households who experience increased public investment in period 1 can increase their private investment in period 2 to benefit from potential dynamic complementarities. Conversely, advance notification of future (period 2) public investment for late treatment households can lead to anticipation effects in the form of greater period 1 private investment by them to benefit from the potential dynamic complementarities when the public investment arrives. The extent of the private investment response also depends on the degree of self-productivity and relative sensitivity of investment in periods 1 and 2, as well as the resources available to households.

These program-influenced investment paths, different for early versus late treatment, go on to affect the evolution and final levels of human capital. Notably, the existence of sensitive periods means that even with positive self-productivity and dynamic complementarities, earlier investment is not necessarily more productive in the longer term, for example if later investment is made in more sensitive periods. Hence, the conceptual model alone cannot be used to determine whether period 1 or 2 public investment is more beneficial for human capital production, making long-term differentials resulting from early versus late treatment theoretically ambiguous. In Section III.B we draw on program design, pre-program human capital (in particular school enrollment patterns and early childhood nutritional status) and the medical literature on sexual maturation for girls to explore which periods might be more sensitive to investments in education and health. The randomized timing of the CCT together with the fixed 3-year exposure window offer an opportunity to examine how different investment shocks during distinct late childhood periods with potentially different sensitivities can affect human capital. Other CCT programs with the possibility of renewed or ongoing eligibility would likely lead to different results given dynamic complementarities and self-productivity.

Last, regardless of which treatment has larger impacts on human capital, improvements in either education or health in periods 1 and 2 can affect labor market outcomes in period 3. Improvements in education and health could both improve labor market outcomes but it is also possible that they could have offsetting effects on labor market outcomes, for example if improved health leads to earlier fertility that in turn lowers labor force participation.

B. Identification of 2010 Long-term Experimental Effects of the Differential Timing of Exposure and Cohorts of Interest

The randomized timing of program implementation generated variation in the ages at which children were exposed to the CCT. Because early and late experimental treatment groups both received the CCT and there is no pure control group, we can estimate differentials rather than absolute program effects. For outcomes in which early and late treatments both have positive absolute effects, the differential will be smaller than the absolute effect in the early treatment. This includes the possibility of similarly sized absolute effects in both treatment groups that offset one another and result in no net differential.

Following the original experimental design, we first examine children 7–12 years old at the start of the program in November 2000, i.e., at baseline. The 7–12 cohort is the largest cohort directly exposed to the education component from the start in early treatment for whom there is detailed survey information in 2010.[11] We then limit to 9–12-year-olds at baseline to focus on periods we expect to be more sensitive to investment in education or health, and consequently for which larger long-term differentials in labor market outcomes might be observed. Limiting the age range to the 9–12 cohort leads to larger potential differences between treatment groups in direct exposure to the education component. This is both because it excludes the younger 7–8 cohort who were potentially exposed to the full three years of the program in either early or late treatment and because older children in the 9–12 cohort would have reached 13 and therefore aged out of eligibility for the education component in late treatment before the CCT became available to them.[12] The more limited 9–12 range also reduces overlap in ages exposed across the early and late treatment groups, sharpening the contrast between the experimental groups.

To understand which period was potentially more sensitive for investment in education in this context, we examine schooling patterns. Investment at ages when children are already enrolled in school and likely to continue, or after they have already been out of school for some time, will likely be less effective than investment at ages when children are still enrolled but at high risk of dropping out (de Janvry and Sadoulet 2006). Baseline enrollment rates in program

[11] We do not include 13-year-olds in the analysis because they were not tracked beyond their original households nor administered the individual-level survey due to budgetary constraints and their limited (at most one year) potential direct exposure to the CCT education component.

[12] Appendix Table C1 summarizes differential exposure by age for study cohorts. Child age mattered for eligibility for the education component but did not matter for exposure to the nutrition and health component for which all households were eligible. Therefore, the CCT could have affected the education of children outside the explicitly targeted ages or grades for the education component if they benefited directly from the nutrition and health transfer. They also could have benefited indirectly from education transfers made to other eligible children in the household.

areas indicate that the risk of school dropout increased sharply for boys starting at about age 11 and for girls at age 13. Boys in the 9–12 cohort in the late treatment group had turned 11–14 years old by 2003 when the CCT began for them and were thus more likely to have already dropped out, with the consequence that for them the CCT might have started too late to affect their schooling. By similar reasoning, the CCT also might have started too late to substantially benefit late-treatment girls 11–12 years old at baseline, as they would have reached 13–14 by 2003. Combined with the age-related program eligibility rules, the dropout patterns make it likely that the early treatment period (or period 1) was more sensitive to education investment, although to a lesser extent for girls because of their older typical dropout ages. Given that the early treatment group benefited from the public investment in period 1, all else equal we expect positive long-term differentials on education human capital, especially for boys.

To understand which period was potentially more sensitive for investment in health, we consider the implications of the CCT providing resources intended for food via the nutrition and health component; children in the 9–12 cohort in beneficiary households likely experienced improved nutrition during the program. Although important for male development, nutrition at these ages is particularly relevant for girls because of its potential effect on sexual maturation.

The nutrition and medical literatures demonstrate that investments in nutrition for pre-pubescent girls can improve health and accelerate sexual maturation (Garn 1987; Cooper et al. 1996; INSERM 2007). Moreover, nutritional shocks (both positive and negative) prior to puberty can have immediate and especially large effects for the onset of puberty and age of menarche for girls who had previously experienced early childhood undernutrition (Mul et al. 2002; Parent et al. 2003; Gluckman and Hanson 2006),[13,14] as was likely in the study population.[15] Therefore, in addition to affecting nutritional status directly (for example as measured by body mass index or

[13] Poor early life and later childhood nutrition and health are associated with delayed menarche, and better nutrition and health with early menarche. Medical research shows menarche is related to a minimum body fat mass, and that nutritional status in childhood affects body fat mass and menarche through the leptin hormone which helps regulate energy balance. Leptin levels fluctuate with nutritional intakes pointing to a mechanism by which positive (negative) nutrition shocks can translate into relatively immediate acceleration (delay) of menarche (Blum et al 1997).

[14] In a dynamic human capital production function starting in early childhood the relationship between early life deprivation and greater sensitivity to investment prior to menarche could be represented as a dynamic substitution. Because our conceptual framework (and data) begins in late childhood, we do not explicitly incorporate this complexity. What is most pertinent in our study, with girls having on average experienced poor childhood nutrition, is the potentially strong and immediate sensitivity of sexual maturation to nutritional shocks in early adolescence.

[15] Early childhood nutritional status for the sample is unavailable, but ~50% were stunted (height-for-age z-score < -2) as young women in 2010 and ~40% of girls < 3 years old in 2000 from the same localities were stunted.

BMI), nutritional investment patterns resulting from the CCT can affect the timing of female sexual maturation, with the interval just prior to puberty likely more sensitive to investment.

Median age of menarche in the sample was 13, pointing to period 2 as the more sensitive period to investment in health and the nutrition shocks for the following reasons. In 2000, a positive nutrition shock may have accelerated menarche in early treatment for some of the older girls in the 9–12 cohort but would have come too early to affect menarche for the majority of them. In 2003, the cessation of transfers for girls in the early treatment group (when they were 12–15 years old) likely meant a negative nutrition shock occurring during the more sensitive ages for the younger girls in the cohort, possibly delaying their menarche. Girls in the 9–12 cohort in late treatment, on the other hand, benefitted from the CCT starting in 2003 when they were 11–14 years old, including many 11–13-year-olds who had likely not reached menarche. This positive nutrition shock in the late treatment group at the beginning of period 2 could have accelerated the onset of menarche for them. Hence, for the girls in the 9–12 cohort overall, the start of the late treatment (period 2) was characterized by opposing nutrition investment shocks at a sensitive period: negative for early treatment, positive for late treatment. Given that the late treatment group benefited from the public investment in period 2, all else equal we expect earlier sexual maturity for late treatment girls (along with earlier fertility and marriage), and therefore negative estimated long-term differentials on these health-related outcomes. Based on the ages outlined above we also hypothesize negative effects would be concentrated among the youngest girls in the 9–12 cohort, 9–10-year-olds at baseline.

The above discussion outlines our expectations regarding sensitive periods for education (period 1) and health (period 2, for girls) investment. The conceptual framework (Section III.A), however, makes clear that final outcomes also reflect self-productivity and dynamic complementarities. For example, strong self-productivity and dynamic complementarities following early public investment in education in period 1 can increase education human capital for those in early treatment and the eventual observed differential between early and late treatment. There also can be later dynamic complementarities in period 2 related to possible synergies between education and health. At the same time, because it was known from the outset that the program would expand to the late treatment, the possibility of period 2 self-productivity and dynamic complementarities means there were incentives for households in late treatment to increase investment in period 1, before the program even began in their localities. This could lead to smaller long-term differentials relative to a context without such possibilities.

The investments in both education and health in periods 1 and 2 ultimately can affect labor market outcomes in period 3, and those effects can differ by sex. For both sexes greater program investment during education-sensitive period 1 could yield positive differentials in labor market outcomes in period 3, through positive differentials in schooling, learning and socio-emotional skills. For women, greater program investment in period 2 nutrition could lead to better nutritional status around the time of adolescent sexual maturation, resulting in earlier sexual maturity and possibly earlier fertility in period 3. Furthermore, the introduction of the adolescent health workshops and contraceptives in period 2 could have reinforced any possible physical sexual maturity-related behavioral changes including earlier sexual activity, potentially also leading to earlier fertility. Since both the nutritional and behavioral pathways described here are likely strongest in period 2, and we estimate program differentials for early minus late treatment, we would expect negative differential effects on fertility in period 3. Because higher fertility is negatively associated with most labor market outcomes for women, the possible fertility-channel related effects on labor market differentials could reinforce any positive education-channel effects on labor market differentials between early and late treatments.

IV. Data

A. Sources

Program Registry Census and Short-term Evaluation Surveys—A May 2000 program registry census done in all 42 localities provides baseline data on schooling, demographics and assets. Using the program census a random sample of 42 households in each of the 42 localities was drawn for the short-term evaluation; because locality sizes differed, we construct short-term evaluation sample weights to provide population estimates of the study area. A comprehensive household baseline survey was conducted in September 2000, with subsequent panel follow-ups in 2002 and 2004. Attrition was ~10% per round (Maluccio and Flores 2005).

2010 Long-term Evaluation Survey—The 2010 survey was conducted 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 evaluation sample plus, to increase power, a random oversample of households with children who were 11 years old in 2000 (specifically, born between January and June 1989). This age group was selected because they were at relatively high risk of dropping out of school in the years immediately following the start of the program, i.e., at ages when education might be particularly sensitive to investment (Section III.B). All long-term estimates account for the short-term evaluation survey sampling methodology and subsequent oversample using long-term evaluation survey sample weights constructed to provide population estimates of the study area. The main cohort of interest, children 9–12 years old at baseline in 2000 (Section III.B), has 2,200 individuals (1,138 males and 1,062 females).

The 2010 survey included separate household- and individual-level instruments. The household survey instrument collected both household- and individual-level information from the best-informed person available for the interview, generally the young adults themselves or the household head or spouse. It included individual schooling outcomes and detailed participation and earnings measures for all economic activities over the last 12 months. A complete labor market history including participation, location and earnings for all non-agricultural self-employment and wage jobs ever held was also collected. Location of activities was collected to distinguish between work performed while residing in the current home community versus work performed during periods of temporary migration. Given the mobility of this young population and temporary nature of many economic activities, such a comprehensive approach was key to characterizing labor market returns.

The individual survey instrument was administered to all individuals born after January 1, 1988. Data was collected through direct interviews of the young adults in their homes and designed to measure individual learning, cognition, socio-emotional, marriage and reproductive health outcomes. Weight and height were collected for females, but not for males.[16] To measure learning we implemented two math and three Spanish language achievement tests. The math tests included numerical fluency and a test of problem solving similar to the Peabody Individual Achievement Test (Markwardt 1989). Spanish tests included word identification, spelling and reading fluency. In addition, we administered two tests that capture both learning and cognitive development: the Spanish version of the Peabody Picture Vocabulary Test (PPVT) (Dunn et al. 1986) assessing receptive vocabulary and a forward and backward digit span test assessing memory. To measure cognition, we administered Raven's colored matrices (Raven et al. 1984).

[16] Anthropometry for males was not collected due to sensitivities related to potential military service in Nicaragua.

The Raven was included to help capture general non-verbal cognitive skills as opposed to the more specific skills likely acquired in the classroom. A strength of all the tests is they provide observed, as opposed to self-reported, measures of learning and cognition regardless of schooling status, substantially reducing concerns about social desirability or selectivity biases. Appendix D provides further detail on test administration protocols. We measured socio-emotional outcomes using the Center for Epidemiologic Studies of Depression (CESD) Scale and the Strengths and Difficulties Questionnaire (SDQ). Lastly, the individual-level instrument included a retrospective sexual and reproductive health module collecting information on prior attendance at and learning in the adolescent health education workshops, age of menarche for women, age of first sexual activity, and fertility and marriage. The module was developed based on qualitative interviews done in preparation for the 2010 survey that revealed perceptions of earlier sexual maturity and increased early sexual activity among teens during the program, the latter a possible unintended consequence related to provision of information and contraception (CIERUNIC 2009).

Nicaraguan Population Censuses—Finally, we also use the 1995 and 2005 Nicaraguan national censuses with individual-level residential, schooling, demographic and migration data.

B. Attrition, Internal Validity and Weighting

Due to extensive national and international tracking, attrition in the 2010 long-term evaluation survey is on par with or lower than in related longitudinal studies of similar populations and time horizons (Appendix E). Individuals not found at their original residences were tracked to new locations throughout Nicaragua and Costa Rica (the destination of 95% of the international migrants). Only 10% of men and 16% of women in the main cohort of interest, 9–12-year-olds at baseline, could not be tracked to their 2010 household location and are therefore missing the labor market outcomes and detailed schooling and marital status information collected in the household-level instrument. Attrition is higher, 19% for men and 22% for women, for the individual-level instrument which included the in-person tests. For individuals who could not be located we collected some schooling, demographic, migration and labor market information through proxy reports from other original household members, reducing the attrition on a set of basic indicators to 4% for men and 6% for women.

There are no significant differences in attrition (or in permanent migration) levels between early and late treatment; all differences are less than |0.015| with p-values greater than 0.6.

Attrition also does not affect the balance of baseline observables (Appendix Table A1). However, attrition is correlated with baseline characteristics associated with migration, and those correlations differ between early and late treatment groups. Consequently, even with balanced attrition sample selectivity remains a potential concern for internal validity. Our preferred estimates account for attrition selection using inverse probability weights (IPW) constructed to allow for differences between early and late treatment groups and incorporating information from the survey tracking process. Individuals who were more difficult to find and interview are given higher weight since they were more similar to individuals not interviewed (Molina-Millán and Macours 2017). Separate IPW weights are calculated by sex for outcomes from household- and individual-level survey instruments, respectively, and all weights also incorporate the long-term evaluation survey sample weights to provide population estimates of the study area. Appendix E presents additional information on tracking, attrition and IPW construction.

C. 2010 Labor Market and Human Capital Families of Outcomes

We examine a wide range of labor market and human capital-related outcomes in 2010, organizing specific measures into different domains, or families of outcomes. Each family is constructed from the average of the z-scores of the individual components of that family.

To characterize the labor market for this predominantly rural, young adult population, we construct two families: labor market participation—capturing participation and temporary migration for work; and earnings—capturing labor market returns via earnings for work off the family farm. We consider two different versions of the earnings family to account for the skewed nature of the unconditional distribution of earnings in which many earn nothing (the small fraction not working as well as those only engaged in unremunerated economic activities) and an extended right-side tail: 1) ranks of earnings (Athey and Imbens 2017); and 2) reported earnings trimmed at the top 5% of values. Specific measures included are defined in results Section VI.B.

To explore various possible mechanisms underlying potential labor market effects, we expand upon the two forms of human capital outlined in the conceptual framework and construct schooling, learning, cognition, fertility and marriage, and socio-emotional families.

For schooling outcomes, the family includes whether the respondent was enrolled in school, highest completed grade and whether they had completed fourth grade, after which they were no longer eligible for the education transfer or subject to its conditionalities. To analyze learning and cognition outcomes we categorize the various tests into three families. The first, learning, focuses on the skills most likely acquired in the classroom and comprises the five achievement tests (math fluency, math problems, word identification, spelling and reading). The second, learning-and-cognition, includes the tests we expect are likely to capture both learning and cognition (receptive vocabulary and memory). The third, cognition, contains only the Raven.

For fertility and marriage outcomes, the components included are different for men and women. For both the family includes whether the individual had sex by age 15, had any children and had ever been married (including informal unions). For women, the family also includes age of menarche. The sign for age of menarche is reversed when averaging the z-scores for the family index so that higher values for the family are associated with higher fertility. We also consider an alternative version of the fertility and marriage family for women that includes BMI for non-pregnant women; unlike the other indicators in the family, BMI has the advantage of being objectively measured rather than self-reported. Higher BMI in this population can be indicative of better nutritional status, as well as earlier onset of puberty or prior childbearing.

For the socio-emotional outcomes, we conducted exploratory factor analysis combining all items from the CESD and SDQ (Appendix F).[17] The analysis reveals four latent factors we interpret as broadly capturing optimism, positive- and negative self-perception, and stress. The signs for negative self-perception and stress are reversed when averaging z-scores for the family index, so that higher values for the family indicate more positive socio-emotional outcomes.

V. Methodology

A. Empirical Specification for 2010 Long-term Experimental Differentials

We estimate intent-to-treat (ITT) effects of the randomized timing of exposure to the CCT using the following linear specification:[18]

$$Y_{il} = \alpha T_l + \boldsymbol{\beta X}_{il} + \varepsilon_{il}, \qquad (2)$$

[17] We use exploratory factor analysis because the correlations among items in the CESD and the SDQ suggested standardized scoring was unlikely to reflect the intended latent traits, similar to Laajaj and Macours (2021).

[18] Because the intervention and short-term evaluation began in 2000 and explicit plans for longer-term follow-up were not made until after the program had ended, there is no formal pre-analysis plan. Analyses adhere to the original experimental design and examine hypotheses outlined in proposals developed for the long-term follow-up.

where Y_{il} is the 2010 (i.e., period 3) outcome for individual i in baseline locality l. T_l is an ITT indicator equal to one for localities randomly assigned to early treatment and zero for late treatment. It measures the impact of early versus late exposure to the entire CCT package. Analyses use all respondents from both treatment groups in the age cohort being considered, regardless of initial completed schooling or actual program participation. All regressions also include strata fixed effects to account for randomization stratified by the locality-level marginality index. Given the randomized assignment, our main specification limits the other control variables in $\boldsymbol{X}_{il}$ to age of the individual at the start of the program in November 2000 using indicators for 3-month age groups, for whether they had 0, 1, 2, 3 or 4+ grades completed at baseline and regional fixed effects. Regressions are weighted to account for sampling and attrition providing population estimates of the study area. We assess robustness of the main specification to alternative controls, samples and approaches for addressing attrition selection in Section VI.B. The latter include estimates: 1) without IPW (using only the long-term evaluation survey sample weights with no correction for attrition); 2) with an alternative IPW that does not incorporate tracking information; 3) of Lee bounds; 4) of Kling and Liebman sensitivity bounds; and 5) of effects on a set of basic indicators using proxy reports for which attrition is lowest.

To reduce concerns related to multiple hypotheses testing, we organized individual variables into families as described above, and construct an index for each by averaging the z-scores of the specific components of the family (Kling et al. 2007). We assess robustness of this approach by also constructing alternative indices using inverse covariance weighting, which assigns less weight to individual indicators that are highly correlated within the families. Standard errors are adjusted for clustering at the locality level. We adjust for multiple hypotheses testing of all families at the same time using Anderson's (2008) familywise error rate. We also assess whether accidental imbalance related to the limited number of clusters (N=42) changes the significance, using randomization inference and estimating exact p-values under the sharp null hypothesis that the treatment effect is zero for all participants (Athey and Imbens 2017; Young 2019).

VI. Results

A. 2010 Long-term Experimental Differentials: Pooled Samples

To analyze the effects of the differential timing of exposure we estimate the 2010 long-term experimental differences between the early and late treatment groups, referred to below as the (long-term) differentials. Table 1 presents results for each family pooling men and women. For both the 7–12 (panel A) and 9–12 (panel B) cohorts there are significant positive long-term differentials in labor market participation (0.17–0.20 SD) and in earnings (0.12–0.14 SD). The results demonstrate that earlier exposure to the CCT translated into better labor market outcomes a decade later when individuals were beginning their adult working lives. The differentials in the schooling and learning families are approximately 0.10 SD for the 9–12 cohort, but more modest, 0.06 SD (and insignificant for learning), for the 7–12 cohort. Estimates are close to zero and insignificant for the socio-emotional family.

B. 2010 Long-term Experimental Differentials: 9–12 Cohort by Sex

To shed light on the possible sensitive periods and mechanisms described in the conceptual framework, we focus on the 9–12 cohort by sex (panels C and D). The pooled results mask differences in estimated magnitudes by sex, though only learning is significantly different between them (panel E).[19] Estimated labor market differentials are nearly twice as large for men. Both sexes have similar-sized positive differentials in schooling, but there is a significant learning differential only for men, with an estimated zero differential for women. In contrast, there is a significant negative differential in fertility and marriage for women, but none for men.

To unpack these results, the subsequent tables show the long-term differentials in the components of the outcome families, for each presenting first the men and then the women.

Labor market participation and earnings—As shown in Table 1 (panel C) and reproduced in column 1 of Table 2 (panel A), there is a sizeable 0.27 SD long-term differential in the labor market participation family for the male 9–12 cohort. Off-farm work is 6 percentage points higher (7% of the mean in late treatment) in early treatment. Men in early treatment are 9 percentage points more likely to have migrated temporarily for work in the last 12 months, nearly one-third higher. Possibly reflecting a migration for work strategy, they are also 8 percentage points (one-third) more likely to have ever had a salaried non-agricultural wage job, and 7 percentage points (one-half) more likely to have ever had an urban wage job.[20]

---

[19] Results for the 7–12 cohort by sex are qualitatively similar, but with smaller estimated magnitudes (Appendix Table A2). Results for the younger 7–8 cohort by sex are reported in Table 7 and discussed in Section VI.C.

[20] To provide further context for the findings, we note that nearly all men (98%) are working and most combine work on the family farm (89%) with work off the family farm (83%). In line with low average schooling (5.5 grades), most off-farm employment is unskilled. Men work as agricultural laborers on farms not belonging to the household or on large plantations, in salaried jobs in the non-agricultural sector (e.g., as construction workers or

In Table 3, we investigate whether differential labor market participation for men is accompanied by differential earnings.[21] To avoid selectivity bias all analyses are unconditional, with zero earnings for each component element when the individual did not report that type of earnings in the reference period. Results are broadly consistent across the various indicators, with an overall differential of about 0.2 SD for both rank (panel A) and 5% trim (panel B) versions of the earnings family. In addition, density functions in Appendix Figure A1 and quantile regressions in Appendix Table A3 demonstrate there are positive differentials across the earnings distribution. Because virtually all men (98%) are working, differential earnings do not stem from increased participation on the extensive margin, but rather from shifts in temporary migration for work and in the mix of economic activities.

Women in the 9–12 cohort exposed to the CCT program earlier also do markedly better in the labor market in 2010. There is a 0.17 SD differential in the labor market participation family (Table 2, panel B) driven by higher off-farm work (7 percentage points or 15%) and a doubling of temporary migration for work (9 percentage points). Correspondingly, there is a marginally significant differential in earnings of about 0.1 SD for both earnings families (Table 3, panels C and D). As was the case for the men, the findings suggest differentials in earnings for women stem from shifts in temporary migration for work and in the mix of economic activities.

Earlier exposure to the CCT led, after 10 years, to better labor market outcomes for both men and women. Results by sex are qualitatively similar but estimated magnitudes are larger for men than for women. One possible explanation for more modest differentials for the women is that less than half had earnings from off-farm labor and analyses are unconditional. Examination of

security guards), and in non-agricultural self-employment. In the poor, remote rural communities where the CCT operated, however, opportunities for even these types of low-skilled employment are often limited. For this reason, we incorporated a measure of seasonal or temporary migration for work in the labor market participation family.

[21] We use several measures to capture earnings from different activities (agricultural or non-agricultural and salaried or self-employment) that are often seasonal or temporary (Table 3). We do not, however, include earnings from the family farm since person-specific individual returns could not be reliably quantified in this context. Only 11% of the cohort of young men are head of their own household so that nearly all work on the farm of an older household member, typically their father or father-in-law. Moreover, we did not collect detailed information on agricultural inputs, precluding calculation of earnings for household farms. Although in theory the positive differentials for off-farm earnings we estimate could be offset by lower on-farm earnings, patterns of participation make it unlikely. First, the estimated differential in participation in on-farm work is small and positive, indicating that higher off-farm work did not substitute for on-farm work on the extensive margin. Second, there is no differential in the total number of months worked off-farm; average months working off-farm during temporary migration increases while average months working off-farm (while residing in the current home community) decreases. These patterns suggest minimal scope for crowding out on-farm work on the intensive margin.

the human capital-related outcomes below suggests that sex differences in other mechanisms also may help explain sex differences in labor market outcomes for the young adults.

Schooling, learning and cognition—There is a ~0.1 SD long-term differential in the schooling family for men (Table 4, panel A), including nearly 0.3 more grades completed (5%). Early treatment men are also 5 percentage points (25%) more likely to still be in school. Overall, a substantial minority (18%) in the late treatment group are still studying in 2010 when they were 19–22 years old, with more than 80% of them in secondary school where they typically enroll in weekend programs that enable them also to work. Average completed grades for early treatment men remain low (5.8), however, and only 78% have completed fourth grade.

Higher schooling for early treatment men is accompanied by higher learning (0.18 SD). In particular, there are significant long-term differentials in exactly the sorts of skills taught in school: 0.16 SD for math and 0.20 SD for Spanish. The findings for Spanish are corroborated by a significant differential on self-reported literacy (reported in column 5 though not included in the learning family which comprises only direct tests). Magnitudes of the differentials are sizable and in line with absolute impacts of education interventions in other similar settings (Evans and Yuan 2022). Together, the results for schooling and learning demonstrate that men exposed to the CCT earlier not only have higher completed grades and do better on achievement tests, but also are more likely to still be studying, suggesting observed positive education differentials may have continued to grow even after 2010.

Differentials are smaller and insignificant for the mixed learning-and-cognition family that arguably captures both classroom and non-classroom skills. The estimate for the Raven, a cognitive test less likely to reflect skills directly acquired in the classroom, is close to zero (-0.02 SD). The lack of differentials for cognition-related measures is as expected given that the intervention for the 9–12 cohort began in late childhood, well after what are often regarded as the most sensitive ages for investment in cognitive development.[22]

For women there is a positive long-term differential in the schooling family the same size as for men, about 0.1 SD (panel B). The differential between early and late treatment in grades completed is more modest (0.18 or 3%) and not significant (though there is a significant

[22] It is for this reason that the cognition-related families are not included in Table 1. In other work we examine effects on these measures for cohorts exposed to the CCT at younger ages (Barham et al. 2013). The distinction between achievement and cognitive tests was outlined in proposals developed for the 2010 long-term evaluation.

differential in having completed fourth grade), possibly related to lower potential sensitivity to education investment for females in period 1 due to later typical ages of dropout. Indeed, even in 2010, 30% of the young women are still in school, with two-thirds in secondary school and one-third studying at the tertiary level.

In contrast to the men, there are no significant differentials in learning for women. Point estimates are close to zero for the various families (panel B), as well as for the underlying individual tests (Appendix Table A4) and for self-reported literacy (panel B). The lack of a differential could indicate that the CCT did not improve learning for women in either treatment group but is also consistent with positive similarly sized absolute program effects for each treatment group. We provide short-term experimental and medium-term non-experimental evidence supporting the latter possibility in Sections VI.D and VI.E below.

The long-term differentials in schooling and learning are consistent with the hypothesis that period 1 was more sensitive for education investment in the 9–12 cohort, especially for boys.

Fertility and marriage—There is no overall significant differential in the fertility and marriage family for men (Table 5, panel A).[23] Individual components offset one another with a positive differential for ever having sex before age 15, but a negative one for ever married.

For women on the other hand (panel B), there is a significant negative differential of 0.17 SD for both the family with and without BMI (columns 1 and 2). There is, notably, a significant differential in the age of menarche, with the late treatment group having reached sexual maturity three months before the early treatment group. Correspondingly, there is a large negative differential on measured BMI, 0.7 kg/m^2 lower in early treatment.[24] The early treatment group is 11 percentage points less likely to have had sex by the time they were 15 years old, one-third lower than the average in the late treatment. This large differential is mirrored by a 6-percentage point (10%) lower probability of having a child by 2010 (p-value 0.17).[25] Overall the results for the fertility and marriage family for women are in line with greater period 2 sensitivity to

[23] Because anthropometrics were not collected for men, we are unable to evaluate nutritional mechanisms related to health human capital for men to the same extent as for women.

[24] It is possible that recall error affects reports of age of menarche, though there is no obvious reason such measurement error would affect the early and late treatment groups differently. Results on BMI for non-pregnant women, objectively measured and therefore unlikely to suffer from differential measurement error, are significant and in line with the other reproductive health outcomes including age of menarche. The consistency in findings across outcomes suggest systematic recall measurement error on age of menarche is not driving the fertility results.

[25] Analysis of the short-term evaluation data provides evidence consistent with this, as an estimated negative 3-percentage point differential in fertility had already emerged by 2004 when the girls were still only 13–16 years old.

investment and the nutrition shocks (resulting from the end of the program in early treatment and the start in the late treatment) operating through the age of menarche. In Section V.E we discuss non-experimental evidence that this negative differential in fertility does not appear to be driven by an absolute increase in fertility in late treatment.

As outlined in Section III.B, given the median age of menarche, we expect the period 2 sensitivity to be larger for the younger girls in the cohort and therefore examine the fertility and marriage family and its components separately for girls 9–10 and 11–12 years old at baseline in Appendix Table A5. The differential for the family for the 9–10 cohort is indeed larger (-0.24 SD, significant at 1%) than for the 11–12 cohort (-0.10 SD, insignificant). Moreover, although not statistically different, larger differentials are observed for the younger cohort in all underlying components; point estimates for age of menarche indicate the younger girls in the late treatment group reached sexual maturity four months before the early treatment, compared to two months earlier for the older girls. This pattern of results for age of menarche and fertility further supports the possibility of greater sensitivity to nutritional investment and its effects on sexual maturation in period 2 and suggests nutrition is playing an important role. In Section VI.D we use the short-term evaluation surveys to document large positive impacts on food availability and quality providing evidence supporting such program-induced nutritional gains.

In columns 8–9 of Table 5 we examine outcomes related to the adolescent health workshops for the 9–12 cohort. As expected, they demonstrate that early treatment girls were less likely to have attended the workshops which began in 2003, corroborated by their being less likely to know about a key health topic emphasized in them, the Pap smear test. If the workshops (or contraceptive provision) had the unintended positive effects on sexual activity suggested by the qualitative research, differential exposure to them could have exacerbated potential negative differentials on fertility induced by the nutritional shocks and physical maturity described above.

To investigate this possibility, we consider an alternative comparison, contrasting girls 9–12 years old in 2000 (our main 9–12 cohort) in early treatment with a younger cohort of girls 9–12 years old in 2003 in late treatment. This provides a partial test of whether differences (including the workshops or the modestly lower transfer amounts) between the early and late treatment programs drive the differential outcomes in 2010. Appendix Table A6 demonstrates there are only small and insignificant differences between the two groups for age of menarche and whether they had had sex by age 15.[26] Moreover, there are no differences in having completed fourth grade for either sex. These results are consistent with the 2010 long-term differentials in fertility coming from the timing of the nutrition shocks during a sensitive period for sexual maturation in period 2, rather than from any differences in the structure of the program between early and late treatment. Nevertheless, with the available data we cannot rule out the possibility of some differential behavioral effects related to the workshops at older ages.

Socio-emotional Outcomes—Socio-emotional outcomes represent another potential important pathway through which the program could have influenced long-term labor market outcomes. The teenage years have been hypothesized to be a period sensitive to investment in socio-emotional (or non-cognitive) skill formation (Cunha and Heckman 2007), so that differences in timing of program exposure could well lead to differentials in such skills. Interpretation as a possible mechanism for the labor market results is arguably less straightforward than for education or fertility, however, since it is plausible that the causality runs in the opposite direction as well, with labor market outcomes affecting current socio-emotional measures.

Table 6 shows small and insignificant long-term differentials in the socio-emotional outcome families for men and women. Examining the each of the factors separately, however, reveals that for men the average masks offsetting differentials in the various latent traits (panel A). Specifically, men in the early treatment group are more optimistic and have a more positive self-perception (~0.25 SD), possibly reflecting their higher learning and earnings. Meanwhile, early treatment men also are more likely to exhibit greater stress and agree with statements reflecting negative self-perception, although point estimates for these traits are smaller (~0.16 SD). This pattern of offsetting effects echoes Fernald et al. (2008) who find that increased economic opportunities improved some aspects of mental health but at the same time increased stress. For women (panel B), there is also a significant differential for negative self-perception. Overall, the findings do not provide clear evidence that the labor market differentials result from, or for that matter drive, differentials in socio-emotional outcomes for either sex.

Robustness—The main results for the 9–12 cohort by sex for the different families of outcomes are robust to adjustments for multiple hypotheses testing (using Anderson's familywise error rate) and randomization inference (Table 1, panels C and D). For the latter, we

[26] Because this is a comparison of girls who across treatment groups have different ages when measured in 2010, we do not compare age dependent outcomes such as ever married or number of children.

can reject the sharp null hypothesis that all treatment effects are zero for all the outcome families by sex that were significant using conventional standard errors clustered at the locality level. Appendix B also demonstrates that the results are robust to: 1) alternative weights and samples; 2) different assumptions related to attrition, confirming that the main findings are not driven by IPW reweighting; and 3) a family index constructed using inverse covariance weighting.

Finally, the logic that timing of early versus late treatment can affect some age groups differently means it may be possible to learn more about the mechanisms by examining differentials for narrower age cohorts, as done for the female fertility and marriage family above. This strategy comes at the cost of reduced power, particularly for younger cohorts for whom there was no targeted oversample. We present the 2010 long-term experimental differentials for the two-year age cohorts in Table 7 (which also includes the younger 7–8 cohort discussed below in Section VI.C). For men, differentials are similar across the 9–10 and 11–12 cohorts, with point estimates a little larger for most outcomes for the 11–12 cohort including a positive significant differential in the socio-emotional family. Results for women are broadly similar across the two-year cohorts, apart from fertility discussed above and a now significant negative differential in the socio-emotional family for the older 11–12 cohort. The differentials in the socio-emotional families for the 11–12 cohorts by sex are in the same direction as the corresponding differentials for the learning families of this cohort, consistent with learning and socio-emotional skills being sensitive in the same periods. The pattern is also observed for the 9–12 cohort (Table 1, panel E).

C. 2010 Long-term Experimental Differentials: 7–8 Cohort by Sex

Based on the age-eligibility cut-offs, children in the 7–8 cohort (i.e., 7–8 years old in 2000) were potentially eligible for the CCT education component in either treatment group throughout program operation (Appendix Table C1). The logic used in Section III.B for identifying education-sensitive periods for investment therefore points to period 2 for this cohort (when they were reaching ages at which school dropout risk rises), although it is less definitive. With less contrast in the potential direct exposure to the education component of the program across the two experimental groups, the differential could depend relatively more on whether possible gains from earlier public investment, or later positive self-productivity and dynamic complementarities in early treatment, dominate gains from any anticipation effects combined with public investment in period 2 in late treatment. Examining investment in health, period 2 is also arguably more sensitive, with most females in the 7–8 cohort too young to reach puberty in period 1 but 12–13 years old by the final year of transfers in late treatment. Because the program ended in 2005, however, it also may be that there would not have been sufficient exposure at ages close to 13 to have had a substantial influence on sexual maturation.

We present the 2010 long-term experimental differentials for the families for the 7–8 cohort by sex in Table 7 (panels A and D). Differentials for both males and females are relatively small, and all are statistically insignificant. The lack of differentials is likely a result of this cohort being exposed to the full three years of the two core program components in both the early and in the late treatment groups as described above. It does not necessarily mean there were no positive absolute effects on these cohorts. We provide short-term experimental and medium-term non-experimental evidence examining this possibility below in Sections VI.D and VI.E.

D. Short-term Experimental Impacts on Schooling and Nutritional Investments in Periods 1 & 2

To help unpack the dynamics underlying the 2010 long-term experimental differentials, we use the short-term evaluation surveys to document program compliance and estimate short-term experimental impacts on schooling and household food availability in 2002 and 2004, corresponding roughly to periods 1 and 2 in the conceptual framework.

School enrollment during the program—Enrollment patterns during program implementation demonstrate that compliance in both treatment groups was high and in line with the timing of the public investment and application of the education component age criteria and conditionalities. As shown in Table 8, in early treatment localities in 2002 (period 1) enrollment was nearly universal for the 7–12 cohort. In late treatment in 2004 (period 2), after the program had been operating there for almost two years, enrollment was similarly close to 100% for the 7–8 cohort, suggesting high compliance for that age-eligible cohort. Enrollment was lower in late treatment for the 9–12 cohort, most of whom had aged out by 2004 (when 13–16 years old).

Short-term experimental effects on schooling—To examine the short-term effects of the public investments during program implementation, we use equation (2) and estimate the experimental differential effects for the 9–12 cohort by sex in 2002 and 2004. In 2002—when early treatment localities had received transfers for two years but late treatment localities had not yet benefited—boys in the 9–12 cohort have 0.36 additional grades (15% more) and are more likely to have completed fourth grade (Table 9, panel A). Moreover, they are 18 percentage points more likely to be enrolled (25%), 36 percentage points more likely to be regularly attending (66%) and 15 percentage points more likely to be literate (20%). Estimated effects in 2002 for girls (panel C) are qualitatively similar and also significant, but smaller. The larger effects for boys in the short-term are consistent with period 1 having been more sensitive to investment in education for them, as hypothesized.

The short-term results for 2002, rather than simply measuring the absolute effect of early treatment compared to a pure experimental control, also could be influenced by anticipatory behavior in the late treatment. In the conceptual framework we argued that with positive dynamic complementarities advance notice of future public investment in period 2 could incentivize households in late treatment to increase investment in period 1. This could lead to smaller period 1 differentials relative to a pure control group. In contrast, however, because of the program eligibility rules it is also conceivable that late treatment households would not enroll or would hold back children so that they would not surpass fourth grade before period 2, particularly if dynamic complementarities are weak. This could lead to larger period 1 differentials. Enrollment and repetition rates for the 9–12 cohort in late treatment show that, if anything, higher earlier investment was more likely. Enrollment for late treatment children was 77% at baseline in 2000 and reached 80% by 2002, increasing for both sexes but with larger increases for boys. It also does not appear that children were systematically being held back to avoid completing fourth grade (thus remaining eligible for transfers longer); repetition rates for the same cohort were 12% in 2000 and 10% in 2002. Therefore, positive differentials in 2002 likely underestimate the absolute program effects.

Between 2002 and 2004 the differential in grades completed (as well as the fraction completing fourth grade) continues to widen, reaching 0.49 grades for boys (panel B) and 0.57 for girls (panel D). Alongside these increases, however, and in line with the program having ended in the early but begun in the late treatment, effects on enrollment and regular attendance in 2004 are reversed. The 2004 negative differentials between early and late treatment for enrollment (14 percentage points or 18%) is significant for girls, suggesting that late treatment children were already beginning to catch up on schooling. Notably, this occurs even though most girls in the late treatment in the 9–12 cohort had already aged out and were no longer themselves directly eligible for the education component or subject to its conditions. The negative differential on enrollment for girls in 2004, therefore, likely stems from other program components that households of late treatment girls were receiving or is related to dynamic production function technology such as synergies between education and health human capital production. These short-term patterns are consistent with expectations about effects from the public investment and point to the potential for considerable catch-up in grades completed for the late treatment girls, indicating that for them more so than for the boys, long-term differentials on schooling could underestimate absolute effects. This is corroborated by the 2010 survey showing that both the early and late treatment groups continued to advance in school after 2004 (when the average was 4.4) reaching 6.8 grades completed for the late treatment group in 2010 (Table 4).

Short-term experimental effects on nutritional investments—Turning to the effects of the public investment on nutrition-related indicators, in Table 10 we use short-term evaluation data to confirm that exposure to the CCT is related to large experimentally induced shocks in the availability of nutritious food in periods 1 and 2 for the cohorts of interest. Both the quantity and the quality of household food consumption changes significantly. In 2002, per capita food consumption expenditure in households with a boy (panel A) or with a girl (panel C) in the 9–12 cohort is ~35% higher in early treatment. Moreover, there is improved nutritional quality of food with higher consumption shares of animal proteins and fruits and vegetables, alongside lower shares of staples or other less nutritious foods.[27] By 2004 (period 2), when the CCT had been operating for more than a year in late treatment but had been phased out in early treatment, the patterns reverse, with 13% or more higher food consumption in late treatment.[28]

E. Non-experimental Differences-in-differences (DD) Absolute Effects

As discussed above, the absence of long-term differentials for some measures (grades completed and learning for the female 9–12 cohort) or some cohorts (both the male and female 7–8 cohort) could result from similar sized absolute effects that cancel one another out in the estimated differential. The short-term experimental evidence presented above (and for the 7–8 cohort in Appendix Table A7) points to this possibility for certain outcomes. In 2002 before the late treatment had begun to receive the program there were positive effects on schooling and

[27] Short-term improvements in quantity and quality of household food consumption are consistent with findings indicating significant reductions in stunting among children in the first years of life (Maluccio and Flores 2005).

[28] Absolute values of the differential effects 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 observed for a related Nicaraguan CCT program (Macours et al. 2012).

learning (measured by literacy) for boys and girls in both age cohorts, but by 2004, after it had begun receiving the program, the late treatment showed signs of starting to catch up.

To help further interpret what underlies the differentials, we explore (less well-identified) non-experimental differences-in-differences (DD) estimates of 5-year absolute effects comparing treatment municipalities to adjacent similarly poor non-treatment municipalities using 1995 and 2005 Nicaraguan population census data. Results point to large absolute increases in schooling outcomes after five years for both boys and girls in the 7–8 and 9–12 cohorts, with effects on grades completed twice as large as the experimental differentials (Appendix Table G1).[29] Together with the 2010 differentials, the findings suggest that late treatment children were largely able to catch up on schooling and learning with the early treatment, possibly as a result of program exposure in late treatment through 2005. Consequently, the experimental differentials may represent lower bounds for the absolute program effects.

Using the national census data and the same DD specification, we also examine early fertility for the 9–12 cohort of girls (14–17 years old at the time of the second census in 2005). Results indicate a 2-percentage point reduction in both marriage and fertility, equivalent to a 25% reduction for early fertility. Finally, in Appendix G we compare the 9–12 cohorts from the study with a same-age cohort of women from the 2011–12 Nicaraguan Demographic and Health Survey. The non-experimental comparison suggests the CCT is associated with an overall increase in BMI alongside a reduction in early fertility for the entire cohort (early and late treatment groups combined), making it unlikely that the negative experimental differential reflects an absolute increase in fertility in the late treatment group.

F. Interpretation and Implications for Long-term Labor Market Results

Although the randomized design does not allow fully disentangling the various mechanisms potentially underlying period 3 labor market outcomes, the sex-specific results for schooling, learning and fertility suggest those mechanisms differ for men and women.

For men, period 1 exposure to the CCT led to more schooling and learning, along with changes in labor market participation and higher earnings. The pattern supports a principal

[29] Appendix G provides the detailed methodology and results for the DD, demonstrates that results are robust to alternative comparison groups, and provides falsification test evidence in support of the common trends assumption.

rationale for CCTs and provides evidence of their potential to reduce the intergenerational transmission of poverty through improved (education) human capital.

Higher earnings for the men, however, were not the result of transformative increases in their schooling or changes toward skilled or formal employment. This is unsurprising considering the short-term nature of the program and that the improvements in basic skills were for a population with initially low levels of education. Findings from qualitative work suggest basic skills attained might have led to higher labor market returns via seasonal or temporary migration. Although such migration has associated costs, there are large wage differentials across regions in Nicaragua, and wages are substantially higher in Costa Rica. Improved math skills may have helped the young men assess the cost-benefit trade-offs of temporary migration, and improved reading comprehension skills may have enabled them to more easily complete required paperwork, particularly relevant for international migration to Costa Rica (CIERUNIC 2009).

For women, period 1 exposure also led to positive differentials in schooling, driven by completing the program-targeted fourth grade, but no significant differentials in total grades completed or learning. The absence of differentials in grades completed or learning is consistent with positive absolute program effects for both the early and late treatment groups (Section V.E).

The negative differential in the fertility and marriage family suggests that program-related nutrition shocks in period 2, along with modest positive differential in schooling described above, may have both contributed to the differential labor market outcomes for women in period 3. A fertility-related mechanism is consistent with the challenges Nicaraguan women face when combining motherhood and the responsibilities of marriage with labor market activities in a context where participation often requires commuting to nearby urban centers or migrating to larger cities or to Costa Rica.

G. Cost-benefit assessment

The positive differential effects on earnings suggest the potential for recouping the costs of the CCT. A full cost-benefit analysis would require monetary valuation of all resource costs and all potential absolute benefits of the program for each individual household member, including those not in the study, and is beyond our scope. The clearest evidence of positive returns is for men in the 9–12 cohort. Even for that group, however, differences in earnings between those with early versus late exposure likely understate longer-term annual returns because: 1) they only capture the differential earnings effects; 2) they do not reflect final schooling differentials which may have continued to widen, further increasing earnings differentials; and 3) they are measured at the beginning of the beneficiaries' working lives and therefore have potential to increase over time. With these caveats, using only the earnings gains for this cohort and under conservative assumptions regarding cost and benefit flows, the net present value with a 10% discount rate turns positive within 2–3 decades (Appendix I).

VII. Conclusions

Within the framework of a dynamic human capital production function, in this paper we provide experimental evidence that earlier exposure to a CCT in rural Nicaragua translated into sustained positive differences in labor market participation and earnings a decade after the program began, for both men and women. Results suggest, however, that the mechanisms through which the CCT program components generated the positive differentials differed by sex.

For males, exposure to the CCT during primary school ages when they were vulnerable to school dropout led to substantial education and learning gains compared to exposure a few years later. Subsequent employment gains were made possible by higher levels of temporary migration. A plausible interpretation is that with more education and learning the men developed core competencies enabling them to find higher paid work away from home when they were beginning their adult working lives, possibly setting them on a higher future earnings trajectory.

Results on sustained learning for the young men contrast with some findings in the literature that point to limited effects of CCTs on learning, findings that led a recent high-level panel to conclude CCTs may not be cost-effective for addressing learning gaps (World Bank 2020). Positive learning and income differentials in Nicaragua were substantial. At the same time, given low initial education levels and other remaining constraints, the program did not fully erase learning gaps or wholly transform the lives of the young men. Nevertheless, cost-benefit analysis suggest the earning gains were sufficient to recover investments within a few decades.

For females, earlier exposure to the CCT also led to positive differentials in schooling attainment, but not in learning. In addition, the timing of program-related nutrition shocks led to later onset of menarche and sexual activity, and lower BMI in the early compared to the late treatment group. These differences in reproductive health-related outcomes help explain the positive labor market differentials for the young women. For those in late treatment, the effect of the nutrition shocks on sexual maturation may have partly offset possible delayed fertility resulting from higher schooling, while for those in early treatment the opposite shock may have complemented any such schooling effects. The results underscore the importance of accounting for the nutrition and health implications—and not just education—of cash transfer or related programs when analyzing longer-term effects. 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, so the specific ages at which interventions target girls can affect subsequent marriage, fertility and labor market outcomes.

Taken together, the evidence suggests that the specific timing of nutrition, health and education interventions during primary school ages can have substantial consequences for their effectiveness and long-term returns. Moreover, at these ages different types of human capital can have similar or reinforcing effects on subsequent outcomes but can also potentially affect important outcomes such as fertility and labor market participation in opposite directions. Future research should aim to further disentangle the roles of nutrition and education during these sensitive years, and their interactions, to ensure optimal returns.

VIII. 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):1481–95.

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.

Araujo, M. C., and K. Macours, 2021. "Education, income and mobility: Experimental impacts of childhood exposure to *PROGRESA* after 20 years," *PSE working paper* 2021/10.

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.

Attanasio, O., C. Meghir and E. Nix. 2020. "Human Capital Development and Parental Investment in India." *Review of Economic Studies* 87: 2511–41.

Avitabile, C, 2012. "Spillover Effects in Healthcare Programs: Evidence on Social Norms and Information Sharing" *IDB Working paper* 380.

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.

Baird, S., C. McIntosh, and B. Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126(4): 1709–53.

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: 169–85.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman. 2020 "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa", *American Economic Journal: Applied Economics*, 12(1):210–59.

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): 811–70.

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): 467–71.

Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E. (2019). "Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia." *American Economic Journal: Applied Economics,* 11(3): 54–91.

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*, 219–70. 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): 439–77.

Behrman, J. R., S.W. Parker and P.E. Todd. 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Follow-up of *PROGRESA*/Oportunidades." *Journal of Human Resources* 46(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): 2904–10

Buchmann, 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, R.A. Prima, E. Satriawan, and E. Syamsulhakim. 2020. "Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia." *American Economic Journal: Economic Policy*, 12(4): 88–110.

Carneiro, P., I. Lopez Garcia, K.G. Salvanes and E. Tominey, 2021. "Intergenerational Mobility and the Timing of Parental Income", *Journal of Political Economy*, 129(3) 757–88.

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 of menarche." *Br. J. Obstet. Gynaecol*. 103(8): 814–7.

Cunha, F. and J. Heckman. 2007. "The Technology of Skill Formation." *American Economic Review Papers and Proceedings* 97(2): 31–47.

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

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): 2757–97.

Duflo, E., P. Dupas and M. Kremer. 2021. "The Impact of Free Secondary Education: Experimental Evidence from Ghana. *NBER Working Paper 28937*.

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.

Evans, D., and F. Yuan, 2022. "How Big Are Effect Sizes in International Education Studies?", *Educational Evaluation and Policy Analysis*, 44(3):532-540.

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* 371:828–37.

Fernald, L., R. Hamad, D. Karlan, E. Ozer and J. Zinman. 2008. "Small individual loans and mental health: a randomized controlled trial among South African adults", *BMC Public Health* 8: 409.

Field, E., and A. Ambrus. 2008. “Early Marriage, Age of Menarche, and Female Schooling Attainment in Bangladesh.” *Journal of Political Economy* 116(5): 881–930.
Filmer, D., and N. Schady. 2014. “The Medium-Term Effects of Scholarships in a Low-Income Country.” *Journal of Human Resources* 49(3): 663–94.
Fiszbein, A., and N. Schady. 2009. “Conditional Cash Transfers: Reducing Present and Future Poverty.” *World Bank Policy Research Report*. Washington, DC: World Bank.
Garn, S.M., 1987. “The secular trend in size and maturational timing and its implications for nutritional assessments: a critical review.” *Journal of Nutrition* 117, 817–23.
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.
Gertler, P., S. Martinez, and M. Rubio-Codina, 2012. "Investing cash transfers to raise long term living standards." *American Economic Journal: Applied Economics,* 4(1): 164–92.
Gluckman, PD and M.A. Hanson. 2006, “Evolution, Development and Timing of Puberty” *Trends in endocrinology and metabolism* 17(1): 7–12.
Khana, Madhulika. 2020. “The Precocious Period: Impact of Early Menarche on Schooling in India.” Unpublished manuscript, Georgetown Unitervisty.
Handa, S. L. Natali, D. Seidenfeld, G. Tembo and B. Davis, 2018. “Can Unconditional Cash Transfers Raise Long-term Living Standards? Evidence from Zambia”, *Journal of Development Economics*, 133: 42-65.
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): 411–16.
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.
Khan, M.E., A. Hazra, A. Kant and M. Ali, 2016. “Conditional and Unconditional Cash Transfers to Improve Use of Contraception in Low and Middle Income Countries: A Systematic Review”, *Studies in Family Planning*, 47(4): 371-383.
Kling, J., J. Liebman and L. Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica* 75(1): 83–119.
Laajaj, R. and K. Macours. 2021. “Measuring Skills in Developing Countries.” *Journal of Human Resources* 56(4): 1254–95.
Lamadrid-Figueroa, H. G. Angeles, T. Mroz, J. Urquieta-Salomon, B. Hernandez-Prado, A. Cruz-Valdez and M.M. Tellez-Rojo et al 2010. “Heterogeneous Impact of the Social Programme Oportunidades on Use of Contraceptive Methods by Young Adult Women Living in Rural Areas”, *Journal of Development Effectiveness*, 2(1): 74-86.
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: 247–73.
Macours, K. and R. Vakis. 2016. “Sustaining Impacts When Transfers End: Women Leaders, Aspirations, and Investment in Children.” *NBER Working Paper* No. 22871.
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., 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): 34(1): 119–59.
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, K. Macours, J.A. Maluccio and L. Tejerina. 2020. "Experimental long-term impacts of early childhood and school-age exposure to a Conditional Cash Transfer Program." *Journal of Development Economics* 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.
Stecklov, G., P. Winters, J. Todd, and F. Regalia. 2007. "Unintended effects of poverty programmes in childbearing in less developed countries: Experimental evidence from Latin America," *Population Studies* **61**(2): 125–140.
Todd, J. E., P. Winters, and G. Stecklov. 2012. "Evaluating the impact of conditional cash transfer programs on fertility: The case of the *Red de Protección Social* in Nicaragua," *Journal of Population Economics* **25**: 267–290.
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): 557–98.

TABLE 1: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR ALL FAMILIES OF OUTCOMES, 7–12 AND 9–12 COHORTS, MALES AND FEMALES

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Schooling Family Z-score	Learning Family Z-Score	Fertility and Marriage Family Z-Score	Socio-Emotional Family Z-score
		Rank	Earnings (5 % Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 7–12-Year-old Males and Females							
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: 9–12-Year-old Males and Females							
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: 9–12-Year-old Males							
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)
FWER p	[0.005]	[0.005]	[0.011]	[0.023]	[0.014]	[0.116]	[0.067]
RI p	[0.002]	[0.001]	[0.008]	[0.036]	[0.030]	[0.442]	[0.257]
N	1,006	1,006	997	1,007	907	907	900
Panel D: 9–12-Year-old Females							
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)
FWER p	[0.066]	[0.072]	[0.076]	[0.066]	[0.361]	[0.066]	[0.141]
RI p	[0.038]	[0.072]	[0.098]	[0.020]	[0.936]	[0.025]	[0.325]
N	888	888	878	888	827	809	821
Panel E: Difference between Male and Female 9–12 cohorts							
P-value	0.288	0.268	0.263	0.964	0.043	0.276	0.102

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument, in columns 5 and 7 using the 2010 individual survey instrument, and in column 6 using both. Regressions include males and/or females 7–12 or 9–12 years old 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 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 tables 2–6), with individual z-scores calculated using the mean and standard deviation of the late treatment group. The fertility and marriage family has different components for males and females and therefore is not combined. Controls include three-month age group indicators and indicator variables for grades completed 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 hypotheses 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 2: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR LABOR MARKET PARTICIPATION FAMILY COMPONENTS, 9–12 COHORT BY SEX

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components			
		Worked Off Family 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 Late Treatment		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 Late Treatment		0.463	0.074	0.312	0.234

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females 9–12 years old 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 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 completed 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: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR EARNINGS FAMILY COMPONENTS, 9–12 COHORT BY SEX

	Earnings Family Z-Score	Earnings Family Components (all unconditional)			
		Earnings Per Month Worked (Last 12 Months)	Annual Earnings (Last 12 Months)	Maximum Monthly Earnings (Last 12 Months)	Maximum Monthly Non-agricultural Salary (Ever)[1]
	(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 Late Treatment		497.2	502.5	486.9	498.3
Panel B: Earnings (C$) 5% Trim, Males					
ITT	0.192***	201.152***	595.013	211.421***	142.260*
	(0.067)	(63.624)	(619.322)	(69.318)	(71.919)
N	997	956	956	956	955
Mean Late Treatment		1436	8222	1619	227.5
Panel C: Rank of Earnings, Females					
ITT	0.116*	33.758**	28.195*	4.148	31.313*
	(0.061)	(15.706)	(14.011)	(15.038)	(15.544)
N	888	888	888	883	888
Mean Late Treatment		414.8	417.4	434.9	415.5
Panel D: Earnings (C$) 5% 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 Late Treatment		464.2	2628	492.9	309.4

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Regressions include males or females 9–12 years old 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 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 5% of values. Earnings shown in Nicaraguan Cordobas (C$); the exchange rate was approximately 20 C$ per U.S. dollar 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 completed at baseline, stratification and region. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

[1] Because relatively few men ever held a non-agricultural salaried job, the unconditional mean in Panel B, column 5 is much lower than earnings per month worked in column 1.

TABLE 4: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR SCHOOLING, LEARNING AND COGNITION FAMILY COMPONENTS, 9–12 COHORT BY SEX

	Schooling Family Z-Score	Schooling Family Components			Read and Write =1	Learning Families			Mixed Cognition and Learning	Cognition (Raven)
		Grades Completed	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 Late Treatment		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 Late Treatment		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 in columns 6–10 using the 2010 individual survey instrument. Regressions include males or females 9–12 years old 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 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 for tests 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 completed 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 5: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR FERTILITY AND MARRIAGE FAMILY COMPONENTS, 9–12 COHORT BY SEX

	Fertility and Marriage Family Z-Score	Fertility and Marriage Family Z-Score (Excluding BMI)	Fertility and Marriage 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 (BMI)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Males									
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 Late Treatment			0.269	0.311	0.225			0.863	
Panel B: Females									
ITT	-0.167***	-0.166**	-0.109**	-0.039	-0.064	0.249**	-0.656***	-0.058**	-0.063*
	(0.060)	(0.063)	(0.041)	(0.043)	(0.047)	(0.119)	(0.236)	(0.028)	(0.037)
N	809	809	809	809	809	806	765	749	792
Mean Late Treatment			0.287	0.612	0.527	13.11	23.64	0.815	0.751

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 or females 9–12 years old 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 ITT results compare early to late treatment groups in 2010. Age of menarche is reversed when it is included in the fertility and marriage family. BMI is not available for men. BMI included only for women not pregnant at the time of measurement. 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 completed 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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR SOCIO-EMOTIONAL FAMILY COMPONENTS, 9–12 COHORT BY SEX

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

Notes: Outcome data source: 2010 long-term evaluation individual survey instrument. Regressions include males or females 9–12 years old 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 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 (after reversing signs for negatively oriented self-perception and stress). 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 completed 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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR ALL FAMILIES, TWO-YEAR COHORTS BY SEX

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Schooling Family Z-Score	Learning Family Z-Score	Fertility and Marriage Family Z-score	Socio-Emotional Family Z-Score
		Rank	Absolute (5 % Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: 7–8-Year-old 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: 9–10-Year-old Males							
ITT	0.213***	0.173*	0.183**	0.126**	0.173*	-0.030	0.001
	(0.079)	(0.090)	(0.079)	(0.058)	(0.092)	(0.072)	(0.061)
N	466	466	463	467	430	430	427
Panel C: 11–12-Year-old Males							
ITT	0.341***	0.220***	0.208*	0.066	0.187**	-0.062	0.094**
	(0.110)	(0.071)	(0.106)	(0.078)	(0.079)	(0.082)	(0.045)
N	540	540	534	540	477	477	473
Panel D: 7–8-Year-old 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 E: 9–10-Year-old Females							
ITT	0.119	0.120	0.146*	0.100	0.085	-0.240***	0.011
	(0.077)	(0.079)	(0.081)	(0.077)	(0.069)	(0.076)	(0.061)
N	403	403	399	403	380	374	377
Panel F: 11–12-Year-old Females							
ITT	0.184*	0.088	0.026	0.100*	-0.086	-0.095	-0.118*
	(0.099)	(0.087)	(0.088)	(0.058)	(0.069)	(0.069)	(0.060)
N	485	485	479	485	447	435	444
Panel G: P-value of Difference between 7–8 and 9–12 cohorts							
Males	0.002	0.019	0.049	0.012	0.012	0.329	0.390
Females	0.068	0.193	0.272	0.024	0.911	0.032	0.939

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using the 2010 household survey instrument, in columns 5 and 7 using the 2010 individual survey instrument, and in column 6 using both. 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 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 completed 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: 2002 AND 2004 AVERAGE ENROLLMENT RATES, TWO-YEAR COHORTS BY SEX

	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. Means are based on males or females with the indicated age at the start of the program in November 2000, weighted to account for sampling in the short-term evaluation survey providing population estimates of the study area.

TABLE 9: 2002 AND 2004 SHORT-TERM EXPERIMENTAL DIFFERENTIALS FOR EDUCATION, 9–12 COHORT BY SEX

	Grades Completed (years)	Completed Grade 4 =1	Enrolled =1	Regularly attended School > 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 in the short-term evaluation survey providing population estimates of the study area. (Results for other age groups shown in Appendix Table A7). The differential 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. Mean LT refers to mean in the late treatment group. Grades completed reflect the completed school year prior to each survey year. Attended school for more than 85% 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 completed 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: 2002 AND 2004 SHORT-TERM EXPERIMENTAL DIFFERENTIALS FOR CONSUMPTION, HOUSEHOLDS OF 9–12 COHORT CHILDREN BY SEX

	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: Males 2002					
ITT	0.354***	0.046**	0.041***	-0.082***	-0.005
	(0.085)	(0.018)	(0.007)	(0.020)	(0.012)
N	500	500	500	500	500
Mean LT	7.309	0.206	0.054	0.534	0.206
Panel B: Males 2004					
ITT	-0.273***	-0.001	-0.000	0.009	-0.005
	(0.068)	(0.016)	(0.008)	(0.016)	(0.012)
N	490	494	494	494	494
Mean LT	8.012	0.201	0.085	0.512	0.196
Panel C: Females 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 D: Females 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 in the short-term evaluation survey providing population estimates of households with boys (girls) this age in the study area. The differential 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. Mean LT refers to mean in the late treatment group. 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 completed at baseline, stratification and region. Standard errors clustered at the locality level 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

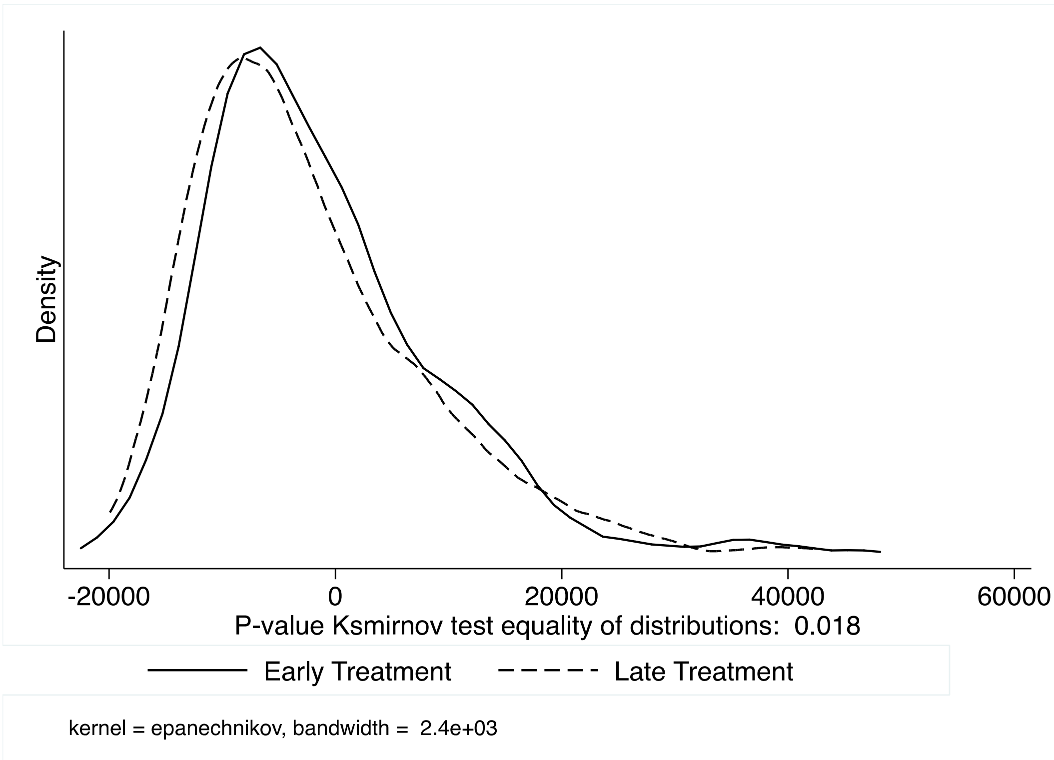
By TANIA BARHAM, KAREN MACOURS AND JOHN A. MALUCCIO

August 2023

List of Appendices

APPENDIX A: ADDITIONAL TABLES AND FIGURES

FIGURE A1: OFF-FARM EARNINGS PER MONTH WORKED, 9–12 COHORT, MALES



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 5% of values.

TABLE A1-A: BASELINE BALANCE, 9–12 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 years	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 (=1)	0.08	0.28	516	0.07	0.33	490	0.01	0.43	0.03
Father not living in same household (=1)	0.22	0.66	516	0.18	0.40	490	0.04	0.16	0.11
Child of household head (=1)	0.86	0.41	516	0.88	0.35	490	-0.02	0.39	-0.06
Mother has no grades attained (=1)	0.45	0.81	516	0.49	0.77	490	-0.04	0.32	-0.05
Mother has 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 in years	44.8	14.6	516	44.4	13.2	490	0.40	0.58	0.03
Has no grades attained (=1)	0.53	0.80	516	0.50	0.55	490	0.04	0.33	0.06
Has 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									
Predicted log expenditures (per capita)	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 ages 0–8	2.10	2.31	516	2.08	2.57	490	0.02	0.92	0.01
Number children ages 9–12	1.76	0.89	516	1.80	1.38	490	-0.04	0.53	-0.03
Logarithm 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 agric (=1)	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
Lives in Tuma region (=1)	0.52	2.74	516	0.29	2.40	490	0.23	0.13	0.09
Lives 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 (=1)	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 9–12 years old 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 providing population estimates of the study area. The mean divided by the standard deviation uses the standard deviation for the late treatment group. Standard deviations (and standard errors for the p-value calculation which also controls for strata) are clustered at the locality level. The predicted logarithmic per capita expenditures are calculated from the 2000 program census data based on a proxy means method developed using the 1998 Nicaraguan Living Standards Measurement Survey (Maluccio 2009). The wealth indices are constructed using principal component analysis for housing characteristics and assets. School characteristics are from program administrative data collected for monitoring conditionalities. Schools under local governance participated in Nicaraguan's school autonomy reform that provided schools and parents a degree of autonomy over management and operations.

TABLE A1-B: BASELINE BALANCE, 9–12 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 years	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 (=1)	0.08	0.25	448	0.08	0.22	440	0.01	0.61	0.03
Father not living in same household (=1)	0.27	0.79	448	0.28	1.00	440	0.00	0.96	0.00
Child of household head (=1)	0.86	0.43	448	0.84	0.49	440	0.02	0.47	0.04
Mother has no grades attained (=1)	0.44	0.68	448	0.40	0.68	440	0.04	0.45	0.06
Mother has 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 in years	45.40	15.87	448	43.43	16.25	440	1.97	0.03	0.12
Has no grades attained (=1)	0.55	0.53	448	0.48	0.52	440	0.07	0.02	0.14
Has 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									
Predicted log per capita expenditures	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 ages 0–8	2.11	2.76	448	2.18	2.56	440	-0.07	0.66	-0.03
Number children ages 9–12	1.75	0.84	448	1.77	1.19	440	-0.02	0.73	-0.02
Logarithm 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	202.8	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 agric (=1)	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	116.8	448	22.72	71.40	440	8.29	0.24	0.12
Lives in Tuma region (=1)	0.52	2.51	448	0.36	2.58	440	0.17	0.30	0.06
Lives 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 (=1)	0.31	2.03	441	0.31	2.12	434	0.00	0.89	0.00

Notes: Outcome data sources: 2000 program census. Results include females 9–12 years old at the start of the program in November 2000 with information in the 2010 household survey. See Table A1-A for additional notes.

TABLE A2: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR ALL FAMILIES OF OUTCOMES, 7–12 COHORT BY SEX

	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Schooling Family Z-Score	Learning Family Z-Score	Fertility & Marriage Family Z-score	Socio-Emotional Family Z-Score
		Rank	Absolute (5 % Trim)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel C: 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: 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, in columns 5 and 7 using the 2010 individual survey instrument, and in 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 and marriage 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 A3: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIAL QUANTILE REGRESSIONS ON 5% TRIM EARNINGS FAMILY (Z-SCORE), 9–12 COHORT, MALES

	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 9–12 years old at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The ITT differential 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 3), 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 A4: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR LEARNING AND COGNITION BY TEST (Z-SCORES), 9–12 COHORT BY SEX

	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
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

Notes: Outcome data source: 2010 long-term evaluation individual survey. Regressions include males or females 9–12 years old at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The ITT differential results compare early to late treatment groups in 2010. Individual component z-scores for tests 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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR FERTILITY AND MARRIAGE FAMILY COMPONENTS AND MECHANISMS, 9–10 AND 11–12 COHORTS, FEMALES

	Fertility and Marriage Family Z-Score	Fertility and Marriage Family Z-Score (Excluding BMI)	Fertility and Marriage 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 (BMI)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 9–10 Cohort									
ITT	-0.240***	-0.252***	-0.114**	-0.129**	-0.110	0.327**	-0.698*	-0.064	-0.120*
	(0.076)	(0.089)	(0.051)	(0.059)	(0.052)	(0.128)	(0.290)	(0.045)	(0.060)
N	374	374	374	374	374	371	354	348	368
Mean Late Treatment			0.293	0.564	0.439	13.01	23.26	0.814	0.710
Panel B: 11–12 Cohort									
ITT	-0.095	-0.091	-0.098*	0.035	-0.034	0.164	-0.534	-0.055	0.013
	(0.069)	(0.073)	(0.054)	(0.052)	(0.059)	(0.163)	(0.369)	(0.051)	(0.048)
N	435	435	435	435	435	435	411	401	424
Mean Late Treatment			0.281	0.661	0.617	13.22	24.03	0.816	0.793

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 9–10 or 11–12 years old at the start of the program in November 2000, weighted to account for sampling and attrition providing population estimates of the study area. The ITT differential results compare early to late treatment groups in 2010. Age of menarche is reversed when it is included in the fertility and marriage family. BMI included only for women not pregnant at the time of measurement. 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 completed 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 Long-term Experimental Differentials for Individuals Exposed to Treatments at the Same Ages, 9–12 Years Old When First Exposed by Sex

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

Notes: Outcome data source: 2010 long-term evaluation survey. Age of menarche and sexual activity measured using the 2010 individual survey instrument and schooling using the 2010 household 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. The ITT differential results compare early to late treatment groups in 2010. Controls include indicator variables for stratification and region. The outcome variables are also age-adjusted to control for linear time trends related to possible cohort effects, using the residual of the regression of the outcome variable 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 A7-A: 2002 (ABSOLUTE) AND 2004 (DIFFERENTIAL) SHORT-TERM EXPERIMENTAL EFFECTS BY COHORTS, MALES

	Grades Attained (years)	Completed Grade 4 =1	Enrolled =1	Regularly Attended School > 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 differentials in 2004. Grades attained reflect the completed school year prior to each survey year. Attended school more than 85% 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 A7-B: 2002 (ABSOLUTE) AND 2004 (DIFFERENTIAL) SHORT-TERM EXPERIMENTAL EFFECTS BY COHORTS, FEMALES

	Grades Attained (years)	Completed Grade 4 =1	Enrolled =1	Regularly Attended School > 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 A7-A. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX B: ROBUSTNESS—ALTERNATIVE SPECIFICATIONS AND SAMPLES

In Tables B1-A (men) and B (women), we examine the sensitivity of the 2010 long-term experimental differentials for the 9–12 cohort by sex to alternative specifications, samples and approaches for addressing attrition. Results for the main specifications in Table 1 are reproduced in panel A.

First, in panel B we estimate differentials controlling only for strata as per the stratified randomized design of the program, removing from the main specification the three-month age group indicators and indicator variables for grades completed at baseline and region. Without the additional controls standard errors increase, as expected. For men the results are largely consistent with the main results with the exception of smaller and marginally insignificant (p-value=0.11) point estimates for labor market participation and slightly larger but insignificant point estimates for schooling. The differential for the schooling family for women is also smaller and no longer statically significant.

Second, in panel C we reincorporate the main specification controls and additionally include an extended set of controls to test sensitivity to characteristics not well balanced at baseline (Appendix Table A1). For men, the additional controls are an indicator for whether the child worked at baseline and binary indicators for the following variables above the mean: the productive assets component of the wealth index, family network size and village population. For women, the additional controls are indicator variables for the household head not having any education, household head age above the mean and the population size of the village above the mean. Results are similar to the main findings for both sexes, although for women the earnings rank family is now marginally insignificant (p-value=0.11).

Third, in panel D we exclude all children from households that were oversampled in the 2010 long-term evaluation survey, only retaining children in the short-term baseline 2000 evaluation survey sample. This reduces the sample size by about half. Point estimates are broadly similar to the main results for both sexes and statistical significance the same for men despite the smaller sample sizes. For women, however, results are no longer significant with the important exceptions of the labor market participation and fertility and marriage families. Overall, these results confirm that oversampling in the 2010 long-term follow-up was important and provided sufficient statistical power to estimate differentials.

Fourth, we explore the sensitivity of the analyses to the weighting. The main results are weighted to account for attrition using information from the survey tracking process. Specifically, IPW attrition correction weights are constructed using observations from the intensive tracking phase and then multiplied by long-term survey sampling weights providing population estimates of the study area (see Appendix E). Alternatively, we can estimate results without attrition correction, using only the long-term evaluation survey sampling weights. This approach ignores the possibility of differential selected attrition by treatment group but can be justified because attrition rates and baseline characteristics of the final interviewed samples are balanced across treatment groups. Results for families are presented in panel E and for the components of each family in Tables B2–B6: labor market participation (B2), earnings (B3), schooling (B4), learning (B5) and fertility and marriage (B6). We further present a different alternative, in panel F, using a different IPW attrition correction weight constructed using observations from both the regular and intensive tracking phases (and then multiplied by long-term survey sampling weights providing population estimates of the study area). All findings are robust to only using sampling weights (panel E and Tables B2–B6) and to using alternative attrition weights (panel F).

Fifth, to account for potential correlation between outcomes within the same family we present results for an alternative index using inverse covariance weighting (Anderson 2008) in panel G. Significant differentials for men are the same or larger. For women, earnings families are no longer marginally significant but differentials for labor market participation and fertility and marriage families are larger.

Sixth, we present two types of attrition bounds: Lee (2009) bounds (panel H); and Kling and Liebman (2004) sensitivity bounds (panel I). Lee bounds assume monotonicity, i.e., that the tracked sample is either entirely negatively or entirely positively selected. 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 the estimation of the differential effects. Lee bounds are estimated omitting covariates and strata controls—i.e., without any covariates—and therefore are not directly comparable to other panels in the table. Given the limited sample sizes, we are unable to trim the samples using covariates to obtain tighter bounds. For men, omitting covariates and strata controls increases point estimates for completed grades variable alone (not shown) and learning results, with both lower and upper bounds significantly different than zero. Both bounds for earnings results are also significant. On the other hand, it reduces the size of the point estimates for the labor market participation and schooling families, and their lower bounds are insignificant. For women, omitting covariates and strata controls increases point estimates for the labor market and earnings families, with both lower and upper bounds significantly different than zero. For schooling, however, both the lower and upper bounds are insignificant. It also reduces the magnitude of the point estimates for the fertility and marriage family, and the upper bound is insignificant. (Lee bounds for the age-of-menarche variable alone suggest it is bounded between 0.17 and 0.38, not shown).

In panel I we present the Kling and Liebman sensitivity bounds for 0.1 SD. The lower/upper bounds are calculated assuming the early treatment group attritors are 0.1 SD above/below the mean of non-attritors and late treatment attritors are 0.1 below/above the mean for non-attritors. For men, the findings are similar to the main results. Probably because of higher attrition, results for women are more sensitive though qualitatively similar. For all but the labor market participation family, point estimates are smaller (though the direction the same) in absolute value and not statistically significant for both lower and upper bounds.

As on further alternative check on attrition selection bias we estimate results for select variables using the proxy measures for which there were much lower rates of attrition (4% for men, 6% for women) but also likely greater measurement error. Results are similar to those in the paper (Table B7).

TABLE B1-A: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR ALL FAMILIES OF OUTCOMES (Z-SCORES), ALTERNATIVE SPECIFICATIONS AND SAMPLES, 9–12 COHORT, MALES

	Labor Market Participation	Earnings		Schooling	Learning	Fertility & Marriage	Socio-emotional
		Rank	5% Trim				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (from Table 1)							
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 Oversample 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 Only (No IPW 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: IPW Attrition Correction Using Observations from Both Intensive & Regular Tracking							
ITT	0.282***	0.195***	0.181**	0.089*	0.164**	-0.0132	0.045
	(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: Families Constructed Using Inverse Covariance 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 H: Lee Attrition 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)
Panel I: Kling and Liebman Attrition Sensitivity bounds (+/- 0.1 SD)							
Lower Bound	0.231***	0.149***	0.144**	0.094**	0.124**	-0.018	0.004
	(0.072)	(0.047)	(0.059)	(0.040)	(0.049)	(0.053)	(0.029)
Upper Bound	0.253***	0.198***	0.195***	0.112***	0.200***	-0.001	0.058*
	(0.071)	(0.048)	(0.058)	(0.039)	(0.050)	(0.052)	(0.031)

Notes: Outcome data source: 2010 long-term evaluation survey. Individual-level variables in columns 1–4 measured using 2010 household instrument, in columns 5 and 7 using 2010 individual instrument and in column 6 using both. Regressions include males 9–12 years old 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 ITT differential 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 completed at baseline, stratification and region, unless otherwise noted in the text (panels B, C, H and I). Standard errors clustered at the locality level are in parentheses except for Lee bounds for which bootstrapped 1,000 repetitions. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B1-B: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR ALL FAMILIES OF OUTCOMES (Z-SCORES), ALTERNATIVE SPECIFICATIONS AND SAMPLES, 9–12 COHORT, FEMALES

	Labor Market Participation	Earnings		Schooling	Learning	Fertility & Marriage	Socio-emotional
		Rank	5% Trim				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Main Specification (from Table 1 in paper)							
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 Oversample 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 Only (No IPW 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: IPW Attrition Correction Using Observations from Both Intensive & Regular Tracking							
ITT	0.204***	0.113**	0.108*	0.085**	0.020	-0.106*	-0.067
	(0.0637)	(0.0559)	(0.0600)	(0.0322)	(0.0557)	(0.0581)	(0.0489)
N	888	888	878	888	827	809	821
Panel G: *Families Constructed Using Inverse Covariance 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 H: Lee Attrition 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)
Panel I: Kling and Liebman Attrition Sensitivity Bounds (+/- 0.1 SD)							
Lower	0.160**	0.049	0.042	0.059	-0.031	-0.112**	-0.103***
Bound	(0.073)	(0.043)	(0.044)	(0.041)	(0.053)	(0.044)	(0.035)
Upper	0.185**	0.114**	0.110**	0.079*	0.048	-0.055	-0.049
Bound	(0.072)	(0.043)	(0.045)	(0.040)	(0.051)	(0.045)	(0.034)

Notes: Regressions include females 9–12 years old at the at the start of the program in November 2000. See Table B1-A for all other notes. *** p<0.01, ** p<0.05, * p<0.10.

TABLE B2: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR LABOR MARKET PARTICIPATION FAMILY COMPONENTS USING SAMPLING WEIGHTS, 9–12 COHORT BY SEX

	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 Late Treatment		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 Late Treatment		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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR EARNINGS FAMILY COMPONENTS USING SAMPLING WEIGHTS, 9–12 COHORT BY SEX

	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 Late Treatment		493.3	492	494.5	482.6
Panel B: Earnings (C$) — 5% Trim, Males					
ITT	0.188***	203.325***	555.966	196.323**	157.602**
	(0.069)	(68.572)	(630.163)	(72.834)	(72.671)
N	997	956	956	956	955
Mean Late Treatment		1437	7996	1626	219
Panel C: Rank of Earnings, Females					
ITT	0.107*	27.841*	23.562*	12.873	25.052*
	(0.053)	(14.874)	(12.443)	(12.630)	(14.559)
N	888	888	888	888	883
Mean Late Treatment		419.7	421.3	427.8	420.8
Panel D: Earnings (C$) — 5% 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 Late Treatment		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 5% 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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR SCHOOLING FAMILY COMPONENTS AND LITERACY USING SAMPLING WEIGHTS, 9–12 COHORT BY SEX

	Schooling Family Z-Score	Schooling 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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES) USING SAMPLING WEIGHTS, 9–12 COHORT BY SEX

	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 LONG-TERM EXPERIMENTAL DIFFERENTIALS FOR FERTILITY AND MARRIAGE FAMILY COMPONENTS AND MECHANISMS USING SAMPLING WEIGHTS, 9–12 COHORT BY SEX

	Fertility & Marriage Family Z-Score	Fertility and Marriage 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
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 and marriage 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.

TABLE B7: 2010 LONG-TERM EXPERIMENTAL DIFFERENTIALS USING PROXY VARIABLES, 9–12 COHORT BY SEX

	Completed Grades (1)	Completed Grade 4 =1 (2)	Enrolled =1 (3)	Read and Write =1 (4)	Ever Married =1 (5)	Worked Off Family Farm =1 (last 12 months) (6)
Panel A: Males						
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: Females						
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

Notes: Outcome data source: 2010 long-term evaluation household survey instrument. Columns 1–4 correspond to columns 2–5 in Table 4; column 5 to column 4 in Table 5; column 6 to column 2 in Table 2. Regressions include males or females 9–12 years old at the start of the program in November 2000, weighted to account for sampling (but not attrition) providing population estimates of the study area. The differential experimental ITT results compare early to late treatment groups in 2010. . If respondent no longer member of their original household nor member of a split-off household that was interviewed, 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.

APPENDIX C: TARGETING AND DESIGN OF THE NICARAGUAN CCT PROGRAM[1]

The *Red de Protección Social* (RPS) CCT program was designed to alleviate both current and future poverty through cash transfers targeted to poor and extremely poor households in rural Nicaragua. Its stated objectives were: 1) supplementing household income for up to three years to increase expenditures on food; 2) increasing the nutritional status and healthcare of children under five; and 3) reducing dropout rates during the first four years of primary school. A randomized evaluation was incorporated into the program design and 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. The first phase lasted three years with a budget of $11 million (M). In late 2002, based in part on the positive findings of short-term impact evaluations, the Government of Nicaragua and IDB agreed to a continuation and expansion for a second phase through 2005, with additional budget of $22M.

Program Targeting—The CCT first targeted the rural areas in six municipalities from three regions in central and northern Nicaragua, selected based on their low health and educational outcomes and extreme 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% of Nicaraguans identified as poor in 1998, 75% resided in rural areas (World Bank 2001). Although the six targeted municipalities were not the poorest in the country, the proportion of impoverished people living in them was well above the national average (World Bank 2003). At the same time, there was relatively good access to the selected municipalities (for example, situated less than a one-day drive from Managua where the central program administrative office was located), relatively strong governmental institutional capacity and local coordination, and mostly adequate primary school coverage. Health services coverage was less complete and therefore the program increased local supply of services in parallel with the program as described below.

In the second stage of targeting, a marginality index was constructed for all 59 of the rural census *comarcas* (hereafter localities)[2] within the six municipalities. The index was the weighted average of a set of locality-level indicators (including literacy, family size, and access to water and latrines)—and higher marginality index scores were associated with greater poverty.[3] 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 set of localities used in the paper. Finally, although the original program design envisioned all households in the 42 targeted localities to be eligible for the CCT, prior to the start of the program the government excluded ~3% determined to have substantial resources (including a vehicle or large landholdings) based on information from the 2000 program registry census. These households are excluded from the analyses in the paper.

[1] The appendix expands on Section II, based on IFPRI (2005), Maluccio and Flores (2005) and Maluccio (2009).

[2] Census *comarcas* were administrative areas based on the 1995 Nicaraguan National Population and Housing Census and included as many as 10 small communities (or villages), totalling approximately 250 households.

[3] Specifically, the marginality index for each locality included the percent of individuals over five years old who were illiterate (0.3 weight), average family size (0.1 weight), the percent without piped water in the home or yard (0.5 weight) and the percent without a latrine (0.1 weight)—all calculated from the 1995 national census.

Although not statistically representative of rural Nicaragua, the 42 localities comprising the randomized evaluation area were 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 Nicaragua had marginality index scores in the same range as the program localities. Moreover, poverty rates in the targeted localities were ~10 percentage points higher than rural national rates: 80% versus 69% poor, and 42% versus 29% extremely poor.

Over time the CCT expanded to the remaining 17 (= 59 – 42) rural localities within the six municipalities not initially targeted and not part of the randomized evaluation. In the 17 localities, which were less poor according to the marginality index, the CCT began in late 2001 and was offered to ~ 80% of the population based on a household-level proxy means targeting model. Consequently, from 2000–5 the three-year CCT was implemented (at different times and to modestly different degrees) in all 59 rural localities in the six municipalities and more than 90% of the overall population had been eligible. In December 2005, the program ended.

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 healthcare and education behaviors, with conditions monitored by specially contracted healthcare providers and by teachers and school administrators. Benefits and conditionalities were explained to families in the early treatment group during registration assemblies in September and October 2000 and transfers began in November. The designated household representative (in Spanish the *titular*) received the cash transfers; where possible, the program appointed the mother or another female caregiver to this role. As a result, more than 95% of household representatives were women. The CCT also worked with volunteer coordinators (in Spanish *promotora*), local beneficiary women chosen by the community to help implement the program. The volunteer coordinators organized their group of household representatives and kept them informed about upcoming program activities, upcoming transfer payments and any failures to fulfill conditions. Conditions were monitored for compliance and when they were not met the relevant transfer was withheld by the program. The CCT also had a strong social marketing message that the money was meant for food, health and educational expenditures. The household representative was asked to sign an agreement that they would comply with the conditions and that the money would be spent as intended, although actual expenditures were not monitored or part of the formal program conditionalities.

<u>Food security, nutrition and health component</u>. Each household was eligible for a cash transfer every other month (bimonthly) known as the nutrition and health transfer that was a fixed amount per household, regardless of household size or composition. The transfer was conditional on the household representative attending bimonthly health education workshops and bringing any children under five years old for scheduled preventive healthcare appointments. The workshops were held within the communities and covered household hygiene and sanitation, nutrition, sexual and reproductive health, and other related topics. The required preventive healthcare appointments were scheduled monthly for children under two years old and bimonthly for those 2–5 years old. Healthcare services at the scheduled visits included growth monitoring, vaccination and provision of iron (ferrous sulfate) supplements and anti-parasite medication.

To ensure the increased demand for healthcare services could be met without overcrowding or reductions in service quality, the CCT augmented the supply of those services in program areas. Specifically, it contracted and trained private healthcare providers to deliver the program-related services to eligible beneficiaries free of direct charges (Regalia and Castro 2007).

Beneficiaries were required to use those providers to fulfill the conditions. The providers visited program areas on pre-arranged dates and operated in existing health facilities, community centers or other convenient locations. They recorded all services delivered using specially designed scan forms that were regularly submitted to the central program office where they were digitized and used to assess compliance, verify formal conditions and determine transfer amounts.

Starting in 2003 contraceptive methods and prenatal consultations were made available to beneficiaries through the healthcare providers and, as a condition for the nutrition and health transfer, adolescents required to attend specially designed workshops covering among other things sexual reproductive health and contraception,[4] organized separately for several narrow age groups. The additional services were implemented in both early and late treatment in 2003, but afterward attendance was an ongoing condition only for late treatment. Thus, at some point all children 7–12 years old in 2000 (7–12 cohort) were eligible for the workshops, but attendance was an on-going condition for three years only for those in late treatment. Administrative records indicate that in practice the workshops were not fully implemented in all early treatment localities, probably because there were fewer synergies with other ongoing program components in localities where the program was ending in 2003. Consequently, adolescent attendance at the workshops was lower in the early compared to late treatment group and related services taken up less intensively in early treatment localities.[5]

Education component. Each eligible household received a bimonthly cash transfer (school attendance transfer), contingent on enrollment and regular school attendance of children 7–13 years old 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 (in January) intended for uniforms, shoes and other supplies (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 per child.

To provide incentives to the teachers and increase resources available to the schools, there was an additional small cash transfer (the teacher transfer). Throughout rural Nicaragua, school parents' associations often requested small monthly contributions to support the teachers and the school. The teacher transfer was intended to substitute for this informal fee. The transfer was per child and delivered bimonthly directly to households, who were required to pass the funds along to the teachers. The program recommendation was that the teachers 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.[6] Similar to the healthcare providers, teachers and school administrators completed specially designed scan forms capturing enrollment, attendance and delivery of the teacher transfer that

---

[4] These topics also were covered in the health information workshops for the adult household representative required throughout the program for both treatment groups. Most CCTs, including *PROGRESA*, included similar information sessions (Fiszbein and Schady 2009). Prenatal consultations for pregnant women were also added in late treatment, but only a small fraction of the adolescent females in this study would have been eligible (those pregnant by 2005).

[5] Also starting in 2003, an additional minor program modification for children 6–9 years old was administration of a single tetanus shot; provision was monitored but not a verified requirement for transfer receipt.

[6] The small teacher transfer was budgeted to continue to the end of 2005 in all program areas, including the early treatment where the other household transfers ended in 2003, but was incompletely delivered. In practice this would have affected only the youngest children in the cohorts we study, 7-year-olds at the start of the program in 2000.

were regularly submitted to the central program office where they were digitized and used to assess compliance, verify formal conditions and determine transfer amounts.

Having targeted the program to areas with generally adequate schooling infrastructure, there was no explicit supply-side intervention for education such as new school construction. However, the centrally administered CCT employed a multisectoral approach promoting inter-ministerial cooperation through specially formed committees at local, municipal and national levels. The coordination proved helpful in directing ad hoc supply-side responses to increased demand, including the placement of additional government teachers in some areas in response to increased demand. Starting in 2003 the CCT also provided some limited teacher training.

In Table C1 we summarize the potential for transfer receipt by age and treatment group for children 7–13 years old at baseline. Apart from its relation to early or late treatment timing, child age mattered for the education transfer but not for the nutrition and health transfer.

Transfer Sizes—Based on the September 2000 average exchange rate of 12.85 Nicaraguan Córdobas the initial annual transfer amounts in US dollars were: nutrition and health transfer, $224 a year; school attendance transfer, $112; school supplies transfer, $21; and teacher transfer, $25 ($5 bimonthly during the academic year). On its own, the potential nutrition and health transfer was ~13% of pre-program average total annual household expenditures. A household with one child benefiting from the education component would have received additional transfers of ~8%, yielding a total potential transfer of 21% of pre-program expenditures. On average, transfers made were 18% of pre-program expenditures in the early treatment.

Nominal transfer amounts remained constant during the first budgetary phase, though the real value of the transfers declined by ~8% due to inflation. The total nominal transfer amount was reduced for households in late treatment who entered in 2003. 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 modestly, to $90 per year, while the school supplies and teacher transfer amounts increased from $25 to $40 per eligible child. These represent the potential transfer amount received upon complying fully with all associated conditions, the amount relevant for ITT.

Administrative transfer data for all households of boys and girls 9–12 years old at the start of the program in November 2000 reveals that take-up and average transfers received per household were lower in late treatment. (Results are nearly identical for households with 7–12-year-olds at baseline and for girls versus boys.) Consistent with children aging out in late treatment by the time the program reached their localities, average take-up over the three program years based on receipt of at least partial transfers was 92% in early treatment versus 84% in late treatment for the nutrition and health component and 87% versus 66% for education. Even if take-up was 100% in both treatments, because nominal transfers were lower in late treatment, we would expect average transfers to have been lower. For example, using the schedule of transfers, a household with one child eligible for the education transfer in late treatment would have received 88% of the transfers provided to a household with identical structure and compliance in early treatment. Average transfers received for households of 9–12-year-olds at baseline, which reflect take-up, household structure, eligibility and compliance over the program years, were ~$950 in early treatment and ~$690 in late treatment. Therefore, late treatment households received 73% of the nominal transfers received by early treatment households, lower than the full take-up potential.

Beneficiaries did not receive the nutrition and health transfer or, separately, education transfer in a period when they failed to comply with all relevant conditions for that component. Repeated violations, including two consecutive periods of non-compliance, led to households losing their overall program eligibility permanently.

TABLE C1. POTENTIAL TRANSFER RECEIPT BY AGE AND TREATMENT GROUP: 7–13 YEAR-OLDS AT BASELINE

		Ages of potential eligibility for each transfer based on age at baseline			
		Early treatment (2000–2003)		Late treatment (2003–2005)	
Age at baseline survey	Age at 2010 follow-up evaluation survey	Nutrition and health transfer	Education transfer	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: Table does not incorporate the requirement that the child had not yet completed fourth grade. Early treatment corresponds to November 2000 through October 2003; late treatment from January 2003 to end-2005.
[1.] The household in which the child resided could still be eligible for the education transfer for other children.

APPENDIX D: 2010 INDIVIDUAL SURVEY TEST ADMINISTRATION

The 2010 long-term evaluation survey included a household- and a separate individual-level instrument; all standardized tests (as well as the socio-emotional assessments) were administered as part of the latter. Each test was 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 comparable populations in Latin America, including for evaluations of CCTs in Ecuador (Paxson and Schady 2010), Mexico (Behrman, Parker and Todd 2009a; Fernald, Gertler and Neufeld 2009), and a different CCT in Nicaragua (Macours, Schady and Vakis 2012). A strength of all the tests is that they provide 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.

All test administrators were female and selected for their background (trained as psychologists, as 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 in the administration of all tests, and also responsible for survey questions in the individual-level instrument. During training, emphasis was placed both on gaining the trust of the respondents before starting and on the standardized application of the tests. Test administrators also learned strategies to motivate the respondents to participate, helping to keep final non-response to a minimum.

Tests were administered in the home (or compound) and privacy and confidentiality of the results assured throughout the process. Continued quality and standardized application of the tests was monitored closely, and because of the long survey period, several re-standardization trainings carried out. All tests were scored using standard practice.

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

APPENDIX E: 2010 TRACKING PROTOCOLS AND CONSTRUCTION OF INVERSE PROBABILITY WEIGHTS (IPW) FOR ATTRITION

In this appendix we describe the fieldwork tracking protocols and final attrition. We then present additional analysis of attrition and the methodology for construction of inverse probability weights (IPW) used to correct for possible attrition selection, focusing on the 9–12 cohort; the same methodology was used for 7–8 cohort.

Tracking Protocols and Final Attrition

The 2010 long-term evaluation survey sample frame has a total of 1,330 households from the early and 1,379 households from 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 born after 1 January 1988 (i.e., under 22 years old in 2010). We put particular emphasis on tracking all permanent and temporary migrants. In the first phase of fieldwork (lasting ~6 months), we interviewed households and individuals residing in or near the original localities—we call this the regular tracking phase. Regular tracking was followed by intensive tracking (lasting ~1.5 years), during which we made exhaustive efforts to locate and interview all individuals born after 1 January 1988 not interviewed during regular tracking, through repeat visits to original locations and tracing to reported locations in Nicaragua or Costa Rica, the destination of 95% of international migrants from the sample. For target individuals who were not resident members of households being interviewed, we collected proxy reports of some information on them while interviewing other original household members. Refusal was negligible.

We distinguish between three sources of information for outcomes described in Section IV.C: 1) individual outcomes from the individual-level survey instrument; 2) individual outcomes from the household-level survey instrument where the individual was a member, reported by the individual or by another informed household member if the individual was temporarily absent; and 3) individual outcomes from the household-level survey instrument collected by proxy from another original household member if the individual was never found.

With the oversample, the target sample of 9–12-year-olds in 2000 includes 2,200 individuals: 1,138 men and 1,062 women. Attrition is highest for outcomes measured in the individual-level instrument that required direct, in-person interaction with the respondent. For men attrition is 19% for the individual instrument, 10% for the household instrument and 4% for outcomes measured by proxy report. For women, the attrition rates are 22%, 16% and 6%.[7] Attrition of all types is lower for the 7–8-year-olds in 2000. Consequently, attrition is similar to or lower than in related studies analyzing the longer-term experimental effects of CCTs.[8]

[7] By 2010, 15 males and seven females in the 9–12 cohort had died. Deceased individuals are not used to predict the probability of attrition in the calculation of weights described below since selection for them is likely driven by other factors. Final attrition rates are marginally higher including them.

[8] Behrman, Parker and Todd (2011) analyze the PROGRESA 6-year follow up, with attrition of 40% for individual-level information on comparable age groups. The PROGRESA 10-year follow-up (Kugler and Rojas 2018) has more than 60% attrition. Attrition is 13–16% for a 5-year follow-up of young women in Malawi (Baird, McIntosh and Özler 2019), 19% for a 10-year follow-up of a much younger cohort in Ecuador (Araujo, Bosch and Schady 2018), and 7–15% for a 20-year follow-up (including phone interviews) of PROGRESA (Araujo and Macours 2021).

Given the relatively small sample sizes, we intensively tracked all migrants in the 2010 long-term evaluation survey, in contrast to some studies where due to cost considerations only a subsample is tracked. As reported above the final tracking rate for individuals measured in the household-level instrument is 90% for men (84% for women). This is comparable to other long-term follow-ups of RCTs with both regular and intensive tracking.[9]

Analysis of Attrition

Table E1 (column 1) demonstrates the fieldwork strategies led to balanced attrition between early and late treatment for the 9–12 cohort for both sexes, with the coefficients on the ITT indicator on the probability of interview smaller than |1.5| percentage points for the household and individual instruments with p-values > 0.6.[10] Table E2 demonstrates this was not the case, however, after regular tracking. Characteristics for men interviewed during regular tracking (columns 3–4, 7–8) are not well balanced across treatments but adding those interviewed during intensive tracking resulted in final samples that are (columns 5–6, 9–10), underscoring the importance of intensive tracking. Only 6 (4) of 42 baseline variables differ at a 10% significance level for the household (individual) instrument.[11] Moreover, they are the same variables for which there are significant differences across treatments at baseline (columns 1–2), and hence the differences do not appear to be driven by selective attrition. Results for women are qualitatively similar (not shown).

To further assess the possibility of selective attrition, for each baseline variable X we estimate where, as in the paper T is an ITT indicator equal to one for localities assigned to early treatment and zero for late treatment (Table E3). Three of the estimated coefficients on T are significant for men (none for women) and only a few coefficients on $X*T$ are. Therefore, not only are attrition levels balanced across treatments as described above, but on the whole so are the observable characteristics of attritors. At the same time, the estimated coefficients on X make clear that those who attrited are different from those who were interviewed along a number of dimensions. The differences imply that even if internal validity is not jeopardized, differential ITT estimates may not be representative of the target population, threatening external validity. This is particularly important if treatment heterogeneity is correlated with attrition selection. Because marriage and labor market opportunities are main reasons for migration and attrition of young adults in this context, as well as key outcome variables we study, this is a source of

[9] For example, attrition is comparable to studies tracking a random subsample during intensive tracking and for which effective rather than actual tracking is reported: 1) an 82% effective tracking rate for a 4-year follow-up survey of young adults in Uganda (Blattman, Fiala and Martinez 2014); 2) an 84% effective tracking rate for a 10-year follow-up of the Kenya Longitudinal Panel Survey (Baird et al. 2016); and 3) an 88% effective tracking rate for children after 5–7 years in the Moving to Opportunity evaluation in the US (Orr et al. 2003). Our tracking was less successful, however, than 98% for an 8-year follow-up of a scholarship program in Ghana (Duflo, Dupas and Kremer 2021), who also carried out 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 starting from baseline, including maintaining regular contact with respondents throughout the duration of the study.

[10] Attrition rates from the oversample individuals also are not significantly different than for the rest of the sample from the original short-term evaluation survey and are also equally balanced (p-values of the differences > 0.5).

[11] Variables in Table E2 are those used in IPW construction described below. The variables go beyond those presented in Table A1, to include a set of variables specifically selected for being possible predictors of attrition.

potential selection bias. Our preferred approach to correct for such bias is to apply inverse probability weights (IPW) we describe next.

Construction of Inverse Probability Weights for Attrition Selection Correction[12]

Because several baseline characteristics are correlated with the probability of interview in 2010, at least some of any potential attrition selection is likely related to observables. It also turns out that individuals who were difficult to find, and therefore interviewed during intensive tracking, are more similar on baseline observable characteristics to those never interviewed. For example, administrative data indicate that those more difficult to find or not interviewed had lower compliance rates with the CCT, suggesting that correcting for attrition may be important to account for heterogeneity in ITT differential effects. These patterns are qualitatively similar for females.

To account for potential selection bias due to attrition we calculate inverse probability weights. A typical approach for constructing such weights is to use all sample individuals, regardless of when during fieldwork they were interviewed, and estimate the probability of interview. Although we also estimate such weights (and report sensitivity analyses using them in Table B1, panel F), in our preferred specifications we use a modified version of this typical approach to more fully exploit information from the intensive tracking phase and put higher weight on individuals who were more difficult to interview, since they are more similar to those never interviewed. The approach estimates the probability of interview but only using observations on individuals tracked and interviewed during the intensive tracking phase, i.e., not interviewed during regular tracking. The key assumption underlying this strategy is that the probability of being interviewed 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 calculate IPW separately for each survey instrument and by sex. A large set of socio-economic variables observed in or generated from the program census data are considered for predicting the probability of interview, informed by the nature of migration from the area. These include all baseline variables shown in Table E3 capturing individual-, household- and locality-level characteristics. Since connectedness to the locality is a potential predictor of tracking success, we include detailed measures of household structure, variables capturing social networks of the individual (village size and family network size]), and variables reflecting the possible temporary or transitory nature of residential location for some.[13] We similarly consider locality-level characteristics that could represent push or pull factors for migration: remoteness (measured using distance to night lights and altitude), location in a coffee producing area and having been affected by hurricane Mitch in 1998, which caused residential displacement. Finally, as additional proxy measures for locations more likely to have temporary residents, we introduce variables capturing the level of attrition between the program census (May 2000) and baseline short-term evaluation survey (August 2000): the share of individuals in the locality that attrited and whether any 9–12-year-old (in 2000) attrited. Individuals from localities with higher initial

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

[13] In particular, we include a set of indicators for whether the household comprised temporary workers on one of the large coffee plantations or *haciendas* in the area. Although interviewed in the 2000 program census, these households were not likely to have been permanent residents of the regions.

attrition could be more difficult to trace as contacts with the origin community are more limited. Because there are many variables to consider alongside relatively few individuals not found after the intensive tracking phase, we follow Doyle et al. (2017) to select a reduced set of predictors. This is done for the sample interviewed during intensive tracking or not interviewed. (Construction of the alternative IPW uses the same approach for the full sample.)

First, we estimate bivariate regressions examining whether for each potential predictor there is a significant difference between the means for those interviewed or not. This testing is conducted separately by sex for the early and late treatment groups and for the household- and individual-level instruments. All estimates use the long-term sample survey weights and standard errors are clustered at the locality level.

Second, we retain all variables found to be statistically significantly different for either early or late treatment. We then carry out a preliminary estimate of the probability of having been interviewed during intensive tracking on this set of baseline predictor variables for each treatment group, with strata and regional fixed effects, as well as 6-month age dummies included as fixed predictors in all models. The set of predictors is then reduced by conducting stepwise selection of variables with backward elimination and using adjusted R^2 as the information criterion. Third, we estimate the probability interview during intensive tracking for both early and late treatment combined, keeping predictors retained by the stepwise procedure, and fixed predictors, all interacted with the treatment variable. Table E4 presents these linear probability models, which have good predictive power for both sexes.

The probability of having been found during intensive tracking is then re-estimated via probit using the specifications in Table E4. Observations interviewed during intensive tracking are assigned a weight of 1/Prob and those interviewed during regular tracking a weight of one. Last, the weights are then multiplied by the long-term sample survey weights.

Final IPW for men range from 1–35 for the household instrument and 1–86 for the individual, with corresponding averages of 3.4 and 3.9. For women, the final IPW range from 1–32 for household and 1–89 individual, both averaging about 4. Despite these wide ranges, the overall distribution of weights is not highly skewed and there are only a small number of outliers; 97% of household and 95% of individual weights are < 2 for men and < 9 for women.[14]

In summary, in our preferred specifications we overweight individuals interviewed during the intensive tracking phase, as these individuals were, by definition, more difficult to find, and therefore a priori more likely to be similar to those not found. 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.

To examine sensitivity, we also constructed the more typical IPW based on predictions of the probability of interview using the entire sample (both regular and intensive tracking). We follow a parallel process for covariate selection in estimation of weights for this alternative IPW.[15]

[14] Two observations for each sex have weights > 32 and omitting them does not alter results. With the exception of those outliers, the distribution of weights is similar after calculating IPW using the full sample.

[15] The distribution of weights constructed from the approach using full sample is similar to that used above.

TABLE E1: ATTRITION AND TRACKING IN THE EARLY AND LATE TREATMENT GROUPS, 9–12 COHORT BY SEX

Panel A	Probability of interview: *Household-level instrument*			
	(1)	(2)	(3)	(4)
	Interviewed	Interviewed during regular tracking	Interviewed during intensive tracking (intensive tracking subsample)	Interviewed including proxy information
Males				
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	1,138	1,138	297	1,138
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
	1,062	1,062	444	1,062

Panel B	Probability of interview: *Individual-level instrument*		
	(1)	(2)	(3)
	Interviewed	Interviewed during regular tracking	Interviewed during intensive tracking (intensive tracking subsample)
Males			
ITT	0.008	0.017	0.004
	(0.037)	(0.044)	(0.071)
Mean late treatment	0.793	0.455	0.62
N	1,138	1,138	611
Females			
ITT	-0.014	0.011	-0.034
	(0.041)	(0.039)	(0.067)
Mean late treatment	0.786	0.444	0.615
N	1,062	1,062	585

Notes: Outcome data source: 2010 long-term evaluation survey used to determine when and if interviewed. Males or females 9–12 years old at the start of the program in November 2000. The intensive tracking subsample includes all living individuals not interviewed during regular tracking. Standard errors clustered at the locality level are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

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

	Baseline Sample: 9–12 Cohort Males		Household Instrument Outcomes				Individual Instrument Outcomes			
			Found During Regular Tracking		Final Sample		Found During Regular Tracking		Final Sample	
	N=1,138		N=826		N=1,006		N=527		N=907	
	ITT Coef	SE	ITT Coef	SE	ITT Coef	SE	ITT Coef	SE	ITT Coef	SE
Dependent variable	(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 (‘000 meters)	-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)
Permanence of Residence in Village										
Owns house (=1)	-0.052	(0.043)	-0.021	(0.046)	-0.064	(0.043)	-0.008	(0.041)	-0.048	(0.041)
House in exchange 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 & rent house (=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: 2000 program registry census information (except night lights). Regressions include males 9–12 years old at the start of the program in November 2000, weighted to account for sampling. Table shows coefficient and standard error on ITT indicator controlling for strata fixed effects. Distance to night light (meters) is linear distance from household to an area with stable night light detected by satellite (DMSP-OLS Nighttime Lights). Address in hacienda is indicator for households whose address refers to a large plantation. 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 INTERVIEW, BASELINE COVARIATES AND TREATMENT, 9–12 COHORT, MALES

	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 ages 0–8	0.018	(0.012)	-0.058**	(0.022)	0.108*	(0.055)
Number of children ages 9–12	0.023	(0.015)	-0.071*	(0.039)	0.111	(0.068)
Household Characteristics: Economic Activities & Assets						
Household head main occupation is agric (=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)
Logarithm 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)
Predicted logarithmic per capita expenditures	-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)
Lives in Tuma region (=1)	-0.123***	(0.032)	-0.013	(0.042)	0.017	(0.018)
Lives 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)
Permanence of Residence in Village						
Owns house (=1)	0.146**	(0.064)	-0.080	(0.091)	0.057	(0.082)
House 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 & rent house (=1)	-0.289**	(0.14)	0.231	(0.16)	-0.026	(0.029)

Notes: Outcome data source: Outcome data source: 2000 program registry census information (except night lights). Regressions include males 9–12 years old at the start of the program in November 2000, weighted to account for sampling. For each baseline variable *X*, the table presents coefficient estimates across the row of a linear probability model: . 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 INTERVIEW, BASELINE COVARIATES AND TREATMENT, 9–12 COHORT, FEMALES

	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 has 3 plus grades attained (=1)	-0.0166	(0.047)	-0.0204	(0.074)	-0.0001	(0.047)
Mother has no grades attained (=1)	0.0610	(0.050)	0.0566	(0.067)	-0.0287	(0.058)
Mother has 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 in years	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 ages 0–8	-0.0076	(0.012)	-0.00673	(0.020)	0.0087	(0.057)
Number of children ages 9–12	-0.0367	(0.034)	0.00138	(0.061)	-0.0082	(0.10)
Household Characteristics: Economic Activities & Assets						
Household head main occupation in 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)
Logarithm 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)
Predicted logarithmic per capita expenditures	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)
Permanence of Residence in Village						
Owns house (=1)	0.175**	(0.074)	-0.0368	(0.089)	0.0294	(0.093)
House 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: See Table E3-A. Regressions include females 9–12 years old at the start of the program in November 2000, weighted to account for sampling. *** p<0.01, ** p<0.05, * p<0.10.

TABLE E4-A. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF INTERVIEW DURING INTENSIVE TRACKING PHASE, 9–12 COHORT, MALES

	Household Instrument		Individual Instrument	
	Coef. (SE)	Coef. * T Interaction (SE)	Coef (SE)	Coef. * T Interaction (SE)
ITT (=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 has no grades attained (=1)	0.159*	0.022		
	(0.094)	(0.13)		
Mother has 3+ grades attained (=1)			-0.032	-0.040
			(0.061)	(0.078)
Household head has 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 0–8 years old	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 lights (km)			0.007*	-0.009
			(0.004)	(0.007)
Tuma region (=1)	-0.283***	0.379***	-0.276***	0.246

	(0.082)	(0.11)	(0.079)	(0.15)
Madriz region (=1)	0.179	-0.080	0.162**	-0.257**
	(0.13)	(0.19)	(0.070)	(0.10)
Village population size	0.001	0.001	0.000	0.001*
	(0.001)	(0.001)	(0.000)	(0.001)
Owns house (=1)	0.371**	-0.331*		
	(0.15)	(0.18)		
House obtained in exchange for service/labor (=1)	0.407***	-0.356*		
	(0.12)	(0.21)		
Address in hacienda & rent house (=1)	-0.231	0.427	-0.204	0.306*
	(0.23)	(0.27)	(0.16)	(0.18)
Probability of attrition prior to baseline survey in locality			-0.133	-0.475
			(0.31)	(0.97)
No 9–12-year-old males attrited prior to baseline survey in locality (=1)	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	

Notes: Outcome data source: 2010 long-term evaluation survey interview status. Regressions include males 9–12 years old at the start of the program in November 2000 for sample not interviewed during regular tracking, weighted to account for sampling. The first column shows the coefficient of variable alone, the second column the coefficient of the variable interacted with ITT. 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 INTERVIEW DURING INTENSIVE TRACKING PHASE, 9–12 COHORT, FEMALES

	Household instrument		Individual instrument	
	(a) Coef (s.e.)	(b) Coef * T interaction (s.e.)	(a) Coef (s.e.)	(b) Coef * T interaction (s.e.)
ITT (=1)	0.249		-0.0501	
	(0.30)		(0.36)	
Household head has no grades attained (=1)	0.128**	-0.112	0.117**	-0.000
	(0.062)	(0.082)	(0.049)	(0.082)
Mother has 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 (years)	-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)		
Logarithm 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 ('000 meters)	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)
Tuma region (=1)	-0.130	0.0761	-0.121**	-0.00271
	(0.095)	(0.14)	(0.059)	(0.099)
Madriz region (=1)	-0.0254	0.0456	0.120	-0.0302
	(0.15)	(0.18)	(0.073)	(0.11)
Village population size	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 9–12-year-old females attrited prior to baseline survey in locality (=1)			-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	

Notes: See Table E4-A. Regressions include females 9–12 years old at the start of the program in November 2000 for sample not interviewed during regular tracking, weighted to account for sampling. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX F: EXPLORATORY FACTOR ANALYSIS FOR SOCIO-EMOTIONAL OUTCOMES

We administered two standardized instruments to measure socio-emotional outcomes, the Strength and Difficulties Questionnaire (SDQ) and the Center for Epidemiologic Studies Depression Scale (CESD) Scale. The SDQ is a self-reported behavioral screening test consisting of 25 questions aimed at measuring a set of positive and negative behaviors. The CESD Scale (Radloff 1977) is a commonly used mental health scale developed as a screening test for depression and depressive disorder, consisting of 22 questions asking the frequency of both positive and negative self-perceptions. Both tests were available in Spanish. We examined the internal consistency of the different scales separately for the samples of males and females in the 9–12 cohorts using exploratory factor analysis.

For males, the overall Cronbach's alpha for all 25 items in the SDQ together indicated the overall scale was internally consistent (0.70). The alphas were much lower, however, after organizing items into the five standard subdomains of SDQ (emotional symptoms, conduct problems, hyperactivity, peer relationships and pro-social behavior), and ranged from 0.26–0.51, lower than the usual thresholds for statistical validity. Exploratory factor analysis on the 25 items suggested there were only two latent factors (i.e., two factors with eigenvalues above one and supported by the scree plot). Moreover, imposing a five-factor structure did not lead to items grouping along the five standard subdomains, with the first factor having high loads on items from three of the five subdomains.[16] For the CESD, the Cronbach's alpha for all 20 items was also high (0.83); exploratory factor analysis pointed to only one or two distinct factors and did not allow finer differentiation.

Results for females were similar. The overall Cronbach's alpha for all 25 items in the SDQ together indicated that the overall scale was internally consistent (0.72). The alphas were much lower, however, after organizing items into the five standard subdomains, from 0.21–0.54. Exploratory factor analysis on the 25 items yielded the same conclusions as for males. For the CESD, the Cronbach's alpha for all 20 items was high (0.86) but as for the males, exploratory factor analysis pointed to only one or two distinct factors and did not allow finer differentiation.

Because the results from exploratory factor analyses suggested responses should not be pooled using the standard subdomains for SDQ for either males or females, we constructed new indices based on both the SDQ and CESD to characterize the relevant latent traits. Following Cunha, Heckman and Schennach (2010), Attanasio et al. (2020) and others, separately for each sex we pooled together all questions from both the SDQ and CESD scales and identified the latent socio-emotional traits in the sample. Based on both eigenvalues and 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 sex, two with high factor loads on items from SDQ and two with high loads on items from CESD. Notably, questions referring to positive (and separately negative) attitudes or behavior loaded onto similar latent factors in each of the scales. Examination of the factor loadings of the different items all together pointed to a plausible interpretation of the retained four factors as capturing optimism, positive self-perception, negative self-perception and stress.

---

[16] Laajaj and Macours (2021) report similar findings for other measures of socio-emotional outcomes when scales originally designed for high-income countries were used in lower-income settings.

Appendix G: Non-Experimental Difference-in-Differences (DD) Absolute Program Impacts Using National Census Data

To better understand what underlies the experimental differentials after 10 years, we estimate non-experimental difference-in-differences (DD) absolute program impacts after five years. At some point over 2000–5, the 3-year CCT operated in all rural areas of the six municipalities where the evaluation took place, covering ~90% of the rural population.[17] The CCT was the only large education- or health-related program operating in these and adjacent municipalities in the regions during those years. Therefore, it is possible to use national census data to provide suggestive evidence on absolute effects for some schooling and demographic outcomes, recognizing that the findings are less well identified than the experimental results in the main paper.

Specifically, we use the full 1995 and 2005 national censuses and a DD approach to estimate absolute program effects five years after the CCT began in the early treatment group. Together, the two 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. Because the CCT did not operate in urban areas, we limit the census sample to individuals living in rural areas. The censuses provide type (urban or rural) of current residential location and municipality of current residence, of residence five years prior to the census administration date and of residence at birth.[18] 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 some forms of selection bias due to migration than the long-term differentials.[19] We assign all individuals to the municipality where they lived five years prior to each census and assume the type of prior residence (rural or urban) was the same. We calculate DD effects for different age cohorts (calculating ages on November 1, 1990 and 2000, respectively, for consistency with the main analyses using the 2010 long-term evaluation survey) and compare 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 had localities with similar marginality index scores but had never received the CCT. We estimate

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

where Y_{imt} is the outcome for individual i in municipality m measured in census year t, $T_{m,t-5}$ is an indicator for whether the individual 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 DD estimate of the absolute effect of the program on Y, five years after the program began. Because they are based on residential location, all estimates are ITT. Outcomes examined include grades attained, enrollment, literacy, whether ever married and, for women, whether they had had a live birth.

[17] This includes the 42 rural localities randomized into early and late treatment but also the 17 rural localities not initially included because they were less poor according to the marginality index. In the additional 17 localities, the program began in 2001 and was offered to ~80% of the population based on a household-level proxy means targeting model (Maluccio 2009). Consequently, by 2005 the CCT had been implemented (at different times and to modestly different degrees) in all 59 rural localities in the six municipalities, with ~90% of the population eligible.

[18] While each census includes more detailed location information, changes in the boundaries of localities between 1995 and 2005 make it impossible to link localities over time. Municipality boundaries were unchanged.

[19] For the age cohorts under consideration, however, it turns out mobility over this period is relatively limited. For example, only about 3% of the 9–12 cohort moved during the five years previous to each census.

Standard errors are robust to heteroskedasticity. We consider the same age cohorts as in the experimental estimates and assess alternative comparison groups as well as common trends.[20]

We present the DD effects after five years for the 7–8- and 9–12 cohorts by sex in Table G1. With the exception of ever having been married, effects for the 7–8 cohort are similar in magnitude and significance to the 9–12 cohort within each sex, consistent with the CCT having a positive absolute effect on all these children. Boys 9–12 years old at the start of the program (and thus 14–17 years old in 2005) in treatment municipalities have 0.6 more grades than similarly aged boys in comparison municipalities (15% higher than the mean for the comparison group) and are 12 percentage points (22%) more likely to have completed fourth grade (panel B). In addition, they are 9 percentage points (12%) more likely to be literate. Estimates for girls are similar though slightly smaller. Girls in treatment municipalities have an average of 0.5 more grades (11%) attained by 2005, translating into an 11-percentage point (16%) effect on having completed fourth grade, and a 7-percentage point (8%) effect on literacy (panel D). Moreover, in 2005 approximately half of boys and girls in the 9–12 cohort are still enrolled in school (with small estimated absolute effects on enrollment of ~3 percentage points), pointing to the potential for further absolute gains after 2005. While the lack of 2010 long-term experimental differentials in learning for women (and small differentials in schooling) reported in Section V.B could mean the CCT did not lead to any absolute effects on those outcomes in either treatment group, the positive absolute effects after five years suggest it is more likely that gains experienced by the early treatment were offset by equivalently sized impacts in the late treatment by 2010.

For the female 9–12 cohort we also observe absolute effects on early marriage and teenage pregnancy, with reductions of more than 2 percentage points for each. The results represent relatively large absolute effects on early fertility as less than 10% had given birth by 2005. The combination of this negative 5-year absolute effect alongside the 2010 long-term negative experimental differential for fertility suggests there is an absolute reduction in early pregnancy for both groups, with a larger reduction in early treatment, consistent with the late treatment experiencing an earlier age of menarche putting them at higher risk of early pregnancy.

We explore the sensitivity of the DD findings for the 9–12 cohort to different sets of comparison municipalities and definitions for treatment status. First, we expand the comparison municipalities to include rural areas in all non-program municipalities in central regions of Nicaragua where the program was located. Second, we consider whether results differ when we instead use current residence or municipality of residence at birth to determine ITT status. While there are some differences in point estimates, findings (not shown) are robust to the alternatives and point to significant positive absolute effects on educational outcomes for both sexes after five years, and negative absolute effects on fertility and marriage for girls.

Finally, we provide falsification tests of the identifying assumptions, estimating program effects on outcomes for two placebo cohorts unlikely to be affected by the intervention: 1) heads in households of a boy or girl in the 9–12 cohort (Table G2); and 2) 30–34-year-old men or women (Table G3). The same empirical specification reveals no effects on educational outcomes or marriage for either group and point estimates are all close to zero, supporting common trends.

[20] δ_3 is the average ITT effect for the three groups—early treatment, late treatment and the other 17 localities with household targeting—each of which by 2005 had been eligible for the three-year program. It is not possible to isolate the three distinct treatment groups within the census data due to local boundary changes between censuses.

TABLE G1: 2005 NON-EXPERIMENTAL DIFFERENCE-IN-DIFFERENCES ABSOLUTE IMPACTS ON EDUCATION, MARRIAGE AND FERTILITY, 7–8- AND 9–12 COHORTS BY SEX

	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: 7–8 Cohort, Males						
DD	0.479***	0.104***	0.032*	0.084***	-0.001	-
	(0.076)	(0.018)	(0.018)	(0.017)	(0.001)	
N	11,056	11,056	11,068	11,045	11,067	
Mean comparison group in 2005	2.965	0.416	0.709	0.765	0.002	
Panel B: 9–12 Cohort, Males						
DD	0.597***	0.124***	0.037***	0.091***	-0.003	-
	(0.078)	(0.014)	(0.014)	(0.013)	(0.004)	
N	18,399	18,399	18,421	18,403	18,421	
Mean comparison group in 2005	3.922	0.559	0.456	0.779	0.017	
Panel C						
P-value (7–8 vs 9–12)	0.282	0.376	0.833	0.523	0.613	
Panel D: 7–8 Cohort, Females						
DD	0.418***	0.099***	0.028	0.084***	0.000	-
	(0.081)	(0.019)	(0.018)	(0.017)	(0.002)	
N	10,260	10,260	10,271	10,260	10,271	-
Mean comparison group in 2005	3.462	0.525	0.776	0.825	0.005	-
Panel E: 9–12 Cohort, Females						
DD	0.536***	0.105***	0.028*	0.066***	-0.022**	-0.023**
	(0.084)	(0.015)	(0.015)	(0.013)	(0.011)	(0.010)
N	17,061	17,061	17,075	17,061	17,075	14,104
Mean comparison group in 2005	4.542	0.642	0.502	0.830	0.162	0.085
Panel F						
P-value (7–8 vs 9–12)	0.312	0.788	0.996	0.387	0.053	-

Notes: Data source: 1995 and 2005 Nicaraguan national censuses. Regressions include males or females 7–8 or 9–12 years old in November of 1990 (for 1995 census) or 2000 (for 2005 census) living in rural areas of the six program or six comparison municipalities. Information on live births unavailable for 7–8-year-olds (too young at time of census) and missing for 17% of girls in the 9–12 cohort. DD (δ_3) is the DD 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 G2: 2005 ASSESSMENT OF COMMON TRENDS FOR PLACEBO COHORTS: HEADS OF HOUSEHOLDS WITH A CHILD IN THE 9–12 COHORT, BY CHILD SEX

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Head is Female =1	Age of Head in Years
	(1)	(2)	(3)	(4)	(5)
Panel A: Heads of Households with 9–12-Year-old Male					
DD	-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 in 2005	1.673	0.226	0.507	0.221	46.901
Panel B: Heads of Households with 9–12-Year-old Female					
DD	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 in 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 9–12 years old in November of 1990 (1995 census) or 2000 (2005 census) living in rural areas of the six program or six comparison municipalities. DD (δ_3) is the difference-in-differences estimate of the absolute program effect in 2005 on the outcome. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

TABLE G3: 2005 ASSESSMENT OF COMMON TRENDS FOR PLACEBO COHORTS: 30–34-YEAR-OLDS BY SEX

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Ever Married =1
	(1)	(2)	(3)	(4)
Pa*nel A: Males*				
DD	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: Females*				
DD	-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 (1995 census) or 2000 (2005 census) living in rural areas of the six program or six comparison municipalities. DD (δ_3) is the difference-in-differences estimate of the absolute program effect in 2005 on the outcome. Heteroskedasticity-robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

APPENDIX H: INSIGHTS FROM NUTRITIONAL AND MEDICAL LITERATURE ON AGE OF MENARCHE AND IMPLICATIONS FOR THE STUDY

Nutrition and medical literatures provide scientific basis for understanding the role of nutritional status and nutritional shocks for the onset of puberty and age of menarche, a defined marker of puberal maturation. Poor female childhood nutrition is associated with delays and correspondingly better childhood nutrition and health with earlier menarche (Garn 1987; Cooper et al. 1996; INSERM 2007).[21]

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). Numerous studies document the close correlation between body fat mass and the onset of puberty, although it is not entirely clear whether excess body fat induces sexual maturation, or sexual maturation triggers excess body fat (INSERM 2007). The links between body fat and the female reproductive system have been interpreted as biological mechanisms for ensuring that pregnancy cannot occur unless there are adequate fat stores to sustain mother and fetus. Supporting this, studies with rodents indicate mammalian females are able to turn off their reproductive systems when food intake is inadequate (Kaplowitz 2008).

Nutritional intakes during childhood, in turn, is believed to affect body fat mass and thus menarche through the leptin hormone, which helps regulate the body's energy balance. When fat stores are low, decreased leptin leads to increased appetite, helping to restore body fat and body weight. There is some uncertainty whether leptin plays a permissive or triggering role (Shalitlin and Philip 2003; INSERM 2007). Leptin levels fluctuate with nutritional intakes, pointing to the channel through which nutritional shocks translate into relatively immediate delay or acceleration of menarche.[22]

The evidence also indicates that both positive and negative nutritional shocks not only can have immediate, but also particularly large effects on sexual maturation for girls who experienced undernutrition earlier in life (INSERM 2007). The well-documented phenomenon of early onset puberty among migrant and adopted children (Mul et al. 2002; Parent et al. 2003) is often explained by an interaction of prenatal undernutrition with a subsequent enriched nutritional context in later childhood (Gluckman and Hanson 2006). Sloboda et al. (2007) find that low birthweight and subsequent BMI when 8 years old predicted early age at menarche. Koziel and Jankowska (2002) show that among females with high BMI at age 14, those who had also had low birthweight were more likely to have had early menarche.[23] The medical literature does not provide firm conclusions regarding specific ages at which positive or negative nutritional shocks might matter most. A review by Kaplowitz (2008), for example, points to the

[21] Age of menarche declined from 17 to 14 years in the US 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).

[22] Laboratory experiments with rodents and monkeys show that increased leptin leads to immediate changes in the age of menarche (Chehab et al. 1996; Chehab et al. 1997; Cheung et al. 2001; Wilson et al. 2003).

[23] More generally, poor prenatal combined with compensatory postnatal growth during childhood is also thought to lead to adverse health outcomes (Hales et al. 2003). Although we do not study this possibility, it is an example of how later investment has possible negative consequences depending on prior circumstances (dynamic substitution).

need for further study of the relationship between nutrition and puberty from as early as 6–7 years old and up to 13–14 years old. Consistent with this wide age-range, both cross-sectional and longitudinal evidence indicate rises in leptin concentration for females 7–15 years old, that parallel increases in body fat during puberty—as well as strong correlations between higher leptin, greater body fat mass and lower age of menarche (Blum et al. 1997; Garcia-Mayor et al. 1997; Ahmed et al. 1999).

Implications for the study—Based on the above evidence from the medical and nutrition literature, in Section V.A we hypothesize that differential timing of nutritional shocks in early versus late treatment could have led to differences in nutritional status, the onset of puberty and age of menarche, and consequently in adult female fertility and related behaviors. As the 9–12-year-old cohort was reaching puberty at the start of period 2 the positive nutrition shock in late treatment may have accelerated it, while the probable reduction of nutritional inputs for the early treatment (resulting from the cessation of transfers) could have had the opposite effect. Earlier sexual maturation in late treatment (relative to early treatment), then, could have led to relatively earlier fertility.

Both experimental treatments received the CCT. Therefore, it likely improved the nutritional status of girls (at least during program operations), possibly reducing age of menarche in both treatment groups. The direction of potential absolute effects on related 2010 outcomes like sexual debut, fertility and marriage, however, are ambiguous given the other program components. Notably the education transfers and the positive shock to household wealth, by inducing girls to stay in school longer, were likely to have reduced early fertility. Such an effect would have worked in the opposite direction of the possible reduction in the age of menarche described above. The non-experimental DD effects on marriage and fertility after five years reported in Appendix G suggest that the overall effect (for both treatment groups together) was a reduction in early fertility.

We compare the 2010 long-term evaluation survey with the 2011–12 Nicaraguan Demographic and Health Survey (NDHS) to provide further cross-sectional evidence. This can be done for sexual debut, marriage, fertility and BMI but unfortunately not for age of menarche which was not asked in NDHS. We compare 2010 outcomes (including the distribution of BMI) for the 9–12 cohorts in early and late treatment groups with same-age women from NDHS. For a comparable sample we restrict NDHS to young women 19–22 years old living in rural areas of the two departments of Nicaragua where the program operated (Madriz and Matagalpa) but excluding the municipalities with the program. Compared to women in the NDHS subsample, women exposed to the CCT are 10+ percentage points less likely to have had their first sexual encounter by age 15, 4+ percentage points less likely to be married and 10+ percentage points less likely to have had a child.[24] The differences—based on simple cross-sectional non-experimental comparisons—are consistent with the possibility that the CCT led to absolute reductions in key fertility-related behaviors in both early and late treatment groups, in line with the negative non-experimental DD in Appendix G.

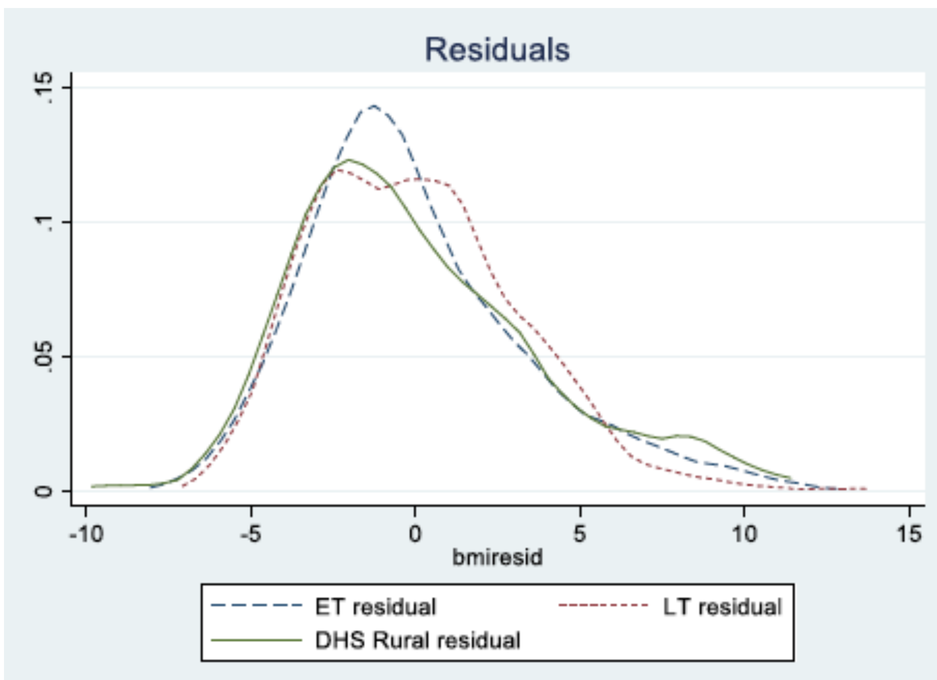
While the CCT may have improved nutrition (and thus BMI) in both treatments, it also may have influenced BMI indirectly through effects on age of menarche or subsequent fertility. Because prior pregnancy influences BMI, we condition on whether a woman had ever been

[24] Moreover, consistent with the positive estimated non-experimental DD absolute effects of the CCT on schooling reported in appendix G, average completed grades in late treatment are higher than for NDHS (6.8 versus 6.3).

pregnant and on age, and then compare the distributions of the predicted residuals for the three samples (Figure H1). Both treatment group distributions appear to match the NDHS distribution closely in the left tail but have visually different shapes moving to the right. This suggests the possibility that BMI (conditional on fertility) increased in the middle range of the sample distribution for both treatment groups, and more so in late treatment. Kolmogorov-Smirnov tests indicate that the early and late treatment group distributions are significantly different (p=0.075). Tests of the other comparisons (early treatment versus NDHS or late treatment versus NDHS) are not significant (p-values > 0.5) for these samples, however, possibly due to the relatively small NDHS subsample size [N=140].

Recognizing the substantial limitations of these simple comparisons, the above patterns are consistent with both early and late treatments having led to an improvement in nutritional status, with greater improvements in the late treatment group. These improvements likely partly offset reductions in early fertility resulting from the other program components.

FIGURE H1: CONDITIONAL COMPARISONS OF BMI FOR WOMEN 19–22 YEARS OLD ACROSS EARLY-TREATMENT, LATE-TREATMENT AND NDHS RURAL MADRIZ AND MATAGALPA



Notes: Data sources: 2010 long-term evaluation individual survey instrument survey and 2011–12 Nicaraguan Demographic and Health Survey sample described in the text. 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 I: COST-BENEFIT CALCULATIONS

A full cost-benefit analysis of the Nicaraguan CCT requires monetary assessment of all resource costs and all potential benefits of the multi-faceted program. This would require estimates of absolute effects of the program not only for the cohorts studied, but also for individuals of other cohorts and for outcomes not examined in the paper, as well as any of the associated costs. A full analysis, therefore, is beyond the scope of the paper. Instead, we carry out a partial, or modified cost-benefit analysis to approximate the net present value (NPV) of the program using the estimated differential earnings for the cohort of boys 9–12 years old at the start of the program in 2000, treating them as a lower bound of the absolute program effects on earnings. Below we outline our approach and describe why it likely provides lower bound estimates of the NPV for the 7–12 cohort as a whole.

All cost and benefit flows are first converted into US dollars using market exchange rates. Then, using the US consumer price index, the flows are deflated into constant dollars in 2000, the first year of the program. Finally, flows are discounted to calculate the NPV in 2000. As recommended for cost-benefit analysis of transfer programs, we include direct costs of implementing the program (including targeting, monitoring, transfer delivery, management and any other administrative costs) but exclude the value of the transfers (Dhaliwal et al. 2012). We also exclude all evaluation costs. We approximate the implementation costs of program components related to the 7–12 cohort in the early treatment group to be 50% of the average per household program costs (Caldés and Maluccio 2005). In nominal terms, these amounted to $60 per year for three years. This approach conservatively assumes the other 50% were for the nutrition and health program components that included among other things the privately contracted comprehensive health services for younger children (Regalia and Castro 2007). Benefits are based only on the estimated average differential increase in monthly earnings for males in the 9–12 cohort and are assumed to have started when we measured them in 2009 (via interviews in 2010 with a 12-month reference period). The increased earnings ($10) are multiplied by the average number of months worked off farm (3.5) and the average number of males 9–12 years old in each household at baseline in 2000 (1.37). As indicated above, US dollar values are first deflated to year 2000 constant dollars and then discounted. Using a 5% discount rate, the NPV turned positive in 2016; at 10%, NPV turns positive in 2027, approximately 25 years after the program ended in the early treatment group.

The above approach does not account for deadweight losses associated with the taxation necessary to raise revenue for the program (including the transfers) or for the possibility that the CCT crowded out other potential government expenditure with higher returns (Harberger 1997). Assuming households in early treatment received the full amount of the education transfers all three years and that the cost of raising the revenue to pay for them was 5%, the NPV with a 10% discount rate would take an additional eight years to turn positive, in 2035.

If there were negative absolute effects of the CCT on outcomes in either the early or late treatment groups our use of differential effects to measure program benefits raises the possibility that we overstate those benefits. Although without a randomized control it is impossible to be certain there were no such negative effects on important outcomes, we argue that the design of the program—offering a wide range of benefits—is such that substantive negative absolute effects are unlikely, especially for the cohort we study. Even for females for whom results indicate negative differential effects on BMI, for example, both early and late treatment groups received program components and it is reasonable to assume the absolute impact on nutritional status for each treatment group on its own was positive or at worst non-negative. This assumption is supported by comparison of BMI measurements in the study with those from the Nicaraguan DHS reported in Appendix H.

Finally, for several reasons it is likely that in practice the methodology we use overestimates costs and underestimates benefits so that at each discount rate the calculated NPV is a lower bound.

Regarding costs, first we include all the education component costs. In addition to costs for males in the 9–12 cohort, these include costs for younger males in the 7–8 cohort, as well as all for females in the 7–12 cohort. Second, we ignore program costs in the late treatment group, which would offset the costs of the program in early treatment in the differential framework.

Regarding benefits, in addition to using differential estimates which may understate absolute earnings gains, several other aspects of our approach make it likely they are underestimated. First, although we include the costs as described above, we do not incorporate potential benefits for the younger males, or for the entire cohort of females. Results in the paper for the younger males and for the females point to positive effects on labor market outcomes for these other groups as well. Because they are not all statistically significant, however, we take a conservative approach and do not include them as benefits. Second, we consider only one type of benefit—increased off-farm earnings. This captures benefits from additional schooling, learning or other possible effects only to the extent that they operate directly through the off-farm labor market. Third, we consider benefits starting in 2009, but for some (e.g., older) in the cohort benefits likely began earlier. Fourth, young men (as well as women) in the early treatment group were more likely to still be enrolled in school in 2010, perhaps leading to even higher earnings differentials when they completed their schooling. And fifth, earnings differentials are measured early in their working lives and thus plausibly increase over time.

In summary, under relatively conservative assumptions regarding cost and benefit flows related to the CCT, estimates in the paper suggest that program related activities for the 7–12 cohort would achieve positive NPV within 2–3 decades. Less conservative assumptions would shorten the payback period.

APPENDIX J: ADDITIONAL REFERENCES

Below we provide appendix references not already cited in the main text.

Ahmed M.L., K.K. Ong, D.J. Morrell, L. Cox, N. Drayer, L. Perry, M.A. Preece and D.B. Dunger. 1999. "Longitudinal study of leptin concentrations during puberty: sex differences and relationship to changes in body composition." *Journal of Clinical Endocrinology Metabolism* 84(3): 899–905.

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.

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

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

Chehab, F.F., M.E. Lim and R. Lu. 1996. "Correction of the sterility defect in homozygous obese female mice by treatment with human recombinant leptin." *Nature Genetics.* 12(3): 318–20.

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

Cheung, C.C., J.E. Thornton, S.D. Nurani, D.K. Clifton and R.A. Steiner. 2001. "A reassessment of leptin's role in triggering the onset of puberty in the rat and mouse." *Neuroendocrinology* 74(1): 12–21.

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.

Frisch, R.E. and R. Revelle. 1971. "Height and Weight at Menarche and a Hypothesis of Menarche." *Archives of Disease in Childhood* 46(249): 695–701.

Garcia-Mayor, R.V., A. Andrade, M. Rios, M. Lage, C. Dieguez and F.F. Casanueva. 1997. "Serum Leptin Levels in Normal Children: Relationship to age, gender, body mass index, pituitary-gonadal hormones, and pubertal stage." *Journal of Clinical Endocrinology Metabolism* 82(9): 2849–55.

Hales, C.N. and S.E. Ozanne. 2003 "The Dangerous Road of Catch-up Growth." *Journal of Physiology* 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: 73–9.

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

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.

Kling, J.R. and J.B. Liebman. 2004. “Experimental analysis of neighborhood effects on youth.” Princeton University, Princeton New Jersey.

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

Kugler, A.D. and I. Rojas. 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 Report.

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

Radloff, L.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.

Shalitin, S. and M. Phillip. “Role of obesity and leptin in the pubertal process and pubertal growth: a review.” 2003. *International Journal of Obesity* 27(8): 869–74.

Sloboda, D.M., R. Hart, D.A. Doherty, C.E. Pennell and M. Hickey. 2007. “Age at menarche: influences of prenatal and postnatal growth.” *J. Clin. Endocr.* 92(1): 46–50.

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

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

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