

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327248586>

Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes

Research · August 2018

CITATION

1

READS

48

3 authors, including:



Tania Barham

University of Colorado Boulder

22 PUBLICATIONS 281 CITATIONS

[SEE PROFILE](#)



Karen Macours

Ecole d'économie de Paris

62 PUBLICATIONS 1,019 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



The Validity and Reliability of Skills Measurement in Rural Household Surveys [View project](#)



Evaluation of productive safety net program in Nicaragua [View project](#)

Experimental Evidence of Exposure to a Conditional Cash Transfer During Early Teenage Years: Young Women's Fertility and Labor Market Outcomes

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

August 24, 2018

Abstract: Conditional cash transfer (CCT) programs are one of most popular policy instruments for increasing investment in nutrition, health, and education in developing countries. For teenage girls, CCTs not only provide incentives and means to remain in school longer, but also may affect fertility outcomes through improved nutrition (with implications for the onset of puberty) or provision of reproductive healthcare information. Therefore, examining the fertility mechanism is crucial for understanding long-term impacts, in particular labor market outcomes, as young women's decisions regarding economic, education, and reproductive activities are closely linked. This paper exploits an experimental design and a survey implemented 10 years after the start of a CCT program in Nicaragua that introduced random variation in program exposure during the early teenage years, ages critical for sexual maturity. Differential exposure to the CCT does not lead to long-term differences in grades attained or learning, but does lead to differential impacts on the age of menarche, young adult BMI, fertility, and subsequent labor market outcomes and income.

JEL Codes: I25, I38, I18, J13

Key words: CCT, long-term effects, education, fertility, labor markets

* 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). Acknowledgments: We thank Alessandro Tarozzi and participants at presentations at PSE for related discussions and ideas. 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). We are indebted to Veronica Aguilera, Enoe Moncada, and the survey team from CIERUNIC for excellent data collection and for their dogged 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 Millan, Olga Larios, Jana Parsons, and Gisella Kagy for help with data preparation and Vincenzo di Maro for numerous contributions to the 2010 survey. All remaining errors and omissions are our own.

I. Introduction

Conditional cash transfer (CCT) programs are one of the most popular policy instruments for increasing investment in human capital. They reach approximately 25 percent of the population in Latin America (Robles, Rubio, and Stampini 2015), and in recent years have expanded rapidly in Africa and Asia. The first generation of these programs promoted childhood nutrition, health, and education, and targeted children throughout different stages of their lives.¹ Some CCTs also directly incorporated preventive and reproductive healthcare components for teens. A principal motivation underlying these programs is that investment in human capital will help improve the lives of the poor in the longer term, with the expectation that it can lead to higher income generating potential in adulthood for beneficiary children.

While the evidence on long-term impacts of CCTs is growing, with much of it demonstrating improvements in schooling, results for labor market outcomes are mixed and inconclusive (Molina Millán et al., 2018a). This is due in part to methodological challenges, but also may be because impacts differ across settings and target populations, reflecting likely heterogeneous constraints to income generating activities. Moreover, CCTs offer a relatively complex set of benefits potentially affecting human capital accumulation at different points during childhood and in different ways. As a result, the mechanisms through which these programs affect long-term outcomes are likely multi-faceted. This may be particularly relevant for understanding impacts on girls, for whom decisions regarding education, reproductive, and economic activities are often linked, with the consequence that different components of CCT programs (nutrition, health, education, and reproductive health) are all likely to affect their outcomes as young adults. To date, the longer-term impacts of early teenage exposure to the nutritional and reproductive health components of CCTs has received relatively little attention in comparison to exposure to the education components. This is an important gap because such components may be as important as schooling when considering later labor market effects for women.

To address this gap in the literature, we collected data 10 years after the start of a randomized CCT program in Nicaragua to examine how nutrition, education, and reproductive health shocks during early teenage years affected later outcomes for girls. We estimate differential intent-to-treat (ITT) effects by exploiting the randomized phase-in of the program which provides experimental variation in the timing of exposure—households assigned to the early treatment group were eligible for transfers from 2000 to 2003, while those assigned to the late treatment group were eligible from 2003 to 2005. The CCT not only provided incentives to remain in school longer, but also may have affected girls through improved food availability and nutrition (with possible implications for the onset of puberty) and exposure to reproductive healthcare information during potentially critical early teenage years. Each of these pathways potentially influence later fertility and marriage outcomes, and thus labor market outcomes. Since both the early and late treatment groups eventually received the program, there is no long-term experimental control group and we cannot estimate experimental absolute long-term impacts for the program. Instead, we estimate differential impacts of having been exposed at different ages in the early versus the late treatment groups. This approach has the advantage of enabling an assessment of whether exposure at different points in a child’s life are critical for later long-term effects, providing insight into the possible causal pathways.

¹ More recent programs, however, are increasingly being designed to focus more narrowly on a single particular objective (e.g., nutrition or health or education) or target group.

We examine the program effects on a cohort of girls for whom the phase-out of transfers in the early treatment group, and the phase-in of transfers in the late treatment group occurred during teenage years around the typical onset of menarche. In particular, we focus on girls aged 9-12 at the start of the program, who were exposed to the nutrition, health, and education transfers for the three subsequent years if they lived in localities in the early treatment group. The same cohort in the late treatment group was exposed to the program nearly three years later, starting when the girls were 11-14. Reproductive health information sessions were added in the third year of program implementation, and as a result the late treatment girls had greater exposure to that information.

The medical and nutritional literature highlights that the different ages at which girls in the early and late treatment areas were exposed to the transfers may have implications for the onset of puberty,² potentially affecting later fertility and labor market outcomes. Several studies demonstrate that poor childhood nutrition is associated with delayed puberty for girls, and conversely that better childhood nutrition and health are associated with earlier menarche (Garn, 1987; Cooper et al 1996, INSERM 2007). Moreover, improved nutrition at different stages of childhood may affect the onset of puberty differentially depending on prior nutritional status. Those who were previously undernourished can experience more accelerated sexual maturity with later good nutrition.³ Conversely, delays in age of menarche can occur when girls suffer a sudden deterioration in nutritional intake (and related hormonal changes) (INSERM, 2007). Finally, numerous studies document a close correlation between body fat mass and the onset of puberty for girls, even if it is not necessarily clear whether excess weight induces early sexual maturity, or vice versa (Blum et al. 1997; INSERM, 2007).

We examine outcomes measured ten years after the start of the program, and five years after households eligible for the program in the late treatment group stopped receiving transfers. By that time, the young women were 19-22 years old with about half working, half married or with a child, and a third still in school. Therefore, the study measures differential impacts while many of the young women were in transition from one lifecycle stage to the next. Although this complicates interpretation of some of the findings, evidence during this key transition is crucial for understanding the mechanisms underlying the long-term outcomes of CCTs. This is particularly salient for female labor market outcomes because, as is the case in many developing countries, in Nicaragua labor force participation of young women (49 percent in the sample) is much lower than that of men (98 percent). Therefore, understanding whether and how early reproductive health outcomes relate to labor market entry is of particular importance.

To establish the differential long-term program effects, we use a 2010 follow-up survey of individuals and households originally interviewed prior to the program start in 2000. Substantial resources were dedicated to minimizing attrition, given the potential (but a priori unknown) relationship between program exposure and migration for such a mobile cohort of young adults. Final attrition rates are relatively low (ranging from 6 to 22 percent depending on the outcome) and balanced between early and late treatment groups, and yield analysis samples that are balanced on baseline observables. As those who migrated but could not be found may still differ

² See appendix I for a more detailed review of the relevant medical and nutrition literature.

³ The non-linear effects of undernutrition depending on the age of exposure are thought to explain, for instance, the high frequency of early puberty among migrant and adopted children (Mul, Oostdijk and Drop, 2002; Parent et al 2003; Gluckman and Hanson (2006). While direct evidence on early childhood nutritional status for the girls in our sample is unavailable, nearly 50 percent are stunted as young women in 2010, and a younger cohort from the same localities (under age three in 2000) had a 40 percent prevalence of stunting, both indicating high undernutrition.

from those found, we consider a variety of complementary estimation strategies to gauge the importance of any remaining selection. In addition, we account for multiple hypothesis testing throughout by grouping outcomes in families, and test the familywise error rate following Anderson (2008). We further test the robustness of the inference using randomization statistical inference to calculate exact p-values for sharp null hypotheses, as recommended by Athey and Imbens (2017) and Young (2017).

After ten years, girls randomly exposed to the CCT in the early treatment group starting when they were 9-12 years old are more likely to be economically active and have higher earnings than girls exposed three years later in the late treatment group. Early treatment girls reached sexual maturity later and had lower BMI compared to late treatment girls. There are also strong differentials in reproductive health outcomes, with early treatment girls starting sexual activity later, resulting in lower overall fertility outcomes when they are 18-21 years old. The differential effect on highest grade attained, however, is insignificant, though there is an effect on completing fourth grade of primary school—the highest grade eligible for transfers. In addition, there are no differential effects on learning. The modest differential impacts on educational outcomes may in part be explained by enrollment patterns, with girls in this context typically still in school during their early teenage years, allowing late treatment girls to also benefit from the program. This evidence suggests that the long-term impacts of the CCT program on labor market outcomes for girls in part reflect changes in reproductive health outcomes, which plausibly were affected directly by the nutrition and the reproductive health components of the program. Impacts on the age of menarche, BMI, and past attendance at the CCT-run reproductive health information sessions further support the likely importance of the reproductive health mechanisms.

The modest long-term differential effects on education and small and insignificant results for learning are consistent with a pattern of limited absolute effects on these outcomes for both treatment groups or, alternatively, larger, but similarly sized absolute effects in the two groups that offset each other. We use two non-experimental strategies to estimate the long-term absolute impacts of exposure to the CCT, both of which provide support for larger (but offsetting) absolute effects. Double difference estimates of the absolute 5-year impact of the program using national census data show absolute gains in education and reductions in early pregnancy and marriage due to the program. Propensity score matching that takes advantage of a separate comparison group that was incorporated into the original evaluation and never received the program further suggests positive and substantial absolute 10-year impacts on education and learning.

This paper complements the evidence on boys in the same age cohort analyzed in a companion paper (Barham, Macours, Maluccio, 2018). While the earnings and labor market participation results are similar between the two sexes, effects on the possible underlying mechanisms are not. Consistent with boys typically dropping out of school at younger ages than girls in this context, there are clearer positive differential impacts on both educational and learning outcomes for boys, suggesting that the causal pathway leading to differential impacts on labor market outcomes for them come primarily through the educational gains. The results for the boys are in contrast to the somewhat more complex pathways for girls presented in this paper, where the nutritional and reproductive health impacts during teenage years also appear to be important for understanding the longer-term impacts of cash transfer programs.

More generally, this paper relates to Field and Ambrus (2008), whose work exploits cross-sectional variation in the age of menarche to demonstrate that later age of menarche is related to

higher schooling and later marriage. In addition, it complements research examining the longer-term effects of cash transfers, scholarships, or educational subsidies on education, fertility, and labor market outcomes for young women (Behrman, Parker, and Todd, 2009, 2011; Filmer and Schady, 2014; Molina Millán et al, 2018b; Parker and Vogl, 2018). In this literature, impacts on fertility outcomes are often directly linked to the educational outcomes and incentives, because inducing girls to remain in school longer can postpone childbearing (Duflo, Dupas, and Kremer 2015, 2017; Baird, McIntosh, and Özler, 2011, 2018). Relatedly, Buchman et al. (2017) show that financial incentives for delaying marriage also can translate into increased schooling. Our paper contributes to the literature by highlighting that the nutrition and health related components of transfer programs are additional channels through which fertility outcomes of teenagers may be affected, which is important for later labor market outcomes.⁴

II. The Nicaraguan *RPS* CCT Program and Its Experimental Design

A. Key program design features of the CCT

Modelled after *PROGRESA*, the *Red de Protección Social (RPS)* was a CCT designed to address both current and future poverty through cash transfers targeted to poor households in rural Nicaragua. The government of Nicaragua implemented the program with technical assistance and financial support from the Inter-American Development Bank (IDB). On average, transfers were 18 percent of pre-program expenditures and delivered bimonthly (every two months). They were paid to a designated household representative (typically a female caregiver) in the beneficiary household and came with a strong social marketing message that the money was intended to be used for educational, food, and health investments.

The CCT had two core components. The first core component focused on nutrition and health with all households eligible for a transfer of a fixed amount per household regardless of the household's size and composition. The nutrition and health transfer was conditional on preventive healthcare visits for all children under age five and on the household representative attending bimonthly health information workshops. The supply of preventive health services was increased in parallel with the program.

The second core component was education and households with children ages 7–13 years who had not yet completed the fourth grade of primary school were eligible for this component. They received an additional fixed bimonthly cash transfer known as the school attendance transfer, which was contingent on enrollment and regular school attendance. For each eligible child, the household also received an annual cash transfer at the start of the school year, intended for school supplies, conditional on enrollment. We refer to the combined school attendance and school supplies transfers as the education transfer. Teachers were required to report enrollment and attendance using forms specifically designed by the program for the verification of the conditions and they also received a nominal per-student cash transfer.

In addition, during the period in which the late treatment group received the program, all adolescents, as well as the designated household representative, were required to attend information sessions focused on reproductive health and contraception, and contraceptive methods were made available to beneficiaries through the healthcare providers. More detailed information on the program is provided in appendix C.

⁴ As such, the evidence also complements studies examining the short-term reproductive health and fertility effects, most of which focus on the adult beneficiaries (Avitabile, 2012; Lamadrid-Figueroa, et al, 2010; Stecklov et al, 2007; Todd, Winters and Stecklov, 2012). See also Khan et al (2016) for a systematic review.

While the CCT, like many other CCTs in Latin America, was modeled after Mexico's original CCT program, there are two differences between the programs important for our analysis. First, in Nicaragua beneficiary households were only eligible to receive transfers for a fixed period of three years, after which it was impossible to renew eligibility. Second, the education conditionalities and transfers in the CCT only applied to the first four grades of primary school, reflecting the low levels of education and high primary school dropout rates in the country.

B. Experimental Design of the CCT

The randomized evaluation was built into the design of the CCT intervention in six rural municipalities from three regions in central and northern Nicaragua, chosen based on their high poverty rates and low education and health indicators. In these six municipalities, 42 out of 59 rural *comarcas* (hereafter, localities) were selected based on a poverty marginality index.⁵ A program census of all residents in these 42 localities was collected in May 2000.

The 42 targeted localities were then randomized into one of two equally sized treatment groups, the early or late group, at a public lottery. To improve the likelihood that the selection of localities in the two experimental groups would be well balanced in terms of poverty levels, the marginality index was used to rank and then classify the 42 localities into seven strata of six localities each. From each stratum, three localities were randomly selected as early treatment and three as late treatment.

The randomization occurred in July 2000, after the program census, and the 21 early treatment localities received their first transfers in November 2000. All but 3 percent of households in these localities were eligible for three years' worth of cash transfers and received the last transfer in late 2003. Households in the late treatment localities were informed at the outset that the program would start in their localities later. The 21 late treatment localities were phased in at the beginning of 2003. They were also eligible for three years' worth of cash transfers. Households in the early treatment group did not receive any transfers after 2003, and correspondingly were not subject to any further conditionalities after that date. Therefore, while the reproductive health services were provided after 2003 in both the early and late treatment groups, attendance was a conditionality only for the late treatment group. Consequently, in practice these services were less well utilized and only partially implemented in the early treatment group where there were no longer other on-going demand-side program components operating. At the end of 2005, all program components were discontinued for both groups and the program no longer operated in these municipalities.⁶

Overall, compliance with the experimental design was high. Short-term evaluations demonstrate that the sample was balanced at baseline and that there was little contamination of the late treatment group (Maluccio and Flores 2005). Appendix Table A1 shows balance among a wide variety of baseline variables for the main sample used in this paper. At the household level, program take-up in early and late treatment localities was approximately 85 percent.

⁵ Census *comarcas* are administrative areas within municipalities (defined by the 1995 Nicaraguan national population census) that included as many as 10 small communities for a total of approximately 250 households. The marginality index was constructed from this national census and included indicators of literacy, family size, and water and sanitation conditions. See appendix C for further details.

⁶ More generally, all CCT programs in Nicaragua were discontinued after 2006. In total they benefitted ~30,000 households. As such, general equilibrium effects on labor and marriage markets are likely to be limited.

III. Data

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

Program Census Data—The 2000 program census provides baseline data on early and late treatment localities including household demographics, grades attained (defined throughout the paper as the highest grade attained, i.e., number of school grades successfully passed) for all household members, housing characteristics, and assets.

Short-term Evaluation Surveys—The first round of the short-term evaluation surveys was conducted in September 2000, with subsequent rounds in 2002 and 2004. The household-level survey instrument was based on the 1998 Nicaraguan Living Standards Measurement Survey, with modules covering education, health, and household expenditures, among others. The sample was drawn from the 2000 program census and includes a random sample of early and late treatment households, 42 households in each of the 42 early and late treatment localities (Maluccio and Flores 2005). Attrition was approximately 10 percent per round. Starting in 2002, a non-experimental comparison group (that never received the program) was added from neighboring municipalities, chosen from a set of localities with marginality scores similar to the original 42 localities. The 2002 household survey targeted for interview 40 households in each of the 21 selected localities, and, since there is no program census data available for this comparison group, provides the baseline for this sample.

2010 Evaluation Survey—Between November 2009 and November 2011, 9 to 11 years after the start of the program, we conducted a long-term follow-up evaluation survey. For convenience, and since the bulk of data were collected in 2010, we refer to this as the 2010 survey. The sample frame consisted of all households in the original short-term randomized evaluation survey sample including the non-experimental comparison group, as well as an additional sample of households who, according to the 2000 program census, had children of critical ages relevant to the long-term evaluation. Specifically, we oversampled households with children born between January and June 1989 in the early and late treatment groups (see section IV for further details and rationale).⁷ All experimental estimates use sampling weights to account for this sampling procedure, and also adjust for attrition as described in appendix F. While the more comprehensive information from the earlier short-term evaluation surveys is not available for the oversampled households, there is information on all of them from the 2000 program census. The initial household-level sample frame has a total of 1,330 households from the early treatment group, 1,379 households from the late treatment group, and another 757 households from the non-experimental comparison group in 2010.⁸ For the main cohort of interest in our study, girls 9–12 years old, the sample frame included 448 observations in the early treatment group, 440 in the late treatment group, and 202 in the non-experimental comparison group. In the 2010 survey, we attempted to interview all 3466 households, as well as all split-off households with at least one original household member under 22 years old in 2010.

We expanded on the short-term household evaluation survey instrument, and also included a separate, individual-level instrument. The household-level instrument included questions on educational attainment and current schooling, and a detailed survey module measuring

⁷ We also oversampled households with children born between November 2000 and mid-May 2001, for analyses on exposure to the CCT during early childhood (Barham, Macours, and Maluccio 2013).

⁸ The 2002 household survey targeted 840 households in the comparison group, of which 823 were interviewed. Of those, 32 would have been ineligible for the program based on asset exclusion criteria. An additional 34 lived in border areas and ultimately received the program, so the target sample in 2010 was 757 households.

participation and earnings in all economic activities each household member had engaged in over the previous 12 months. It asked separately about economic activities conducted at or near the place of residence and about activities engaged in while temporarily absent from the household. For all young adults ages 15 to 22, an additional module also collected their full labor market history, with questions on participation, location, and earnings for all non-agricultural wage jobs and self-employment. All questions in the household-level instrument were answered by the best-informed person in the household available for the interview. Hence, responses were obtained from the young women themselves if they could be located at home, or from the household head or the spouse of the household head if not.

The 2010 individual-level instrument was conducted through direct in-person interviews with the young women in their homes, and designed to measure individual learning and socio-emotional outcomes, as well as marriage and reproductive health outcomes of each individual born after January 1, 1988 (see Appendix E for details). Qualitative interviews were carried out in preparation for the long-term evaluation survey (CIERUNIC, 2009). They revealed a perception by older program beneficiaries of increased early sexual activity as a possible perverse result of increased access to information and contraceptives through the reproductive health sessions targeting adolescents. There was also a perception of earlier age of menarche and sexual maturity. Based on these qualitative findings, we included a specific module that asked recall questions about fertility, use of contraceptives, and prior attendance at reproductive health workshops, as well as age of menarche and age of first sexual activity.

To measure learning, we administered three Spanish language and two math tests to all young adults between the ages of 15 and 22 years (in 2010). The Spanish language tests included a word identification test, a spelling test, and a test of reading fluency. The math tests included a test of math fluency and a test to measure problem solving at various levels of difficulty, which we refer to as math problems and is similar to the *Peabody Individual Achievement Test* (Markwardt 1989). All tests were appropriate for different grade levels. In addition, we administered two tests that could capture both learning and cognitive development: the *Test de Vocabulario en Imágenes Peabody* (TVIP, the Spanish version of the *Peabody Picture Vocabulary Test*; Dunn et al. 1986), and a forward and backward digit span test, in which the respondent is asked to repeat a series of numbers read to her. Finally, the Raven's colored matrices (the 36-item version with sets A, AB, and B) was added to measure cognition (Raven, Court, and Raven 1984). An important advantage of all of the tests is that they provide observed, as opposed to self-reported, measures of learning and cognition, thereby substantially reducing concerns about reporting biases and regardless of current schooling status. Finally, the individual survey instrument also included measures of socio-emotional outcomes. In particular, we administered the 20-item Center for Epidemiologic Studies of Depression (CESD) Scale and the Strengths and Difficulties Questionnaire (SDQ).

Nicaraguan Population Census Data—Finally, we use the two most recent Nicaraguan population censuses in the double difference models for the absolute effects. These national censuses provide repeated cross sections at the individual level and include basic demographic and education information, in 1995 before the start of the program and in 2005, the year the program ended.

IV. Methodology

Given the timing of the initial intervention and of the subsequent data collection, there is no formal pre-analysis plan guiding the analysis in this paper. Rather, the hypotheses and outcomes

investigated broadly follow those outlined in proposals prepared to finance the follow-up data collection.⁹

A. Identification of Experimental Long-term Differential Impacts

To estimate the differential long-term effects of the CCT, we take advantage of the exogenous variation in treatment assignment provided by the randomized phase-in of the program. We focus on the cohort of girls, ages 9 to 12 when the program started in the early treatment group in November 2000, as the experimental variation introduces differential exposure to key program components at critical ages for both reproductive health and education, depending on the treatment group. Transfers were provided to the early treatment group from November 2000 through October 2003, and to the late treatment group from January 2003 to the end of 2005.

Specifically, for the 9-12 cohort the CCT may have directly affected reproductive health outcomes differently in the early treatment than in the late treatment group through two mechanisms—one physical and the other behavioral. First, when the households of late treatment girls in the 9-12 cohort began receiving transfers they were starting their early teenage years (11-14) and a positive shock in nutrition (corresponding with increased expenditures on food) may have accelerated sexual maturity and menarche. In contrast, early treatment girls in the cohort may have benefitted from better nutrition at younger ages and their access to good nutrition was likely reduced after the end of transfers to their households (when they were 12-15), potentially delaying menarche. Either possibility, or the combination of both, leads to a positive differential in the age of menarche and therefore potentially in the age of first sexual activity. Second, all girls in this cohort were exposed to the reproductive health information sessions between 2003 and 2005. Those in the late treatment group, however, were exposed to a greater degree. Qualitative findings suggest this exposure may have led to an increase in early sexual activity.

Differential exposure to the education components of the program also may have affected outcomes differently in the two treatment groups. Based on the program rules and dropout rates, one might expect a stronger positive effect for the early treatment group on education. Children ages 7 to 13 at the start of the school year in January were eligible for the education transfer for that year if they had not yet completed fourth grade. Education transfers were provided and conditionalities applied to the early treatment group from November 2000 through October 2003, and to the late treatment group from January 2003 to the end of 2005. As a result, and abstracting from the grade requirement, the difference in exposure to the education transfer would be up to three school years for 11- and 12-year olds at the start of the program, up to two years for 10- and 13-year olds, and so on.¹⁰ Because some fraction of children reached the maximum grade of eligibility prior to the end of the program in their treatment group, actual exposure differences can differ.¹¹ In fact, actual differences in exposure to the education transfers (based on program administrative data) are largest, approximately two school years, for 9-year-old girls and gradually decline as age increases (Figure A1). For the 9-12 year-old girls, take-up of the education transfer during at least one school year was 94 percent in the early treatment group, and 85 percent in the late treatment group.

⁹ In particular for NSF (SES 11239945 and 1123993), 3ie (OW2.216), and IDB funding. See <https://sites.google.com/site/johnamaluccio/research/nicaragua-cct-proposal-and-analysis-plans-1>.

¹⁰ The announcement of the program in July 2000 means that although the transfers did not begin until late in the school year 2000, the program still had strong potential to influence schooling in that year.

¹¹ Actual differences also may be lower than potential because take-up was less than 100 percent.

Potential positive impacts on schooling of these differences in exposure to the educational transfers and their conditionalities, however, may be diminished because of the pre-existing school dropout patterns for girls, which differ markedly from those of boys (Barham, Macours, and Maluccio 2018). Enrollment is generally high at eligibility ages, though never above 80 percent, with most of the increasing risk of dropout for girls beginning only at age 13 (Figure A1). Most girls 9–12 years old in 2000, therefore, would not yet be at higher risk of dropping out between 2000 and 2003.¹² On the other hand, girls in the same age cohort from the late treatment group would have been 12–15 by the time their households became eligible for the program, and therefore at higher risk. While the education transfer and conditionality did not apply to all of them directly, the education of girls in the late treatment group could still have benefitted if they were living in households receiving transfers and other program benefits. In contrast to our predictions for reproductive health for this cohort of girls, then, predictions for the differential impact on education resulting from differential exposure to the educational components of the Nicaraguan RPS CCT are somewhat ambiguous.¹³

B. Empirical Specification for Long-term Differential Impacts

We estimate the following individual-level model,

$$Y_{il}^k = \alpha^k T_l + \beta^k X_{il} + \varepsilon_{il} \quad k = 1 \dots K, \quad (1)$$

where Y^k is one of the outcomes of interest for individual i in baseline locality l . T is an intent-to-treat (ITT) indicator that takes on the value of one for girls in localities randomly assigned to early treatment and zero for those in localities randomly assigned to late treatment. Principal analyses are carried out on an ITT basis and using all girls from both treatment groups in the 9–12-year-old age cohort, regardless of initial levels of completed schooling or actual program participation. Given randomized assignment, our main specifications limit the set of control variables X to age when the program started in early treatment (dummy indicators for 3-month age groups), completed grades attained prior to the program (dummy indicators for 0, 1, 2, 3, or 4+ grades), and regional fixed effects. All regressions also contain strata fixed effects, to account for the fact that randomization was stratified by marginalization level. All regressions are weighted to account for sampling and attrition providing population estimates (Section IV.D). We assess robustness of the above specification with alternative sets of controls and using alternative weights and samples, in appendix B. These alternative specifications do not substantively alter the main findings.

Standard errors are adjusted for clustering at the locality level. As there are 42 localities, we also assess whether accidental imbalance related to the relatively small number of clusters drives any of the results. We follow the recommendation by Athey and Imbens (2017) and Young

¹² We calculate ages on 1 November 2000, the start of the program, which we define to be when transfers began in the early treatment group.

¹³ Although the length of exposure for 7 and 8 year olds is also high, we do not include them in the main analysis because many of them had not reached the age of menarche before the end of the program (with average age of menarche in the sample over 13 years old). In addition, their risk of dropout is low during the program years in the early treatment group and the potential differential exposure to the education transfers is also lower. We also do not include 13 year olds (for whom we did not collect the individual surveys and whom we did not track beyond the original household) in the main analysis, because many of them had already reached the age of menarche or had completed fourth grade by program start and were therefore ineligible for the education transfer. See section V for discussion and analysis of these other age groups.

(2017) and estimate the exact p-value under the sharp null hypothesis that the treatment effect is null, by calculating all possible realizations of the test statistic and rejecting if the observed realization in the experiment itself is extreme enough. This avoids a dependence on asymptotic theorems that can produce inaccurate finite sample statistical inference sensitive to outliers. In testing the null of no treatment effects, randomization inference is not testing whether the average treatment effect is zero, but rather whether the treatment effect is zero for all participants.

C. Outcomes

CCTs can influence a wide range of behaviors. Therefore, we measured a large set of outcomes, with the main analysis focusing on outcomes in 2010 when the respondents were young women approximately 19–22 years old. To reduce concerns regarding multiple hypotheses testing we follow Kling, Liebman, and Katz (2007) and organize individual outcomes into different domains we refer to as “families of outcomes,” capturing labor market, earnings, fertility, schooling, learning, and socio-emotional outcomes. For each individual outcome we calculate within-sample z-scores, using the mean and standard deviation (SD) of the late treatment group. We then determine the average z-score for each family of outcomes and estimate the ITT effect using this index, which yields the effect size in SD. To further test the robustness of the findings for each family, we show p-values adjusted for familywise error rate, as proposed by Anderson (2008).

For the labor market, we consider two families of outcomes: labor market participation and labor market earnings. The labor market participation family captures participation and temporary migration for work. The earnings family includes labor market returns or earnings for work off the household farm. We present two versions of the earnings family to account for nature of the distribution with a larger number of zeros and long right-side tail: 1) using the rank of earnings instead of actual reported values (following Athey and Imbens, 2017); and 2) using absolute values of earnings trimmed at the top five percent of values. Labor market and earnings data were constructed based on a comprehensive module of labor market activity, covering all activities over the last 12 months, and separating between activities conducted while living in the home community versus those performed during periods of temporary migration. Additional variables are constructed from a separate labor market history module, which reflects all off-farm activities since the young women entered the labor market. Given the high seasonality and temporary nature of many economic activities of the target population, this comprehensive approach is key to accurately reflect labor market returns. Specific variables included in the earnings family are described in further detail in Section V.A.

The fertility family is composed of five variables including three indicator variables (ever married, have any children, and whether had sex by age 15), body mass index (BMI), and age of menarche. As BMI can be affected by having started child bearing, higher BMI can be indicative both of an earlier onset of puberty and of previous child bearing. We separately examine two additional variables to explore whether there was differential access to program interventions related to reproductive health, but do not include these as components of the fertility family.

The education family includes an indicator of whether the girl was enrolled in school, highest grade attained, and whether she had completed fourth grade, after which children were no longer subject to the program’s schooling-related conditionalities. To analyze learning and cognition, we classify the various tests into three categories. The learning family measures skills most

likely acquired in the classroom and comprises the five achievement tests (word identification, spelling, reading and math fluency, and math problems). The learning-and-cognition family includes tests that are likely to capture both learning and cognition (receptive vocabulary and memory test). And finally the cognition family has only one test, the Raven's colored matrices. For the socio-emotional outcomes, we conducted an exploratory factor analysis including all items of the CESD and the SDQ.¹⁴ This analysis points to four latent factors, broadly capturing optimism, positive self-evaluation, stress, and negative self-evaluation (see appendix E). The socio-emotional family is measured as the average of the z-scores for the four factors, after reversing the signs that have opposite meaning (stress and negative self-perception) so that higher values always indicate more positive, or better, socio-emotional outcomes.

D. Attrition and Internal Validity

Considerable effort and resources went into minimizing attrition in the 10-year follow-up for both the individual- and household-level instruments. Individuals not found in their original locations were tracked during an intensive tracking phase to their new locations throughout Nicaragua. Migrants to Costa Rica—the destination of 95 percent of international migrants from the sample—also were tracked. As migration in this context is often temporary, the survey team made multiple return visits to the original localities to interview temporary migrants after they returned. In addition, for permanent migrants who could not be located, information on selected variables (related to educational, marital, migration, and labor market status) was collected through proxy reports in the original household. For those select variables, information is available for all but 6 percent of the girls 9-12 years old at baseline. Of the targeted cohort of young women, 16 percent could not be tracked to their 2010 location and hence has missing information for the main outcomes coming from the household instrument which include education, marriage, earnings and labor market outcomes.¹⁵ Attrition is 22 percent for information collected through the individual-level instrument, which required direct, in-person interviews and included all tests on learning, cognitive, and socio-emotional outcomes, BMI, and reproductive health history. Overall, these attrition levels are on par with or lower than those found in related longitudinal studies with similar time horizons and target populations (see appendix F).

There are no significant differences in attrition levels between early and late treatment groups—differences are smaller than 1.5 percentage points and p-values of the differences are 0.794 for data from the household survey and 0.734 for the individual survey—and attrition does not affect the balance of baseline observables (appendix Tables A1, F1, and F2). Therefore, one approach would be to introduce no further attrition corrections. However, further comprehensive analysis on attrition in this sample demonstrates that it is correlated with baseline characteristics that are associated with migration patterns, and that these correlations differ to some extent between early and late treatment groups. And because attrition rates are non-trivial despite the tracking efforts, and marriage and labor market opportunities are the two main reasons for

¹⁴ We use exploratory factor analysis, as the correlations between items in the SDQ suggests standardized scoring of these items is unlikely to reflect intended latent traits, similar to findings in Laajaj and Macours (2017).

¹⁵ While 93 percent of original targeted households were surveyed, many young women no longer lived in their original locations and attrition was slightly higher for split-off households. Attrition rates for girls of the oversample households are not significantly different than for the rest of the sample and equally balanced (p-value of the difference 0.766 for the household survey and 0.704 for the individual survey).

migration (and consequently for attrition) of young women in this context, sample selectivity could still affect the findings.

To address this possibility, we consider several approaches to account for attrition selection, each relying on a different set of assumptions. In our preferred approach for the main analyses reported in the paper, we account for selection by weighting. Specifically, we use inverse probability weighting (IPW), allowing for differences between early and late treatment groups and incorporating information from the survey tracking process to give higher weight to individuals who were more difficult to find. The rationale underlying this strategy is that those not found are more similar on both observables and unobservables to those who were harder to find than to those more easily found (Molina Millán and Macours 2017). Separate weights are calculated for variables from the household and individual surveys. Appendix F provides further information on tracking and details the construction of these attrition weights, which also incorporate the sampling weights. Estimates without the attrition correction are shown in the appendix B (Tables B2.1-B2.4) and demonstrate that the main results are not driven by the attrition correction. We also estimate Lee (2009) bounds to test the sensitivity of results for corrections of the small differences in attrition rates, but note that their applicability to our study is unclear because the monotonicity assumption underlying them may be violated in this context in which we estimate differential treatment effects. Finally, as an additional robustness test we also use a more standard IPW correction (estimating weights using observations from both the regular and the intensive tracking phases combined). Unless otherwise noted, the various approaches to correct for attrition yield qualitatively similar findings (see discussion in Appendix B and results in Table B1.1).

E. Identification of Non-Experimental Absolute Effects

Double difference estimation using 1995 and 2005 Nicaraguan National Censuses—At some point during the period 2000-05, the three-year CCT operated in *all* rural areas of the six municipalities where the evaluation was implemented. In addition to nearly the entire population of the 42 experimentally assigned localities, 80 percent of the population in the other 17 localities of these municipalities (that were not part of the randomized experiment) also received the program for three years, starting in late 2001 (see Appendix C). Taken together, therefore, the program covered over 90 percent of the rural population in these municipalities. Given this high coverage, it is possible to use national census data to approximate absolute program effects on selected outcomes.

We use a non-experimental double difference approach to estimate absolute program effects with the two most recent Nicaraguan censuses. Together, the 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. The data include current municipality of residence (and whether rural or urban), as well as municipality of residence five years prior to the census administration date. As such, they allow us to account for selection due to domestic migration and therefore provide important complementary evidence that is less vulnerable to selection due to migration-related attrition compared to the 10-year follow-up survey. We calculate double difference impacts using two cohorts of girls (those ages 9–12 in 1990 and in 2000) by comparing education, civil status, and fertility outcomes in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities where the non-experimental comparison group was selected. The 9–12 age cohort in 2000 is the same age cohort examined in the experimental analyses. More specifically, we estimate

$$Y_{imt} = \delta_0 + \delta_1 T_{m,t-5} + \delta_2 C_t + \delta_3 T_{m,t-5} * C_t + \varepsilon_{imt} \quad (2)$$

Where Y_{imt} is the outcome for girl i in municipality m measured in census year t , $T_{m,t-5}$ is an indicator for whether the girl resided in a treatment municipality five years prior to the census year, and C_t an indicator for the 2005 census. δ_3 yields the double difference estimate five years after the program began. Standard errors are robust to heteroskedasticity. We examine alternative age cohorts and comparison groups, and assess common trends in appendix G.

Matching estimators using the non-experimental comparison group—To obtain an estimate of the absolute effects after 10 years, we take advantage of the non-experimental comparison group sample drawn in 2002 and resurveyed in 2010.¹⁶ Since this group never received the CCT, we use it as a counterfactual for the experimental treatment groups and compare 2010 outcomes of 9–12 year olds. While the non-experimental comparison localities were selected based on having similar marginality indices to the treatment localities, and several share a geographical border, the *ex ante* match on locality-level characteristics did not completely balance household and individual characteristics across the groups. To improve the balance on observable characteristics, we estimate propensity scores at the individual level and use the five nearest-neighbor matching with bias adjustment and kernel estimators to determine the absolute treatment effects. To maximize power, in particular for educational and learning outcomes, we pool early and late treatment groups together, an approach justified by the findings that there are no experimental differential effects for those outcomes.

The nearest neighbor estimator is bias corrected and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. Standard errors for the kernel estimates are calculated via bootstrapping, also accounting for clustering at the locality level. The common support is defined by trimming all observations that have a propensity score lower than the first percentile of the treatment group. Appendix H provides further details on the matching procedures, as well as several alternatives, including specifications with alternative definitions of the common support, one and two nearest neighbors matching, and local linear regression matching estimators.

Both the double difference and matching non-experimental estimates are based on stronger assumptions than the experimental estimates. We present them mainly to help distinguish between different possible mechanisms and to facilitate interpretation of the differential effects.

V. Results

A. Results for girls 9-12 years old at the start of the program

A primary pathway through which CCTs have potential to reduce intergenerational poverty is via the labor market. Prior to investigating the possible underlying mechanisms, therefore, we begin with an assessment of the experimental differential impacts of the CCT on labor market participation and earnings (Tables 1 and 2). The findings indicate that young women (ages 19-

¹⁶ Attrition for this non-experimental sample of 9-12 year old girls is 13 percent for the variables measured in the household instrument and 21 percent for variables from the individual instrument. There are 182 young women in the comparison sample with information on education, fertility and labor market outcomes.

22) exposed to the CCT program starting when they were girls between ages 9-12 were doing markedly better in the labor market than those exposed three years later. There is a clear increase in the labor market participation family (0.17 SD) with notable significant increases in work off the family farm (7 percentage points) and (temporary) migration for work in the last 12 months (9 percentage points), with the latter more than double the mean in the late treatment group. Corresponding to these results, there is a differential effect on earnings of approximately 0.11 SD for both versions of the labor market earnings family (Table 2), significant at the 10 percent level. Three of the four rank indicator components for earnings are also significantly positive— young women in the early treatment group have higher ranks for earnings per month worked and annual earnings, as well as for their (monthly) maximum salary in the previous 12 months. While the magnitude of these effects is modest, note that all results are unconditional and incorporate women who are not working and therefore have no earnings, approximately half of the sample.

To disentangle and understand the key potential mechanisms underlying these differential effects, we first analyze reproductive health and then education outcomes. Table 3 shows negative differential impacts on fertility-related outcomes with the overall fertility family decreasing 0.17 SD. The early treatment group is 11 percentage points less likely to have been sexually active by age 15 years, a one-third reduction over the mean in the late treatment. This large difference is reflected in a 6 percentage point lower probability of having had a child by 2010 when the young women were 19-22 years old (p-value 0.17). Table 3 further shows a significant differential impact on the age of menarche, with the late treatment group starting menarche earlier, consistent with better nutrition for them during critical early teenage developmental years and/or the interruption of good nutrition for the early treatment group when their program transfers ended. We also observe a strong differential impact on BMI. BMI for non-pregnant women in the early treatment group is 0.6 kg/m² lower than in the late treatment group, consistent with later ages of menarche and the delays in fertility. As expected, columns 7 and 8 confirm that early treatment girls were less likely to have attended the CCT's reproductive health workshops or to know about the PAP test, a health topic heavily emphasized in those workshops. This is consistent with the hypothesis that there was differential exposure to those sessions and the interpretation that they may have influenced fertility behavior in a perverse manner.¹⁷ Finally, there is a negative but insignificant differential effect on permanent migration, which in this context is often related to marriage.

In Table 4, we use short-term evaluation data to confirm that exposure to the CCT was indeed related to large shocks in available nutrition at the household level *during* program operation as both the quantity and the quality of household food consumption changed significantly. In 2002 when girls in the early treatment had received transfers for two years (but the late treatment group had not yet benefited), per capita food consumption was more than 35 percent higher in their households than in the late treatment group. Moreover, there was significantly improved nutritional quality with higher consumption of animal proteins and fruits and vegetables. By 2004 the program had been operation for a little more than a year in the late

¹⁷ It is possible that respondents could not remember whether certain workshops were organized by the CCT or other actors, so that overall participation in the CCT-organized workshops may be overstated. Recall error also may affect the information collected on age at menarche. For both of these variables, however, it is unclear whether such measurement error would affect the early and late treatment groups differently. Importantly, results for BMI are significant and in line with the other reproductive health outcomes including age at menarche. As BMI is an observed outcome (with weight and height measured using standardized instruments and methods) it is unlikely to suffer from differential measurement error. The consistent findings for BMI and age at menarche, therefore, suggest systematic recall measurement error on age is not driving the fertility findings.

treatment group but had been phased out in the early treatment group. As expected, the pattern reversed with 13 percent higher food consumption, and better nutrition, in the late treatment group (as shown by the negative ITT effect).¹⁸ (Note at this stage in 2004, as in 2010, the experimental estimates no longer provide absolute impacts.)

Of course the CCT also may have influenced fertility-related behaviors by keeping girls in school longer. To explore the time path of program effects on education, we start with the short-term evaluation surveys and estimate the effects during program operation. The results confirm that the program, like other similar CCT programs, had sizable short-term *absolute* impacts on schooling. In 2002 (Table 5, panel A)—when the early treatment group had received transfers for two years but the late treatment group had not yet benefited—girls in the 9–12 year old age cohort had 0.27 additional grades attained, were 5 percentage points more likely to have completed fourth grade, 10 percentage points more likely to be enrolled in school, 18 percentage points more likely to attend school regularly (more than 85 percent of the time), and 10 percentage points more likely to be literate (i.e., able to read and write according to parental report).¹⁹ By 2004 (Table 5, panel B), the transfers to the early treatment had ended and the late treatment group had started receiving transfers. Reflecting that the program now operated in the late treatment area, the sign of the effects on enrollment and attendance in 2004 reverses (with similar magnitudes), and differences between early and late treatment girls are now significantly negative for those variables, suggesting the late treatment girls had begun to catch up with the early treatment girls on schooling. Notably, this is true even though most late treatment girls in the cohort were not themselves eligible for education transfers by the time the program reached their localities. This suggests impacts for them stem from the other components of the program that their household would still have received and possibly also from shifts in social norms regarding investment in education triggered by the program. The patterns suggest a potential for substantial catch-up in grades attained for the late treatment girls, indicating that the long-term differential on education may not necessarily be a good estimate of the long-term absolute program effects. Even with such effects, however, the difference in highest grade attained between early and late treatment in 2004 is positive and significant, and was double the effect in 2002, and the probability of having completed 4th grade shows a large 18 percentage points significant difference.

As the late treatment group continued receiving the program after the 2004 survey, we use the Nicaraguan national census data to shed further light on subsequent schooling progression in these areas, estimating non-experimental, double difference absolute CCT program effects, approximately one year later. Compared to girls in non-program municipalities, girls in program

¹⁸ The differences between treatment groups are smaller in 2004 than in 2002, possibly because transfer sizes were modestly smaller and also because the early treatment group may have continued to invest in better nutrition even after the end of the program as shown by Macours, Schady and Vakis (2012) for a related CCT program in Nicaragua.

¹⁹ These results build on the evidence of the effect of the CCT program on education (Maluccio and Flores 2005; Maluccio, Murphy, and Regalia 2010), but focus on the specific age cohorts relevant for analyses in this paper. We note that intra-household spillovers in the late treatment group could possibly lead to an overestimate of the differential treatment effect, for example if the late treatment households reallocated resources away from the 9-12 year olds towards younger children who would become eligible for the education transfers after the program started in their localities. Unfortunately there are too few children (less than 20%) without younger siblings to consider heterogeneity along these dimensions. That said, as the timing of the start of the program was not announced in advance, and as the data suggests an overall increase in enrollment patterns in the late treatment even prior to their receiving program benefits (a possible anticipation effect), this seems unlikely to have played a major role.

municipalities had an average of 0.5 more grades attained by 2005, translating into an 11 percentage point increase in having completed grade 4, and a 7 percentage point increase in self-reported literacy (Table 6). Moreover, at that point half of the girls in this cohort were still enrolled in school (and the estimated absolute impact on enrollment was nearly 3 percentage points), pointing to the potential for even further absolute gains in the years to come after 2005. Importantly for this cohort of girls, who were 14 to 17 years old at the time of the national census in 2005, we also find absolute differences in early marriage and teenage pregnancy, with a reduction of approximately 2 percentage points in the probabilities of ever having been married and having given birth. These results hence suggest a relatively large absolute impact on early fertility decisions, as less than 10 percent had given birth by these ages.²⁰

Turning to the estimate of the long-term experimental differential impacts on educational outcomes (Table 7), the first important finding is that the girls continued to accumulate additional years of schooling after 2004 and 2005, with a the mean of the highest grade attained at 6.8 grades for the late treatment group in 2010 (compared to 4.4 in 2004). By 2010, the experimentally estimated differential effect on grades attained is relatively small, 0.18, and no longer significant, likely as a result of continued catchup by the late treatment group first evidenced by the higher enrollment rates in 2004. Indeed, in 2010 even when they were 19-22 years old, 30 percent of the girls are still in school, with a third of those enrolled in tertiary education but most of the rest still in secondary school, consistent with accumulated schooling delays into young adulthood. Overall, the education family indicator suggests a modest positive differential, 0.10 SD. Among the individual components, only the probability of having completed 4th grade, notably the CCT program requirement, shows a significant differential.

Table 8 shows there are no significant differential impacts on learning, with the point estimates close to zero. These results could indicate that the CCT did not improve learning for girls in *either* early or late treatment groups, but they also could reflect a pattern in which impacts on learning were, on average, similar for the two groups. The latter would be consistent with the apparent catchup in schooling outcomes. The double difference findings for literacy in Table 6, for example, hint at improved learning at that earlier stage. There are also no significant differential impacts for the socio-emotional outcomes. Regardless of what underlies them, these findings indicate that the differential effects on labor market outcomes and earnings are unlikely to be the result of differences in learning or socio-emotional skills between the two groups. Tables A2 and A3 show these conclusions hold for the variables composing the learning and socio-emotional families.

The overall pattern of impacts, with modest differential effects on education, suggests a potentially important role for the negative differential effects on fertility in explaining the positive differential effects on labor market outcomes. This explanation is also consistent with the reality that early motherhood can be difficult to combine with labor market activities in rural Nicaragua. With job opportunities for young women almost non-existent in most rural villages, off-farm jobs often require commuting to local urban centers, or (temporary or permanent) migration to larger cities or Costa Rica. Appendix B (tables B1.1 and B1.2) shows that the labor market, earnings, and fertility results are generally robust to various alternative definitions of weights or sample, to different assumptions related to attrition, to adjustments for multiple hypothesis testing (using Anderson's familywise error rate) and to randomization inference.

²⁰ Appendix G shows that the double difference results are robust to alternative definitions of the comparison group, and also provides placebo evidence in support of the common trends assumption.

Splitting the 9-12 age cohort provides further evidence on the fertility channel, even if it comes at the cost of statistical power. Table 9 shows that the impacts on the fertility family and its components are driven by girls ages 9-10 years at the start of the program in November 2000 (panel A). The results are consistent with nutrition shocks (affecting girls negatively for early treatment and positively for late treatment) affecting the age of menarche primarily for the girls who had, for the most part, not yet reached menarche at the time of the shock.²¹ The differential impacts are large as the probability of early sexual activity and having started child bearing when they were surveyed at ages 19-20 years is reduced by 10 percentage points, and the probability of having married by 11 percentage points. Impacts are smaller for those aged 11-12 at the start of the program (panel B), but with the exception of marriage not significantly different from the 9-10 year olds (panel C). Results splitting the age cohort for the other families of outcomes are presented in Table A4 and also show no significant differences for labor market or earnings results. In contrast to the fertility results, point estimates for labor market participation differentials are not larger for the 9-10 year olds, likely because 40 percent of the 9-10 year olds in the early treatment group are still in school, compared to 24 percent for the 11-12 year olds, and there is a positive but insignificant point estimate on enrollment of 7 percentage points for the younger group.

We provide additional evidence of the potential fertility channel on labor market outcomes and earnings by examining the differential results for all families of outcomes for young women with, and separately without, children. Clearly having children is endogenous (indeed we argue that fertility-related decisions are the crucial intermediate outcomes for understanding the long-term differentials); as such, these estimates are only meant to be descriptive. Nevertheless, it is strongly suggestive that the significant differentials on labor market outcomes and earnings are nearly all concentrated in women without children. Table 10 shows that the labor market and earnings differentials are larger and significant for women without children, but they have no differentials in education or learning outcomes. Mothers with children, on the other hand, do show a significant differential in education, albeit from a much lower mean.

B. Non-experimental matching results

The lack of clear long-term differential impacts for education and, in particular, learning could mean that the CCT did not lead to any absolute effects on those outcomes for either treatment group, or that changes experienced by the early treatment group are offset by equivalently sized impacts in the late treatment group. The positive double difference absolute estimates from the national census after five years (Table 6) point to the latter interpretation. Non-experimental matching estimates for the 10-year impacts, although based on a stronger set of assumptions and therefore more tentative, further confirm this interpretation. Table 11 presents the estimated absolute effects of the program on all the families of outcomes, as well as on highest grade attained. The estimates compare 2010 outcomes between the early and late treatment groups combined with the non-experimental comparison group (see appendix H for further details and results for alternative matching estimators and common supports). The final four columns of panel A show large absolute effects on highest grade attained (more than one full year) and on learning (about 0.25 SD), alongside an increase of 0.15 SD for the socio-emotional family.

²¹ Attrition for the 9-10 age cohort is 14 percent for the household level variables and 19 percent for the individual level variables, and balanced. Estimated coefficients on the ITT in a model of attrition are -0.013 at the household level (P=0.779) and 0.009 at the individual level (P=0.658).

Taken together, differential and absolute program effects for education, learning, and socio-emotional skills suggest that exposure to the CCT increased these outcomes for both the early and late treatment groups equally. This may be because it kept children in school during the program, but also possibly because households (or the children themselves) continued to invest in education after the program ended.

The matching results in Table 11 further suggest that there are no significant absolute effects on fertility. A possible interpretation is that the delays in fertility due to increased schooling could have been offset by the earlier age of menarche in the late treatment group. We note, however, that the five-year double difference absolute results did show significant declines in early pregnancy, so that it is unclear whether the insignificant absolute effect on the fertility family in 2010 is driven by developments between 2005 and 2010, lack of power, or sensitivity to the identifying assumptions underlying the non-experimental estimates. Taken at face value, however, the matching estimates suggest that overall the program did not have a significant impact on absolute fertility outcomes and, consistent with this, also did not increase women's labor force participation. Finally, estimates for the income rank of the earnings family suggest an approximate 0.18 SD increase. This may point to a possible positive return to the higher education and learning among those that work, though the result is neither significant nor robust to alternative specifications (see Table H3).

C. Results for other eligible ages

To this point we have focused the analysis on girls ages 9-12 at the start of the program, which highlights the potential importance of the nutrition-fertility channel using the random variation in exposure to transfers during arguably critical teenage years. As the design of the education component of the CCT targeted all girls ages 7-13 (who had not yet completed fourth grade), a related but different question of interest is the differential impact for those younger or older than 9-12 and for the entire eligible age cohort.

Non-experimental double difference estimates with the national censuses point to large absolute increases in education for the 7-8 cohort by 2005, and somewhat more modest effects for the 13 year old cohort, although both are statistically indistinguishable from the older 9-12 cohort (Table G1). This raises the possibility that there were also differential impacts for the 7-8 and 13 year olds. Tables A5 and A6 present the experimental differential treatment estimates incorporating these younger and older age cohorts.

Based on their age alone, the 7-8 year old cohort was potentially eligible for three years of the education transfers in both the early and late treatment groups—therefore, we do not *a priori* anticipate a large differential program effect on education for this cohort (though it is likely that actual exposure was lower in the late treatment group as some girls would have completed fourth grade before 2005). While most estimates are positive, differential effects on all of the families of outcomes are relatively small and statistically insignificant (Table A5). Considering these education results alongside the double difference absolute effects in 2005, this suggests the late treatment children were able to catch up with the early treatment, possibly as a result of program exposure through 2005.

The point estimate for the fertility family for the 7-8 year olds is nearly zero, consistent with the fact that this group was likely too young at the time of exposure for the nutritional and reproductive healthcare components of the program to have affected them in the same way as the older 9-12 cohort. The estimates for the other families of outcomes generally point in the same direction as for the 9-12 year old cohort, but are smaller, consistent with the 7-8 year olds in the

late treatment group still being eligible for the education component when the program reached their communities. Even so, with the exception of earnings, panel B confirms that the main findings for the 9-12 cohort are robust to widening the age group to the 7-12 cohort.²²

It is not possible to widen the full analysis of all outcomes to include the 13 year olds as this age group was not tracked beyond their original households nor administered the individual-level instrument.²³ For those 13 year olds living in the original household or in a split-off household interviewed while tracking a younger child (39 percent), we have all of the individual information asked in the household-level instrument. In addition, for 13 year olds not living in their original households, we have proxy information on key variables obtained for younger children when we were unable to track them. Combining these sources there is information for nearly 95 percent of the girls in the 7-13, 9-13, and 13 year old groups; as for the other samples, rates for missing information are similar across early and late treatment groups.²⁴

Recognizing that proxy information is likely to have more measurement error, we use it to compare experimental differential effects for the 9-12 versus 13-year olds. We first re-estimate differential effects for the 9-12 year olds using only sampling weights and including proxy information (Table A6, Panel A)—results are similar to the main findings for this age cohort, though less precise.²⁵ When we combine 9-12 and 13 year olds (Table A6, Panel B), there are significant and substantial program effects for education and marriage outcomes (Table A6, Panel B) that are larger in magnitude than for the 9-12 cohort alone. Although the differential program effects for these outcomes are larger in magnitude for the 13-year olds compared to the 9-12 year olds, in no case is the difference between the two cohorts statistically significant (Panel C). Finally, analysis of the entire 7-13 year old cohort incorporating proxies (Panel D) confirms that including the youngest cohort reduces the magnitudes of the differential impacts on marital status (column 5) and labor markets (column 6), in line with the findings discussed above (Table A5).

Overall the analysis of the younger and older ages supports our focus on the 9-12 age cohort in this paper, allowing us to highlight on the potential role of fertility-related changes for long-term labor market participation and earnings results, without too much loss of statistical power.

²² The comparison of the labor market participation and earnings outcomes of the 7-8 year olds with those of the 9-12 years old is complicated by the fact that transition from schooling and into the labor market would have been less complete for the younger cohort in 2010.

²³ The 13 year old cohort was eligible for the educational component of the program for only one year in the early treatment group. They would have been 16 by the time the program first reached the late treatment group and thus not only ineligible but also much less likely to still be in school. For these reasons, and the substantial budgetary implications of tracking such a mobile population, we cut off tracking at age 12. With respect to the nutrition channel, many of the 13 year-old cohort had already reached the age-of-menarche at the start of the program (as the average age of menarche is 13).

²⁴ For example, the P-value for the difference in attrition between early and late treatment group is 0.343 for the 7-12 age group.

²⁵ The analysis incorporating the proxy information also provides an additional test of possible attrition bias of the main results for the 9-12 year age group, since the overall attrition rates are substantially lower when including the proxy information. Results do not appear to be driven by attrition bias. Splitting the 9-12 age group as before into 9-10 and 11-12, the information with proxy also generally reflects the findings of Tables 9 and A4, with in particular the marriage results being concentrated in the 9-10 year olds (Table A7).

VI. Conclusions

CCTs are often thought to potentially affect labor market outcomes later in life by increasing educational attainment and learning. While such mechanisms are likely important, other components of CCT programs could also contribute to labor market outcomes for young adults. Evidence on such alternative pathways, however, is scarce. We take advantage of the random differential exposure to a CCT program at different ages in childhood, and a 10-year tracking survey to provide evidence on the potential role of program impacts on reproductive outcomes to understand subsequent labor market returns. Random variation in the age of program exposure during early teenage years leads to differential outcomes in the age of menarche and subsequent fertility, together with modest differential effects on education outcomes.

The results suggest that early fertility outcomes may prevent some women from reaping the returns to their education in the labor market, as the treatment group with later onset of sexual activity has higher labor market participation and earnings. While the differential impacts are modest, possibly because the young women are only just transitioning into the labor market, they suggest that nutrition and health components of CCT programs may be important to account for when analyzing long-term impacts on labor market outcomes. In addition, as the results further suggest that transfers and related nutritional shocks in early teenage years can affect the age of menarche, the age at which girls enroll or exit from transfer programs can have important consequences. Overall, the results suggest that, at least for girls, the long-term labor market impacts likely in part reflect changes in reproductive health outcomes.

VII. References

- Abadie, A., and G. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76(6): 1537–1557.
- Abadie, A., and G. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects." *Journal of Business & Economics Statistics* 29(1): 1–11.
- Anderson, M., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects, 2008. *Journal of the American Statistical Association*, 103(484): 1481-1495.
- 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.
- Avitabile, C, 2012. "Spillover Effects in Healthcare Programs: Evidence on Social Norms and Information Sharing" *IDB Working paper* 380.
- Baird, S., C. McIntosh, and B. Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126(4): 1709–1753.
- Baird, S., C. McIntosh, and B. Özler. 2018. "When the Money Runs Out: Do Cash Transfers have Sustained Effects?", mimeo, George Washington University.
- 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 and Proceedings* 103(3): 467–471.
- Barham, T., K. Macours, and J.A. Maluccio, 2018. "Are Conditional Cash Transfers Fulfilling their Promise? Schooling, Learning and Earnings After 10 Years." University of Colorado Boulder, unpublished.
- Behrman, J. R., S.W. Parker and P.E. Todd. 2009. "Medium-Term Impacts of the *Oportunidades* Conditional Cash Transfer Program on Rural Youth in Mexico." In: S. Klasen and F. Nowak-Lehmann, eds. *Poverty, Inequality, and Policy in Latin America*, 219–270. Cambridge, United States: MIT Press.
- Behrman, J. R., S.W. Parker and P.E. Todd. 2011. "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Followup of PROGRESA/*Oportunidades*." *Journal of Human Resources* 46(1): 93–122.
- Blum WF, Englaro P, Hanitsch S, A. Juul, N.T. Hertel, J. Muller, N.E. Skakkebaek, M.L. Heiman, M. Birkett, A.M. Attanasio, W. Kiess, and W. Rascher. 1997. "Plasma leptin levels in healthy children and adolescents: dependence on body mass index, body fat mass, gender, pubertal stage, and testosterone." *J Clin Endocrinol Metab*;82(9):2904–2910
- Buchman, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen, 2017. "Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial", mimeo, Duke University.
- Centro de Investigación y Estudios Rurales y Urbanos de Nicaragua (CIERUNIC). 2009. "Qualitative Findings for the Evaluation of the Long-term Impact of the *Red de Protección Social* in Nicaragua." Unpublished.
- Cooper, C., D. Kuh, P. Egger, M. Wadsworth and D. Barker, 1996. "Childhood growth and age at menarche." *Br. J. Obstet. Gynaecol.* 103(8): 814–817.

- Duflo, E., P. Dupas, and M. Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review* 105(9): 2757–2797.
- Duflo, E. Dupas, P. and Kremer, M. (2017). "The Impact of Free Secondary School Education: Experimental Evidence from Ghana." Mimeo, Stanford University.
- Dunn, Lloyd M., D.E. Lugo, E.R. Padilla, and Leota M. Dunn. 1986. *Test de Vocabulario en Imágenes Peabody*. Circle Pines, Minnesota, United States: American Guidance Service, Inc.
- 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): 663–694.
- Garn, S.M., 1987. "The secular trend in size and maturational timing and its implications for nutritional assessments: a critical review." *J. Nutr.* 117, 817–823.
- Gluckman, PD and M.A. Hanson 2006, "Evolution, Development and Timing of Puberty" *Trends in endocrinology and metabolism* 17(1):7-12.
- INSERM (Institut National de la Santé et de la Recherche Médicale), 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. 2017. "Measuring Skills in Developing Countries." *World Bank Policy Research Paper* No. 8000. Washington, DC, United States: World Bank
- 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.
- Lee, D. S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 76: 1071-1102.
- Macours, K., N. Schady and R. Vakis. 2012. "Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment." *American Economic Journal: Applied Economics* (4)2: 247–273.
- Maluccio, J. A., A. Murphy and F. Regalia. 2010. "Does Supply Matter? Initial Schooling Conditions and the Effectiveness of Conditional Cash Transfers for Grade Progression in Nicaragua." *The Journal of Development Effectiveness* 2(1): 87–116.
- Maluccio, J. A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social*." *Research Report 141*. Washington, DC, United States: International Food Policy Research Institute.
- Markwardt, F. C. 1989. *Peabody Individual Achievement Test-Revised Manual*. Circle Pines, Minnesota, United States: American Guidance Service.
- Molina Millán, T. and K. Macours. 2017. "Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias." Unpublished.
- Molina Millán, T., T. Barham, K. Macours, J.A. Maluccio, and M. Stampini. 2018a. "Long-term Impacts of Conditional Cash Transfers: Review of the Evidence." *World Bank Research Observer*, forthcoming.

- Molina Millán, T., Macours, K., Maluccio, J., and Tejerina, L., 2018. “Experimental long-term impacts of early childhood and school age exposure to a conditional cash transfer.” Mimeo, Nova University.
- 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. and Vogl, T., 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.
- Robles, M., M.G. Rubio, and M. Stampini. 2015. “Have Cash Transfers Succeeded in Reaching the Poor in Latin America and the Caribbean?” *Development Policy Review*, forthcoming <https://doi.org/10.1111/dpr.12365>.
- 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.
- Young, A., 2017. “Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results”, mimeo, LSE.

Tables

TABLE 1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION AND MIGRATION FOR WORK

	Labor Market Participation Family Z-Score	Labor Market Participation Family Components			
	(1)	Worked Off- Farm =1 (last 12 months) (2)	Migrated for Work =1 (last 12 months) (3)	Ever Had a Salaried Non- Agricultural Job =1 (4)	Ever Worked in Urban Area =1 (5)
ITT	0.169** (0.074)	0.069* (0.038)	0.087*** (0.024)	0.020 (0.037)	0.016 (0.031)
N	888	888	887	883	883
R ²	0.081	0.075	0.086	0.084	0.054
Mean late treatment		0.463	0.074	0.312	0.234

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS

	Family Z- Score	Earnings Family Components (C\$)			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Earnings (last 12 months)	Maximum Salary Ever
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Rank of Earnings</i>					
ITT	0.116* (0.061)	33.758** (15.706)	28.195* (14.011)	31.313* (15.544)	4.148 (15.038)
N	888	888	888	888	883
R ²	0.098	0.091	0.095	0.091	0.091
Mean late treatment		430.3	433.4	431	444.2
<i>Panel B: Earnings — Five Percent Trim</i>					
ITT	0.104* (0.060)	97.623 (62.401)	95.031 (328.513)	134.973** (56.948)	16.071 (43.562)
N	878	848	856	848	839
R ²	0.083	0.090	0.058	0.095	0.083
Mean late treatment		464	2628	493	309

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C\$) and the exchange rate is approximately 20. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR FERTILITY FAMILY OUTCOMES AND MECHANISMS

	Fertility Family Z-Score	Fertility Family Components				Attended CCT Workshop on Reproductive Health =1	Knows What a Pap Test is =1	Permanent Migration Out of Municipality =1	
		First Had Sex by Age 15 =1	Ever Married =1	Any Children =1	Age of Menarche				Body Mass Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ITT	-0.167*** (0.060)	-0.109** (0.041)	-0.039 (0.043)	-0.064 (0.047)	0.249** (0.119)	-0.656*** (0.236)	-0.058** (0.028)	-0.063* (0.037)	-0.040 (0.032)
N	809	809	809	809	806	765	749	792	888
R ²	0.101	0.122	0.115	0.141	0.063	0.073	0.105	0.109	0.060
Mean late treatment		0.287	0.612	0.527	13.11	23.64	0.815	0.751	

Notes: *** p<0.01, ** p<0.05, * p<0.10. Age of menarche is reversed when it is included in fertility family. Body mass index does not include women who were pregnant at the time of measurement, and therefore have lower N. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument, with the exception of marriage status.

TABLE 4: 2002, 2004 EXPERIMENTAL IMPACT ON QUANTITY AND QUALITY OF FOOD CONSUMPTION

	Log Consumption per Capita	Log Food Consumption per Capita	Share of Food Expenditure On:			
			Animal Protein	Fruit and Vegetables	Staples	Other Food
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: 2002</i>						
ITT	0.300*** (0.090)	0.345*** (0.083)	0.054*** (0.013)	0.042*** (0.007)	-0.074*** (0.018)	-0.023** (0.009)
N	475	475	475	475	475	475
R ²	0.245	0.240	0.155	0.271	0.225	0.147
Mean late treatment	7.806	7.343	0.216	0.0592	0.508	0.216
<i>Panel B: 2004</i>						
ITT	-0.099 (0.072)	-0.133** (0.065)	-0.037** (0.015)	0.004 (0.011)	0.023 (0.018)	0.015* (0.008)
N	459	459	465	465	465	465
R ²	0.247	0.206	0.163	0.190	0.141	0.106
Mean late treatment	8.391	7.998	0.222	0.082	0.502	0.189

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2002 and 2004 household instruments.

TABLE 5: 2002 AND 2004 EXPERIMENTAL IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Attended School More Than 85% of Time =1	Read and Write =1	Ever Married =1	Any Children or Pregnant =1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: 2002 — Absolute Effects</i>							
ITT	0.266*** (0.048)	0.048* (0.025)	0.099*** (0.019)	0.177*** (0.027)	0.096*** (0.029)	-0.002 (0.002)	--
N	450	450	450	450	450	450	
R ²	0.821	0.755	0.177	0.191	0.215	0.067	
Mean late treatment	3.013	0.260	0.861	0.780	0.834	0.005	
<i>Panel B: 2004 — Differential Effects</i>							
ITT	0.573*** (0.117)	0.184*** (0.036)	-0.141*** (0.051)	-0.149** (0.060)	0.032 (0.022)	0.006 (0.010)	0.028 (0.019)
N	394	394	394	394	394	394	391
R ²	0.671	0.621	0.229	0.236	0.121	0.054	0.143
Mean late treatment	4.395	0.538	0.754	0.672	0.923	0.005	0.005

*Notes:**** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and use sample weights. Compares ITT effects of early versus late treatment groups. The late treatment group started to receive the program in 2003, so 2002 represents absolute effects and 2004 differential effects. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Attended school for more than 85% of the time is zero for those who were not enrolled in school at the time. All variables measured using the 2002 and 2004 household instruments.

TABLE 6: 2005 ABSOLUTE IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY

	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)
Treatment municipality * 2005 (δ_3)	0.536*** (0.084)	0.105*** (0.015)	0.028* (0.015)	0.066*** (0.013)	-0.022** (0.011)	-0.023** (0.010)
N	17,061	17,061	17,075	17,061	17,075	14,104
R ²	0.093	0.073	0.025	0.044	0.001	0.004
Mean comparison group 2005	4.542	0.642	0.502	0.830	0.162	0.085

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use national census data to compare rural areas of program municipalities to rural areas of comparison group municipalities. Birth information unreported (missing) for 17 percent of observations in these age cohorts. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include girls ages 9-12 in November of 1990 and 2000, respectively.

TABLE 7: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY

	Education Family Z-Score	Education Family Components			Read and Write =1
	(1)	Grades Attained (2)	Completed Grade 4 =1 (3)	Enrolled =1 (4)	(5)
ITT	0.096** (0.040)	0.177 (0.141)	0.066*** (0.020)	0.022 (0.029)	0.001 (0.014)
N	888	888	888	885	888
R ²	0.311	0.433	0.244	0.090	0.116
Mean late treatment		6.758	0.825	0.296	0.956

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE 8: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LEARNING AND COGNITION FAMILIES (Z-SCORES)

	Learning Family (1)	Mixed Cognition and Learning (2)	Cognition (Raven) (3)	Socio-Emotional Family (4)
ITT	-0.005 (0.057)	-0.047 (0.061)	-0.011 (0.088)	-0.053 (0.050)
N	827	826	826	821
R ²	0.449	0.328	0.258	0.126

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument.

TABLE 9: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR FERTILITY FAMILY OUTCOMES AND MECHANISMS, AGES 9-10 AND 11-12

	Fertility Family Z-Score	Fertility Family Components					Attended CCT Workshop on Reproductive Health =1	Knows What a Pap Test is =1
		First Had Sex by Age 15 =1	Ever Married =1	Any Children =1	Age of Menarche	Body Mass Index		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: ITT Results By Age</i>								
Age 9–10	-0.219*** (0.071)	-0.104** (0.051)	-0.111** (0.052)	-0.092* (0.053)	0.317** (0.136)	-0.625* (0.331)	-0.070 (0.048)	-0.114** (0.056)
Age 11–12	-0.112 (0.082)	-0.114* (0.063)	0.036 (0.056)	-0.035 (0.069)	0.180 (0.183)	-0.652 (0.422)	-0.045 (0.050)	-0.010 (0.047)
N	809	809	809	809	806	748	749	792
Mean late treatment								
Age 9–10	0.0319	0.293	0.564	0.439	13.01	23.26	0.814	0.710
Age 11–12	0.143	0.281	0.661	0.617	13.22	24.03	0.816	0.793
<i>Panel C: P-value of Difference between 11-12 and 9-10</i>								
P-value	0.263	0.898	0.0332	0.459	0.533	0.963	0.166	0.263

Notes: Notes: *** p<0.01, ** p<0.05, * p<0.10. Age of menarche is reversed when it is included in fertility family. Body mass index does not include women who were pregnant at the time of measurement, and therefore have lower N. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-10 or 11-12 as indicated at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument, with the exception of marriage status.

TABLE 10: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR ALL FAMILIES
BY MOTHERHOOD STATUS

	Labor Market Participation Family Z- Score	Earnings Family Z-Score		Education		Learning Family Z-Score	Socio- Emotional Family Z-Score
		Rank	Absolute (5 % Trim)	Grades Attained	Family Z-Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Panel A: Woman Has No Children</i>							
ITT	0.294*** (0.103)	0.261*** (0.090)	0.243*** (0.081)	-0.047 (0.222)	-0.006 (0.068)	-0.022 (0.073)	0.097 (0.060)
N	447	447	442	447	447	412	407
Mean late treatment				8.147			
<i>Panel B: Woman Has At least One Child</i>							
ITT	0.117 (0.091)	0.017 (0.095)	0.016 (0.085)	0.347* (0.178)	0.180*** (0.060)	0.016 (0.064)	-0.204*** (0.074)
N	440	440	435	440	440	415	414
Mean late treatment				5.326			

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in column 1-3 and 5-6 measured using the 2010 household survey; variables in columns 7-8 measured using the 2010 individual survey.

TABLE 11: 2010 MATCHING ABSOLUTE IMPACTS FOR ALL FAMILIES

	Labor Participation Family Z-Score	Earnings Family Z-score		Fertility Family Z-score	Education		Learning Family Z-Score	Socio- Emotional Family Z-Score
		Rank	Absolute (5% Trim)		Grades Attained	Family Z-Score		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A: NN5 — 1 Percent Trim</i>								
ATT	0.040 (0.218)	0.184 (0.158)	0.119 (0.194)	-0.055 (0.191)	1.137 (0.708)	0.319* (0.177)	0.257* (0.140)	0.144*** (0.041)
N	969	969	960	886	969	969	905	899
<i>Panel B: Kernel Matching — 1 Percent Trim</i>								
ATT	0.013 (0.159)	0.143 (0.100)	0.085 (0.125)	-0.039 (0.133)	1.227* (0.688)	0.336** (0.165)	0.284* (0.167)	0.156*** (0.059)
N	969	969	960	886	969	969	905	899

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Absolute effects compare combined early and late treatment groups to comparison group in 2010. ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors in Panel A and kernel matching (with standard errors calculated via bootstrapping) in Panel B. Z-scores are calculated using the mean and standard deviation of the late treatment group. Variables in column 1-3 and 4-5 measured using the 2010 household survey; variables in other columns measured using the 2010 individual survey. Girls ages 9-12 at the start of the program in November 2000.

Online Appendices for “Experimental Evidence of Exposure to a Conditional Cash Transfer
During Early Teenage Years: Young Women’s Fertility and Labor Market Outcomes”

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

August 2018

APPENDIX A: TABLES AND FIGURES

TABLE A1: BASELINE BALANCE FOR 9–12 COHORT SAMPLE

	Early Treatment			Late Treatment			Diff. in Means		Mean/ SD
	Mean	SD	N	Mean	SD	N	Diff.	P-value	
<i>Individual Characteristics</i>									
Age at start of transfers in months	10.98	0.88	448	10.93	1.23	440	0.05	0.51	0.04
No grades attained (=1)	0.41	0.97	448	0.38	1.02	440	0.03	0.69	0.03
Highest grade attained	1.27	2.31	448	1.36	3.24	440	-0.09	0.69	-0.03
Worked last week (=1)	0.02	0.24	448	0.05	0.52	440	-0.03	0.30	-0.05
Mother not living in same	0.08	0.25	448	0.08	0.22	440	0.01	0.61	0.03
Father not living in same household	0.27	0.79	448	0.28	1.00	440	0.00	0.96	0.00
Child of household head	0.86	0.43	448	0.84	0.49	440	0.02	0.47	0.04
Mother no grades attained (=1)	0.44	0.68	448	0.40	0.68	440	0.04	0.45	0.06
Mother 3 plus grades attained (=1)	0.33	0.73	448	0.39	0.95	440	-0.06	0.30	-0.07
<i>Household Head Characteristics</i>									
Age	45.40	15.87	448	43.43	16.25	440	1.97	0.03	0.12
No grades attained (=1)	0.55	0.53	448	0.48	0.52	440	0.07	0.02	0.14
3 plus grades attained (=1)	0.27	0.47	448	0.31	0.81	440	-0.04	0.35	-0.05
Worked last week (=1)	0.88	0.43	448	0.90	0.38	440	-0.01	0.58	-0.04
<i>Household Characteristics</i>									
Log predicted expenditures (pc)	7.74	0.69	448	7.78	0.54	440	-0.04	0.32	-0.07
Number of household members	8.15	5.08	448	8.27	4.22	440	-0.12	0.74	-0.03
Number of children aged 0-8	2.11	2.76	448	2.18	2.56	440	-0.07	0.66	-0.03
Number children aged 9-12	1.75	0.84	448	1.77	1.19	440	-0.02	0.73	-0.02
Log of size of landholdings	8.20	6.91	448	8.06	8.14	440	0.13	0.84	0.02
Family network size (individuals)	89.84	202.8	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 ag.	0.85	0.72	448	0.84	0.84	440	0.01	0.90	0.01
Wealth index - housing	-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
Live in Tuma region (=1)	0.52	2.51	448	0.36	2.58	440	0.17	0.30	0.06
Live in Madriz region (=1)	0.18	1.84	448	0.19	2.07	440	0.00	0.94	0.00
Village population	612	2552	448	357	1146	440	255	0.02	0.22

Characteristics of Nearest School

Highest grade school offers	4.71	3.63	441	4.69	5.93	434	0.02	0.88	0.00
Student-teacher ratio	37.38	31.15	440	37.98	35.90	376	-0.60	0.57	-0.02
School under local governance	0.31	2.03	441	0.31	2.12	434	0.00	0.89	0.00

Notes: Standard errors and deviations are clustered at the locality level following the stratified randomization design. Means are weighted to account for sampling and attrition providing population estimates. The sample includes girls age 9–12 at program start with information in the 2010 household survey. Balance is similarly obtained for the sample of girls with information in the 2010 individual survey. The p-value for the difference in means includes controls for strata following the program design (difference in means do not include strata controls). The mean divided by the standard deviation uses the standard deviation of the late treatment group. The predicted per capita expenditures are from the program census data and uses the proxy means method developed by the government of Nicaragua for the purpose of household targeting based on the 1998 Nicaraguan Living Standards Measurement Survey (Maluccio, 2009). The wealth indices are constructed using principal component analysis for the assets list (appendix D). School characteristics are constructed from program administrative data collected for monitoring conditionalities. School under local governance refers to schools that participated in Nicaraguan’s school autonomy reform, which provided schools and parents a certain level of autonomy over their own management and operations.

TABLE A2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR
LEARNING AND COGNITION BY TEST (Z-SCORES)

	Learning Family Components					Mixed Cognition and Learning Family Components	
	Math Fluency	Math Problem s	Reading Fluency	Spelling	Word Identification	Receptive Vocabulary	Memory Math
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ITT	-0.035 (0.064)	0.036 (0.059)	-0.005 (0.059)	0.036 (0.075)	-0.043 (0.079)	-0.047 (0.075)	-0.042 (0.069)
N	827	823	824	824	823	825	825
R ²	0.389	0.322	0.464	0.292	0.340	0.326	0.199

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 individual survey.

TABLE A3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR
SOCIO-EMOTIONAL FAMILY BY OUTCOMES (Z-SCORES)

	Family Z-Score	Socio-emotional Family Components			
		Positive Self Evaluation	Optimism	Stress	Negative Self Evaluation
	(1)	(2)	(3)	(4)	(5)
ITT	-0.053 (0.050)	-0.012 (0.092)	0.022 (0.064)	0.069 (0.083)	0.153* (0.084)
N	821	821	821	821	821
R ²	0.126	0.118	0.144	0.084	0.140

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. 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 individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 individual survey.

TABLE A4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FAMILIES, AGES 9-10 AND 11-12

	Enrolled in School (=1)	Labor Market Participation Family Z-Score	Earnings Family Z-Score		Education		Learning Family Z-Score	Socio- Emotional Family Z-Score
	(1)	(2)	Rank	Absolute (5 % Trim)	Grades Attained	Family Z-Score	(7)	(8)
<i>Panel A: ITT Results By Age</i>								
Age 9–10	0.070 (0.045)	0.110 (0.100)	0.082 (0.081)	0.109 (0.084)	0.167 (0.226)	0.084 (0.074)	0.069 (0.075)	-0.023 (0.060)
Age 11–12	-0.049 (0.044)	0.288** (0.124)	0.137 (0.101)	0.095 (0.099)	0.290 (0.201)	0.102 (0.064)	-0.003 (0.064)	-0.134** (0.055)
N	885	888	888	878	888	888	826	820
Mean late treatment								
Age 9–10	0.327				6.875			
Age 11–12	0.261				6.627			
<i>Panel B: P-value of Difference between 11-12 and 9-10</i>								
P-value	0.117	0.272	0.715	0.928	0.710	0.877	0.390	0.184

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-10 and 11-12 as indicated at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in column 1-6 measured using the 2010 household survey; variables in other columns measured using the 2010 individual survey.

TABLE A5: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, BY VARIOUS AGE GROUPS

	Labor Market Participation Family Z-Score	Earnings Family Z- Score		Fertility Family Z-Score	Education		Learning Family Z-Score	Socio- Emotional Family Z-Score
	(1)	Rank	Absolute (5 % Trim)	(4)	Grades Attained	Family Z-Score	(7)	(8)
<i>Panel A: Differential Effects Age 7-8</i>								
ITT	0.097 (0.101)	0.076 (0.097)	0.060 (0.102)	-0.001 (0.079)	0.164 (0.238)	0.046 (0.073)	0.079 (0.096)	0.079 (0.062)
N	505	505	502	473	505	505	478	477
R ²	0.070	0.112	0.092	0.081	0.257	0.171	0.179	0.071
Mean late treatment					6.776			
<i>Panel B: Differential Effects Age 7-12</i>								
ITT	0.136** (0.066)	0.099 (0.059)	0.083 (0.060)	-0.104** (0.049)	0.148 (0.049)	0.070** (0.033)	0.016 (0.063)	-0.011 (0.042)
N	1,393	1,393	1,380	1,281	1,393	1,393	1,304	1,297
R ²	0.060	0.092	0.077	0.081	0.369	0.242	0.323	0.079
Mean late treatment					6.765			

Notes: *** p<0.01, ** p<0.05, * p<0.10. Mean for late treatment for grades attained is 6.78 years. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 7-8 (Panel A) or 7-12 (Panel B) at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables in column 1-3 and 5-6 measured using the 2010 household survey; variables in other columns measured using the 2010 individual survey.

TABLE A6: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS WITH PROXY VARIABLES
BY VARIOUS AGE GROUPS

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

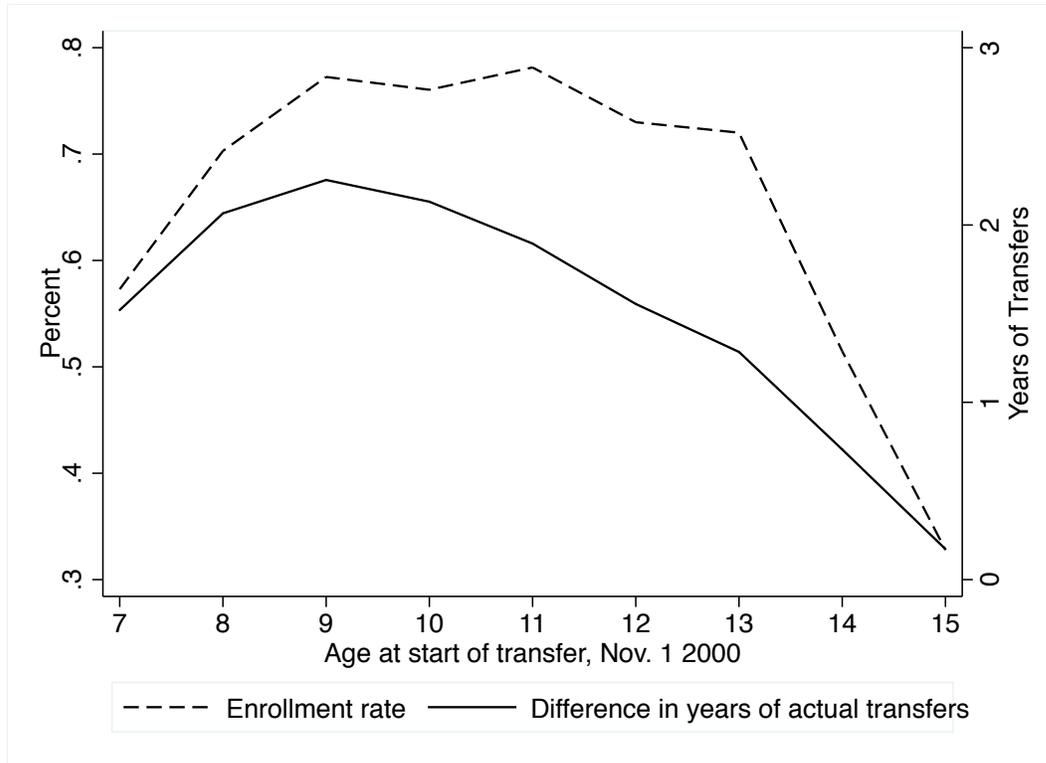
Notes: *** p<0.01, ** p<0.05, * p<0.10. Proxy reports included for all variables except enrolled because they were not collected. Regressions include girls with indicated ages measured at the start of the program in November 2000 and are weighted to account for sampling, but not attrition, providing population estimates. Standard errors are clustered at the locality level and given in parentheses. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household survey. For young women that were no longer member of their original household nor member of a split-off household, parental proxy reports obtained in the original household are included. Because 13-year olds were not intensively tracked, it is not possible to construct attrition weights similar to those done for the younger cohort

TABLE A7: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS WITH PROXY VARIABLES
BY VARIOUS AGE GROUPS

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever Married =1	Worked Off- Farm =1 (last 12 months)
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: Differential Effect Age – Using Sampling Weights and Including Proxy Reports</i>						
ITT Age 9-10	0.167 (0.226)	0.023 (0.031)	0.069 (0.042)	0.007 (0.018)	-0.131*** (0.043)	0.021 (0.043)
ITT Age 11-12	0.290 (0.201)	0.087*** (0.030)	-0.046 (0.041)	0.025 (0.023)	0.044 (0.042)	0.069 (0.046)
N	998	998	993	999	999	996
Mean Late Treatment						
Age 9-10	6.875	0.852	0.327	0.965	0.533	0.443
Age 10-11	6.627	0.794	0.261	0.945	0.647	0.486
<i>Panel B: Difference in ITT effect between Age 9-10 and 11-12</i>						
P-value	0.710	0.206	0.108	0.576	0.002	0.472

Notes: *** p<0.01, ** p<0.05, * p<0.10. Proxy reports included for all variables except enrolled because they were not collected. Regressions include girls with indicated ages measured at the start of the program in November 2000 and are weighted to account for sampling, but not attrition, providing population estimates. Standard errors are clustered at the locality level and given in parentheses. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household survey. For young women that were no longer member of their original household nor member of a split-off household, parental proxy reports obtained in the original household are included.

FIGURE A1: DIFFERENCE IN MEAN YEARS OF EDUCATION TRANSFERS RECEIVED BETWEEN EARLY AND LATE TREATMENT GROUPS AND MEAN ENROLLMENT RATE FOR GIRLS



Notes: Source: Enrollment rates (left vertical axis scale) are calculated using the 2000 program census. The difference in years of education transfers are calculated using program administrative data (right vertical-axis scale). The program administrative data contain detailed information on school enrollment and attendance for beneficiaries in both early and late treatment groups as well as information on transfers and household eligibility for the education transfer between 2000 and 2005. The difference in mean years of education transfers refers to the mean difference in the total number of school years that all children in the early and late treatment localities received, not just girls in the sample. Since eligibility depended on age at the start of the school year, while we calculate ages based on the start of the program in the early treatment group (November 1, 2000), it is possible for children older than 13 to have received transfers.

APPENDIX B: ROBUSTNESS

In Appendix Table B1.1, we examine the robustness of the 2010 differential results to a number of alternative specifications and samples. Results from the main specification reported in the paper are reproduced in panel A for comparison.

First, we estimate effects controlling only for strata as per the stratified randomized design of the program. The results in panel B are consistent with the main results with the exception of the education family, which is smaller and no longer significant.

Second, in panel C we include additional binary controls for the household head not having any education, and for the age of the household head and the population size of the village being above the mean for the population, to test sensitivity for the few factors that were not balanced at baseline. Results are similar to the main findings although the earnings rank family is only marginally significant (p-value=0.11).

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

Fourth, we explore the sensitivity of the analyses to the attrition correction weights (see Appendix F for more details). The main results in the paper are weighted to account for attrition, by multiplying the attrition corrected weights and sampling weights. To examine the results without the attrition weights, we present results using only the sampling weights in panel E. Tables B2.1-B2.4 repeat the more detailed analysis of the variables in each family (labor market participation, earnings, fertility, and education) without attrition weights. Alternatively, we also calculate a different set of attrition weights using individuals found during both the regular and the intensive tracking phases, and show inverse probability weighted regressions with these alternative attrition weights multiplied by the sampling weights in Panel F. All findings are robust to only using sampling weights (Panel E and Tables B2.1-B2.4). The labor market and fertility results are also robust to using alternative attrition weights (Panel F), but increased standard errors turn the earnings and education result insignificant in this specification.

Fifth, in panel G we re-estimate effects for the outcome variables derived from the household instrument for the subsample of girls for whom there is also information from the individual-level instrument, and use the weights designed to account for attrition and sampling in the individual survey. Results are robust to this sample restriction. (Variables to construct the fertility family are from the individual-level instrument and therefore not repeated here.)

Finally, panel H shows Lee (2009) bounds, where the bounds assume the tracked sample is either entirely negatively or entirely positively selected (the monotonicity assumption). The Lee bounds are estimated based on ITT estimations without any covariates (for example, without controls for strata). Given limited sample size, we are unable to trim the samples using covariates to obtain tighter bounds. Omitting covariates and strata controls increases point estimates for the labor market and earnings results, with both upper and lower bounds significantly different than zero. On the other hand, omitting covariates and strata controls reduces the size of the point

estimates for the fertility family, and the upper bound becomes insignificant. (Lee bounds for the age-of-menarche variable suggest it is bounded between 0.17 and 0.38, not shown). For education, both the upper and lower bounds are insignificant.

Appendix Table B1.2 further examines the robustness of the inference. Following Anderson (2008), we adjust the p-values for the possibility of multiple hypotheses testing using the familywise error rate in panel A. All differential outcomes are significant at 10 percent or below.

Panel B shows exact p-values obtained through randomization inference, by calculating all possible realizations of the test statistic and rejecting if the observed realization in the experiment itself is extreme enough, using Young (2017). This is potentially important given the relatively small number of randomized clusters (42) in the experiment. Results confirm we can reject that the Fisher exact null-hypotheses of no differential treatment effects for all families of outcomes except education.

TABLE B1.1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ALTERNATIVE SPECIFICATIONS

	Labor Market Participation Family Z-score (1)	Earnings Family Z-Score		Education Family Z-score (4)	Fertility Family Z-score (5)
		Rank (2)	Absolute (5% Trim) (3)		
<i>Panel A: Age 9–12 Main results</i>					
ITT	0.169** (0.074)	0.116* (0.061)	0.104* (0.06)	0.096** (0.040)	-0.166*** (0.06)
N	888	888	878	888	809
<i>Panel B: Age 9–12 Strata Controls Only</i>					
ITT	0.193*** (0.07)	0.171** (0.07)	0.148** (0.071)	0.055 (0.069)	-0.141* (0.074)
N	888	888	878	888	809
<i>Panel C: Age 9–12 Extended Controls</i>					
ITT	0.155* (0.079)	0.100 (0.060)	0.107* (0.057)	0.116*** (0.037)	-0.169*** (0.059)
N	888	888	878	888	809
<i>Panel D: Age 9-12 Excluding Over-Sample Children</i>					
ITT	0.214** (0.098)	0.137 (0.103)	0.139 (0.095)	0.066 (0.049)	-0.136** (0.063)
N	472	472	467	472	431
<i>Panel E: Age 9–12 Sampling Weights (No attrition correction)</i>					
ITT	0.186*** (0.067)	0.107* (0.053)	0.103* (0.053)	0.078** (0.034)	-0.117* (0.061)
N	888	888	878	888	809
<i>Panel F: Age 9-12 Attrition Correction using Observations From Intensive & Regular tracking</i>					
ITT	0.204*** (0.078)	0.113 (0.074)	0.108 (0.076)	0.086 (0.052)	-0.106* (0.058)
N	888	888	878	888	809
<i>Panel G: Restrict Household Survey Variables to Individual survey Sample and Weight</i>					
ITT	0.181** (0.07)	0.125** (0.059)	0.124** (0.057)	0.095* (0.049)	
N	826	826	818	826	

Panel H: Lee bounds

Lower Bound	0.216*** (0.074)	0.176** (0.072)	0.164** (0.071)	0.005 (0.072)	-0.114** (0.057)
Upper Bound	0.399*** (0.097)	0.208** (0.090)	0.200* (0.111)	0.050 (0.068)	-0.031 (0.062)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 household survey, with the exception of the fertility family, which is based on variables from the 2010 individual survey.

TABLE B1.2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS, ROBUSTNESS INFERENCE

	Labor Market Participation Family Z-Score (1)	Earnings Family Z- Score Rank (2)	Absolute (5% Trim) (3)	Education Family Z-Score (4)	Fertility Family Z-score (5)
<i>Panel A: Multiple Hypothesis Testing – Familywise Error Rate Adjusted P-values</i>					
P-value	0.048	0.081	0.089	0.048	0.044
<i>Panel B: Randomization Inference</i>					
Exact p-value	0.038	0.072	0.098	0.227	0.015

Notes: Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling and attrition providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. Variables measured using the 2010 household survey, with the exception of the fertility family, which is based on variables from the 2010 individual survey. Panel A adjusts the p-values for multiple hypothesis testing using familywise error rates following Anderson (2008) based on five main families included in the table. Panel B shows Fisher exact p-values obtained through randomization inference using Young (2017)'s randomization-t.

TABLE B2.1: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR LABOR MARKET PARTICIPATION AND PERMANENT MIGRATION, SAMPLING WEIGHTS

	Labor Market Participation Family Z-Score (1)	Labor Market Participation Family Components				Permanent Migration Out of Municipality =1 (6)
		Worked Off-Farm =1 (last 12 months) (2)	Migrated for Work =1 (last 12 months) (3)	Ever Had a Salaried Non-Agricultural Job =1 (4)	Ever Worked in Urban Area =1 (5)	
ITT	0.186*** (0.067)	0.033 (0.031)	0.093*** (0.026)	0.040 (0.032)	0.044 (0.028)	-0.015 (0.024)
N	888	888	887	883	883	888
R ²	0.087	0.078	0.094	0.086	0.052	0.034
Mean late treatment		0.485	0.0866	0.298	0.213	0.177

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses.

Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE B2.2: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EARNINGS FAMILY AND COMPONENTS, SAMPLING WEIGHTS

	Family Z- Score	Earnings Family Components (C\$)			
		Earnings Per Month Worked (last 12 months)	Annual Earning (last 12 months)	Maximum Earnings (last 12 months)	Maximum Salary Ever
	(1)	(2)	(3)	(4)	(5)
<i>Panel A: Rank of Earnings</i>					
ITT	0.107* (0.053)	27.841* (14.874)	23.562* (12.443)	12.873 (12.630)	25.052* (14.559)
N	888	888	888	888	883
R ²	0.100	0.098	0.095	0.089	0.098
Mean late treatment		419.7	421.3	427.8	420.8
<i>Panel B: Earnings — Five Percent Trim</i>					
ITT	0.103* (0.053)	82.735 (61.254)	61.230 (278.319)	119.109** (55.517)	26.449 (39.218)
N	878	848	856	848	839
R ²	0.087	0.105	0.046	0.110	0.077
Mean late treatment		469.1	2597	499.1	296.1

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling providing population estimates. Differential ITT results compare early to late treatment groups in 2010. Earnings include wage work off the family farm. Earnings in panel A are trimmed at the top five percent of values. Earnings are in Nicaragua Cordobas (C\$) and the exchange rate is approximately 20. The family z-score is calculated by averaging the z-score for the individual components where the average is determined even if one component is missing. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

TABLE B2.3: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR FERTILITY FAMILY OUTCOMES AND MECHANISMS, SAMPLING WEIGHTS

	Fertility Family Z-Score	Fertility Family Components				Body Mass Index	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			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ITT	-0.117* (0.061)	-0.043 (0.036)	-0.054 (0.042)	-0.044 (0.044)	0.248* (0.128)	-0.424** (0.208)	-0.067*** (0.024)	-0.063 (0.040)
N	809	809	809	809	806	766	749	792
R ²	0.062	0.091	0.079	0.108	0.053	0.060	0.101	0.078
Mean late treatment		0.229	0.573	0.488	13.13	23.40	0.816	0.719

Notes: *** p<0.01, ** p<0.05, * p<0.10. Age of menarche is reversed for fertility family. Body mass index does not include women who were pregnant at the time of measurement. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score is calculated by averaging the z-score for the individual components. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 individual instrument, with exception of the marriage status, which comes from the household instrument.

TABLE B2.4: 2010 DIFFERENTIAL EXPERIMENTAL IMPACTS FOR EDUCATION FAMILY AND LITERACY, SAMPLING WEIGHTS

	Education Family Z-Score	Education Family Components			Read and Write =1
	(1)	Grades Attained (2)	Completed Grade 4 =1 (3)	Enrolled =1 (4)	(5)
ITT	0.078** (0.034)	0.194 (0.134)	0.050*** (0.016)	0.016 (0.025)	0.005 (0.015)
N	888	888	888	885	888
R ²	0.333	0.437	0.225	0.087	0.136
Mean late treatment		6.682	0.837	0.294	0.947

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and given in parentheses. Regressions include girls ages 9-12 at the start of the program in November 2000 and are weighted to account for sampling providing population estimates. Differential ITT results compare early to late treatment groups in 2010. The family z-score averages the z-scores for the individual components and is calculated if any component is available. Z-scores are calculated using the mean and standard deviation of the late treatment group. Controls include three-month age group indicators and indicator variables for grades attained at baseline, stratification, and region. All variables measured using the 2010 household instrument.

APPENDIX C: TARGETING AND DESIGN OF THE NICARAGUAN *RPS* CCT PROGRAM²⁶

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

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

In the next stage of (pre-program) targeting, a marginality index was constructed for all 59 of the rural census *comarcas* (hereafter localities)²⁷ within the six selected municipalities. The index was the weighted average of a set of locality-level indicators (including average family size, lack of access to potable water and latrines and illiteracy rates); localities with higher marginality index scores were considered more impoverished.²⁸ The 42 localities with the highest scores were selected for inclusion in the randomized program evaluation (divided into the early and late treatment groups). Finally, while the initial program design called for all households in these 42 targeted localities to be eligible for the CCT, prior to the start of the program the government excluded approximately three percent of them determined to have substantial resources, in particular those who owned a vehicle or had large landholdings. These households are excluded from the analyses in the paper.

²⁶ This appendix draws on IFPRI (2005), Maluccio and Flores (2005) and Maluccio (2009).

²⁷ Census *comarcas* are administrative areas within municipalities based on the 1995 Nicaraguan national census that included as many as 10 small communities for a total of approximately 250 households.

²⁸ More specifically, the marginality index for each locality included average family size (10 percent), percent without piped water in the home or yard (50 percent), percent without a latrine (10 percent) and percent of persons over age five who were illiterate (30 percent)—all calculated from the 1995 Nicaraguan *National Population and Housing Census*.

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

During its operation, the CCT also expanded to the remaining $59 - 42 = 17$ rural localities not initially targeted. In those 17 localities, which were less poor according to the marginality index, the CCT began in late 2001 and was offered to 80 percent of the population based on a household-level proxy means targeting model. Consequently during the period 2000-05, the three-year program had been implemented (to modestly different degrees and at different times) in all (59) rural localities of the six municipalities, and more than 90 percent of the population had been eligible.

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

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

To provide incentives to the teachers and increase resources available to the schools, there was also a small cash transfer, known as the teacher transfer. In rural Nicaragua, school parents associations often request small monthly contributions from parents to support the teacher and the school; the teacher transfer was, in part, intended to substitute for this type of fee. This transfer was per child as well, and delivered directly to the household, which was then required

to pass the funds along to the teacher. The instruction was that the teacher keep half with the other half earmarked for the school. Although the delivery of the funds by the household to the teacher was a program condition that was monitored, the ultimate use of the funds was not. The teacher transfer was continued in areas even after household transfers had stopped until the end of the program. Teachers and schools completed specially designed scan forms that were regularly submitted to the central office to verify conditions, assess compliance, and determine transfers.²⁹

While there was no explicit supply-side intervention for education such as a school building program (having targeted the program to areas with adequate schooling infrastructure), the centrally administered CCT had a multisectoral approach promoting inter-institutional cooperation through specially formed committees at the national, municipal, and local levels. This coordination proved useful in some areas for directing ad hoc supply-side responses to increased demand, including the placement of additional government teachers.

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

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

In 2003, as the early treatment group was phasing out and the late treatment group phasing in, a number of additional services and corresponding conditions were added. Chief among them were vaccination for school-age children, family planning services for women of childbearing age, prenatal care consultations, and an additional set of health education workshops for adolescents. All adolescents were required to attend the additional workshops (segregated by age groups) and they focused on healthy living and reproductive health, including contraception. At the same time, modern contraceptive methods were made available to beneficiaries through the healthcare providers. These additional services were designed to be implemented after 2003 in both the early and the late treatment groups, but attendance was a conditionality for transfer payments only for the late treatment group as transfers to the early group were ending in 2003. Thus all girls 9-12 in 2000 were eligible for these sessions in the late treatment group. In practice, the services were only partially implemented in the early treatment group because there

²⁹ Although not an explicit component, the CCT administration did work with localities on an ad hoc basis to alleviate bottlenecks in assignment of new teachers and in the second phase offered some limited teacher training.

were fewer synergies with other ongoing program components. Consequently, attendance of adolescents at the health education workshops was lower in the early compared to late treatment groups.

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

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

APPENDIX D: WEALTH INDEX

The baseline program census data contain a number of variables to proxy for household wealth, including characteristics of the housing structure and household assets. Following Filmer and Pritchett (2001), we aggregate these characteristics using principal components analysis. The principal components are estimated using the baseline target sample of all 9–12 year old girls regardless of whether they were interviewed in 2010. We retain the first three principal components as they each has an eigenvalue greater than one; jointly, they account for 53 percent of the variation of the nine underlying variables included. The first principal component mostly reflects characteristics of the house, the second productive assets (ownership of work animals and a fumigation sprayer), and the third specific household amenities (zinc roof and latrines).

TABLE D1. PRINCIPAL COMPONENT SCORING COEFFICIENTS

Variable	PC 1	PC 2	PC 3
<i>Household Characteristics</i>			
Work animals (=1)	0.12	0.68	-0.05
Fumigation sprayer (=1)	0.25	0.47	-0.34
Number of rooms in the house	0.39	0.30	0.06
Radio (=1)	0.31	-0.10	0.32
Cement block walls (=1)	0.43	-0.05	0.01
Zinc roof (=1)	0.24	-0.26	-0.62
Dirt floor (=1)	-0.40	0.24	0.31
Latrine (=1)	0.28	0.09	0.49
Electricity light (=1)	0.43	-0.30	0.24

Notes: PC refers to principal component

APPENDIX E: LEARNING, COGNITIVE, AND SOCIO-EMOTIONAL OUTCOMES

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

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

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

Data collection and test administration was organized in such a way that the test administrators would maintain a balance between the number of children visited in early and late treatment localities. Visits to early and late treatment localities were also balanced over time to avoid possible seasonal differences in measurement between the experimental groups. Consistent with these field protocols, results are robust to controls for the identity of the test administrator (not shown).

Exploratory Factor Analysis for Socio-Emotional Outcomes—Two standardized instruments to measure socio-emotional outcomes were applied in the individual instrument. The first was the Strength and Difficulties test (SDQ), a self-reported behavioral screening test consisting of 25 questions aimed at measuring a set of positive and negative behaviors. In addition, we implemented the Center for Epidemiologic Studies Depression Scale or CESD (Radloff 1977), a commonly used mental health scale, developed as a screening test for depression and depressive disorder and consisting of 20 questions asking for the frequency of both positive and negative self-perceptions. Both tests are available in Spanish.

We first analyze the internal consistency of the different scales for the sample of girls 9–12 year olds at baseline. The overall Cronbach alpha of the 25 items of the SDQ together (0.72) indicates that the scale as a whole is internally consistent. But the alphas are much lower when considering the five usual subdomains (emotional symptoms, conduct problems, hyperactivity, peer relationships, and pro-social behavior), and vary from 0.21 to 0.54, hence much lower than the usual threshold for statistical validity. Exploratory factor analysis on the 25 items suggests there are only two factors that can be meaningfully retained (i.e., two factors have eigenvalues above one and the scree plot leads to a similar conclusion). Moreover, when imposing the 5-factor structure, the items do not group along the original five subscales, with the first factor

having high factor loads on items from three of the five subscales.³⁰ When we consider the CESD, the Cronbach alpha for internal consistency of the 20 items is high (0.86) but the factor analysis only points to one or two factors, and does not allow further differentiation.

As the data suggest that we should not pool questions together based on regular subcategories of the SDQ, we construct new indices capturing the relevant latent traits, based on all items in the SDQ and CESD. We pool together all questions from the SDQ and the CESD scales and identify the latent socio-emotional traits in our sample, following, among others, Cunha, Heckman, and Schennach (2010) and Attanasio et al. (2015). Based on both the eigenvalue and the scree plot, and using an oblique quartimin rotation to allow the different factors to be correlated with one another, we retain four factors, two factors with high loads on items from the SDQ scale, and two factors with high loads on items from the CESD scale. Notably, questions referring to positive, respectively negative, attitudes or behavior are pooled in each of the scales. Hence considering the factor loadings of the different items points to a plausible interpretation of these factors as capturing stress, positive self-evaluation, negative self-evaluation, and optimism.³¹

³⁰ Similar findings have been obtained on other measures of socio-emotional outcomes when scales originally designed for developed country settings are used in developing country settings (Laajaj and Macours 2017).

³¹ All results are qualitatively similar with or without inclusion of the non-experimental comparison group.

TABLE E1. FACTOR LOADINGS OF SOCIO-EMOTIONAL QUESTIONS

	Factor 1 Stress	Factor 2 Positive Self- Evaluation	Factor 3 Negative Self- Evaluation	Factor 4 Optimism
CESD				
During the last 7 days, how many days ...				
... were you bothered by things that usually don't bother you?	0.4613	0.0399	-0.1496	0.0869
... did you not feel like eating? (your appetite was poor)	0.4832	0.0362	0.0662	0.0303
... did you feel that you could not shake off the blues even with help from your family and friends?	0.7012	0.0913	0.0137	0.0104
... did you feel that you were just as good as other people?	0.6809	0.0055	0.0085	0.0922
... did you have trouble keeping your mind on what you were doing?	0.5838	0.0097	0.0101	0.0932
... did you feel depressed?	0.6341	0.0004	0.0062	0.0366
... did you feel that everything you did was an effort?	0.5041	0.0105	0.0233	0.0926
... were you hopeful about the future?	0.1383	0.0142	0.022	0.5384
... did you think your life had been a failure?	0.6466	0.0316	0.044	0.0849
... did you feel fearful?	0.5176	0.0054	0.0467	0.038
... was your sleep restless?	0.4503	0.0522	0.018	0.0028
... were you happy?	0.107	0.0104	0.0042	0.4808
... did you talk less than usual?	0.3319	0.0076	0.0027	0.165
... did you feel lonely?	0.6623	0.0045	0.0454	0.0235
... people were unfriendly?	0.5992	0.0633	0.0358	0.0729
... did you enjoy life?	0.4853	0.026	0.0514	0.1792
... did you have crying spells?	0.6082	0.0406	0.0624	0.0036
... did you feel sad?	0.724	0.0029	0.0081	0.0355
... did you feel that people disliked you?	0.501	0.1107	0.1072	0.0188
... could you not get 'going'?	0.432	0.0678	0.1227	0.1087
... did you feel you were moving ahead in life?	0.0587	0.0673	0.0301	0.5831
... where you thinking about the way to move ahead in life?	0.0325	0.04	0.0154	0.5876
SDQ				
I try to be nice to other people. I care about other people's feelings	0.027	0.1987	0.1787	0.2344
I am restless, I cannot stay still for long	0.0147	0.1482	0.1949	0.1999
I get a lot of headaches, stomach-aches or sickness	0.163	0.2042	0.1018	0.0414
I usually share with others, for example food, pencils/	0.0759	0.2529	0.2303	0.1165
I get very angry and often lose my temper	0.1143	0.1187	0.3129	0.1335
I would rather be alone than with other people	0.1157	0.1151	0.2168	0.0395
I usually do as I am told	0.0118	0.4762	0.1756	0.043
I worry a lot*	0.1101	0.3957	0.0942	0.0161
I am helpful if someone is hurt, upset or feeling ill	0.031	0.4608	0.0597	0.0202
I am constantly fidgeting or squirming*	0.0219	0.2641	0.1661	0.0639
I have one good friend or more	0.055	0.2266	0.0753	0.0274
I fight a lot. I can make other people do what I want	0.0405	0.1317	0.4273	0.0739
I am often unhappy, depressed or tearful	0.1983	0.1194	0.3189	0.0582

Other people my age generally like me	0.0951	0.3974	0.0103	0.0627
I am easily distracted, I find it difficult to concentrate	0.0205	0.2409	0.12	0.1546
I am nervous in new situations. I easily lose confidence	0.1256	0.1177	0.2997	0.1418
I am kind to younger children	0.0463	0.4616	0.0356	0.0808
I am often accused of lying or cheating	0.0009	0.0441	0.5141	0.0149
Other young people pick on me or bully me	0.064	0.0598	0.317	0.1938
I often offer to help others (parents, teachers, children)	0.0156	0.4591	0.0649	0.073
I think before I do thing	0.0446	0.4945	0.0465	0.0056
I take things that are not mine from home, school or elsewhere	0.0341	0.1157	0.2126	0.0621
I get along better with adults than with people my own age	0.0105	0.3656	0.0446	0.0318
I have many fears, I am easily scared	0.1621	0.1804	0.275	0.0805
I finish the work I'm doing. My attention is good	0.0067	0.5053	0.0689	0.0763

Notes: * denotes items that are meant to capture negative traits (difficulties) in English but in the Spanish translation may have been interpreted as positive by the respondents.

APPENDIX F: TRACKING PROTOCOLS AND ATTRITION CORRECTIONS

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

We hence distinguish between three sets of outcomes based on their source—from the household-level instrument, from the individual-level instrument, or by proxy from the household-level instrument. Outcomes on individuals collected in the household instrument include educational attainment, all labor market and earnings outcomes, and marital status, self-reported by the individual or—in cases when she was resident but temporarily absent—another informed household member. Outcomes collected in the individual-level instrument include BMI, age at menarche, fertility, reproductive health behaviors, achievement and cognitive tests, and socio-emotional outcomes. And outcomes on which proxy information was collected in the original household include highest grade attained, marital status, migration, and labor market status. Attrition is highest for the outcomes that required direct in-person interactions with the respondents themselves during the individual-level instrument. For the main sample of 9-12 cohort of girls examined in this paper, 22 percent could not be tracked for the individual instrument, 16 percent for the household instrument, and for 6 percent we are also missing a proxy report.³²

Given the relatively small sample sizes, we intensively tracked all migrants rather than a random subset. Tracking rates in the intensive phase are comparable to intensive tracking rates obtained in other studies for random subsamples, resulting in a final tracking rate of 84 percent for the household instrument, comparable to other long-term follow-ups of RCTs with both regular and intensive tracking phases.³³ Attrition rates in our study are similar to or lower than related studies analyzing the long-term experimental impacts of CCTs.³⁴

³² Seven individuals in the target 9-12 age group were deceased by 2010. These individuals are not used to predict the probability of attrition as selection for them is most likely driven by other factors. There are also 18 observations for whom the reproductive health module in the individual survey is missing, leading to lower numbers of observations for the variables in the fertility family.

³³ It is, for instance, comparable to Blattman, Fiala, and Martinez (2014) who report an 82 percent effective tracking rate for a 4-year follow-up survey of young adults in Uganda; the 10-year follow-up of the Kenya Longitudinal Panel Survey with an effective tracking rate of 84 percent (Baird et al., 2016); and the 88 percent effective tracking rate for children after 5-7 years in the Moving to Opportunity evaluation in the United States (Orr et al., 2003). Our tracking efforts, however, were less successful than the 8-year follow-up of a scholarship program in Ghana (Duflo, Dupas, and Kremer, 2017), who like us use intensive tracking for the entire sample of those not found during regular tracking, but where special protocols to track respondents were incorporated into the RCT design from baseline including maintaining regular contact with respondents throughout the duration of the study, resulting in a 98 percent tracking rate.

³⁴ Behrman, Parker, and Todd (2011) and Adhvaryu et al. (2018) use the 6-year follow up of Mexico’s PROGRESA evaluation sample, with an attrition rate of 40 percent for individual-level information on comparable age groups. The 10-year PROGRESA follow-up (used for instance in Kugler and Rojas, 2018) has more than 60 percent

Table F1 shows that attrition rates are well balanced between the early and late treatment groups with the coefficients on the ITT indicator on the probability of having been found, i.e., interviewed, smaller than |1.5| percentage points for both the household and individual survey instruments. Table A1 confirms that this resulted in a sample that is balanced on baseline characteristics, with only 4 out of the 52 baseline variables (from the 2000 program census) examined significantly different at the 10 percent level, marginal differences that were also observed for the full sample. To further assess the possibility of selective attrition, for each baseline variable we estimate $Prob(found) = F(X, T, XT)$. As seen in Table F2, none of the estimated coefficients on T are significant and only a few on XT are, confirming that overall, both the rates of attrition and the observable characteristics of individuals who attrited are similar between the two treatment groups. At the same time, the coefficients on X make clear that those who attrited are different than those found along a number of dimensions. This implies that, even if internal validity is not jeopardized, the differential ITT estimates may not be representative of the effects for the full target population, i.e., external validity is a potential concern. This is particularly relevant if treatment heterogeneity is correlated with attrition selection. Because marriage and labor market opportunities are the two main reasons for migration (and consequently for attrition) of young women in this context, and also principal outcome variables in this paper, this is a potential concern for analyses.

In our preferred specifications, therefore, we specifically account for such selection using inverse probability weights constructed as described below. We also present a number of alternative estimations to examine the sensitivity of the findings to different assumptions about attrition. First, because attrition rates and baseline characteristics of the interviewed samples are balanced across treatment groups, we re-estimate the main results without any attrition corrections (appendix B, Tables B2.1-B2.4). Results from that exercise demonstrate that the main findings are not driven by the attrition correction and point in the same direction for all the families of outcomes (as also summarized in panel E of Table B1.1). Second, we estimate Lee bounds (Table B1.2). These rely on the monotonicity assumption, implying that assignment to the treatment group can affect attrition in only one direction. Because attrition may be correlated with marriage, labor market, and migration outcomes, and each of those could be affected by early and late treatment exposure in various ways, the validity of the monotonicity assumption in this setting is unclear (but Lee bounds are nevertheless presented for completeness). Third, for selected available outcomes we present results using proxy measures, which leads to much lower rates of missing information, but also likely introduces measurement error in the outcome variables (Table A.5 and A.6).

Attrition selection correction with inverse probability weights³⁵

Because several baseline characteristics are correlated with the probability of having been found, at least some of the potential attrition selection is likely related to observables. To account for such selection, we calculate attrition-corrected weights. In our preferred specification, we use a modified version of the standard inverse probability weighting adjustment that more fully exploits information obtained during the intensive tracking phase, and allows putting higher

attrition. Attrition rates in Baird, McIntosh, and Özler (2018) are 13-16 percent after five years for young women, and the 10-year follow-up of a much younger cohort in Ecuador has 19 percent attrition (Araujo, Bosch, and Schady, 2018).

³⁵ This section draws from Molina Millán and Macours (2017), which contains a more detailed explanation of the approach taken and further rationale for the selection correction.

weight on individuals who were more difficult to find. The key assumption underlying this strategy is that the probability of being found during the intensive tracking phase is explained by observable characteristics. Overweighting individuals whose observed characteristics predict they were more difficult to find corrects for the sample selection. We assign a weight of 1 to individuals found during the regular tracking phase, and only estimate the weights for those found during the intensive tracking phase. To determine the weights, we estimate the probability of being found for those found during the intensive tracking phase.

We calculate attrition-correction weights separately for each survey instrument. A large number of socio-economic variables observed in or calculated from the program census was considered for predicting attrition, informed in particular by the nature of migration from the regions. These include all of the baseline variables shown in Table F2, capturing individual-, household-, and locality-level characteristics. As connectedness to the locality could be a good predictor of tracking success, we included two variables to capture the social network of the individual (village size and family network size), some more detailed household structure variables, and a set of proxy variables meant to capture the possible temporary nature of residency in the village at baseline for some households.³⁶ We similarly consider locality-level characteristics that could be push or pull factors for migration: remoteness (measured using distance to night light and altitude), location in a coffee producing area, and having been affected by hurricane Mitch, a severe storm in the area in 1998. Finally, as further proxy measures for locations with a concentration of more temporary residents, we introduce two variables capturing the level of attrition between the program census and the first baseline survey (i.e., between May and August of 2000): 1) the share of individuals in the locality that attrited; and 2) whether any individual in the target age group attrited.³⁷ Individuals from such localities not only were more likely to attrit, but also could be more difficult to trace, as contacts with the community of origin could be limited.

Because of the large number of potential variables to consider and because there are relatively few individuals not found after intensive tracking, we follow the approach of Doyle et al. (2017) to select a reduced set of predictors. Separate estimates are carried out to model attrition for the household and individual instruments. First, we estimate bivariate regressions in which each potential predictor is examined to determine whether a significant difference exists between the means for those found and not found during intensive tracking. All estimates use the survey sample weights and standard errors are clustered at the locality level. This testing is conducted separately for the early and late treatment groups. Results are shown in the first four columns of Table F3 for attrition in the household instrument and in the final four columns for the individual instrument. The correlates of having been found during intensive tracking differ between treatment groups and also between the survey instruments.

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

³⁶ In particular, we include a set of indicators as proxies for whether the household comprised temporary workers on one of the large coffee plantations known as *haciendas*. These households were captured by the program census but not likely to have been permanent residents of the areas.

³⁷ The baseline survey was conducted shortly after the public lottery and before the start of the transfers.

information criterion. Strata and regional fixed effects, as well as 6-month age dummies are included as fixed predictors in all models.

In the final step, we estimate the probability of having been found during the intensive tracking phase for both early and late treatment group together, keeping only the predictors as indicated by the stepwise procedure, as well as the strata and regional fixed effects and the 6-month age dummies, all interacted with the treatment variable. Table F4 presents the linear probability model estimates for each survey. The resulting regression has good predictive power (R^2 of 36 percent for household level attrition and 34 percent for individual level attrition).

The probability of having been found during intensive tracking (conditional on not having been found during regular tracking) estimated via a probit for the specification shown in Table F4 is then used to determine the weights for the attrition selection. All observations found during regular tracking are assigned a weight of 1, while those found during the intensive tracking are assigned a weight $1/\text{Prob}(\text{found during intensive tracking}|\text{not having been found during regular tracking})$. Finally, these weights are then multiplied with the sample weights. Final attrition-correction weights vary between 1 and 32 for the household instrument, and 1 and 89 for the individual instrument, with the average being 4.³⁸

We overweight individuals interviewed during the intensive tracking phase, as these individuals were, by definition, more difficult to find, and therefore more likely to be similar to those not found at all. Empirically, observable characteristics are also better predictors for the subsample that was intensively tracked than for the full sample, indicating that selection on observables for this subsample is a more plausible assumption than for the full sample. Nevertheless, we also present results with the standard inverse probability weighting (with weights estimated using the probability of being found in the entire sample) in Table B1.1, for comparison. Estimations of weights for the standard IPW followed a similar process of covariate selection.

³⁸ With the exception of a few outliers, the distribution of weights is not highly skewed with 97 percent of household and 95 percent of individual weights less than nine. Only two observations have individual weights higher than 32 and omitting these two observations from the analysis does not alter any of the findings. The distribution of weights is similar when using conventional IPW estimates.

TABLE F1. ATTRITION AND TRACKING IN THE EARLY AND LATE TREATMENT GROUPS

<i>PANEL A</i>	Probability of having been interviewed: Household instrument			
	(1)	(2)	(3)	(4)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)	Found (incorporating proxy information)
ITT	-0.010 (0.038)	-0.005 (0.038)	-0.019 (0.073)	-0.021 (0.022)
Mean late treatment	0.849	0.584	0.636	0.958
Observations	1062	1062	444	1062
<i>PANEL B</i>	Probability of having been interviewed: Individual instrument			
	(1)	(2)	(3)	(4)
	Found (i.e., interviewed)	Found during regular tracking	Found during intensive tracking (intensive tracking subsample)	Found and with reproductive health module
ITT	-0.014 (0.041)	0.011 (0.039)	-0.034 (0.067)	-0.014 (0.043)
Mean late treatment	0.786	0.444	0.615	0.769
Observations	1062	1062	585	1062

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. Standard errors are clustered at the locality level and given in parentheses. Girls age 9-12 at the start of the program in November 2000.

TABLE F2. RELATIONSHIP BETWEEN THE PROBABILITY OF BEING FOUND, BASELINE COVARIATES, AND TREATMENT

	<i>X</i>		<i>X*T</i>		<i>T</i>	
<i>Individual Characteristics</i>						
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	-0.00504	(0.023)	0.00735	(0.028)	-0.0152	(0.056)
Worked in last week (=1)	-0.0171	(0.10)	-0.0987	(0.22)	-0.00241	(0.038)
Participated in some economic activity (=1)	-0.00645	(0.069)	0.00932	(0.11)	-0.00610	(0.038)
<i>Household Characteristics: Education</i>						
Distance to nearest school (minutes)	-0.00131*	(0.00077)	0.00151*	(0.00085)	-0.0423	(0.044)
Household head no grades attained (=1)	0.114***	(0.031)	-0.0869	(0.055)	0.0331	(0.053)
Household head 3 plus grades attained (=1)	-0.0166	(0.047)	-0.0204	(0.074)	-0.0000973	(0.047)
Mother no grades attained (=1)	0.0610	(0.050)	0.0566	(0.067)	-0.0287	(0.058)
Mother 3 plus grades attained (=1)	0.000570	(0.061)	-0.102	(0.093)	0.0312	(0.048)
<i>Household Characteristics: Demographics</i>						
Father not living in same household (=1)	-0.00617	(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)	-0.00307	(0.038)
Child of household head (=1)	-0.00397	(0.034)	0.122	(0.096)	-0.108	(0.10)
Number of children of household head	0.00197	(0.0065)	0.0134	(0.0097)	-0.0711	(0.071)
Female household head (=1)	-0.00858	(0.050)	-0.0441	(0.074)	0.00202	(0.043)
Age of household head	0.00288**	(0.0013)	-0.000745	(0.0027)	0.0228	(0.13)
Number of household members	0.00551	(0.0055)	-0.00469	(0.0089)	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)	0.00529	(0.083)	-0.00588	(0.055)
Other household structure (=1)	0.00110	(0.053)	-0.130	(0.10)	0.0123	(0.038)
Number of children aged 0-8	-0.00761	(0.012)	-0.00673	(0.020)	0.00872	(0.057)
Number of children age 9 to 12	-0.0367	(0.034)	0.00138	(0.061)	-0.00817	(0.10)
<i>Household Characteristics: Economic Activities&Assets</i>						
Household head main occupation is agriculture (=1)	0.00682	(0.043)	0.000379	(0.061)	-0.00536	(0.072)
Size of landholdings ('000 sq meters)	-0.000439	(0.0013)	0.00145	(0.0016)	-0.0316	(0.058)
Log of size of landholdings	0.00979	(0.0073)	0.0179	(0.011)	-0.149	(0.11)

Number of parcels of land	0.0472	(0.035)	0.124*	(0.065)	-0.120	(0.087)
Log predicted expenditures (pc)	0.0591	(0.057)	-0.0643	(0.073)	0.493	(0.57)
Wealth index - housing characteristics	0.00389	(0.017)	-0.0255	(0.027)	-0.00467	(0.040)
Wealth index - productive assets	0.0160	(0.012)	0.0161	(0.029)	-0.00393	(0.039)
Wealth index - other assets	0.0302*	(0.018)	-0.0246	(0.025)	-0.00380	(0.039)
<i>Village Characteristics</i>						
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)	-0.00575*	(0.0029)	-0.00203	(0.0034)	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)	-0.000365	(0.043)
<i>Social Capital</i>						
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)	0.00452	(0.061)
<i>Proxy's of Permanent Residence in Village</i>						
Own house (=1)	0.175**	(0.074)	-0.0368	(0.089)	0.0294	(0.093)
House is obtained in exchange for service/labor (=1)	-0.287***	(0.078)	0.0970	(0.12)	-0.0130	(0.036)
Address in hacienda (=1)	-0.185***	(0.059)	0.130	(0.091)	-0.0282	(0.042)
Address in hacienda & house rented (=1)	-0.334***	(0.087)	0.247**	(0.12)	-0.0233	(0.040)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Girls 9-12 years old at the start of the program in November 2000.

TABLE F3. CORRELATES OF THE PROBABILITY OF BEING FOUND DURING THE INTENSIVE TRACKING PHASE

	Found in intensive phase household instr.				Found in intensive phase individual instr.			
	Treatment		Control		Treatment		Control	
	N=		N=		N=		N=	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual Characteristics								
Age at start of transfer in months	-0.072	(0.18)	-0.498**	(0.23)	-0.227	(0.15)	-0.481***	(0.16)
No grades attained (=1)	0.071	(0.10)	-0.132	(0.086)	0.050	(0.096)	-0.040	(0.074)
Highest grade attained	-0.130	(0.24)	0.039	(0.36)	-0.144	(0.25)	-0.027	(0.28)
Worked in last week (=1)	-0.028	(0.045)	0.006	(0.037)	-0.003	(0.036)	0.011	(0.027)
Participated in some economic activity (=1)	-0.037	(0.045)	-0.004	(0.031)	0.008	(0.038)	-0.005	(0.041)
Household Characteristics: Education								
Distance to nearest school (minutes)	-0.136	(4.10)	-7.417	(5.01)	-5.924	(6.30)	-10.03**	(4.03)
Household head no grades attained (=1)	0.030	(0.097)	0.193***	(0.066)	0.060	(0.075)	0.122**	(0.056)
Household head 3 plus grades attained (=1)	-0.055	(0.097)	-0.011	(0.074)	-0.033	(0.090)	0.012	(0.080)
Mother no grades attained (=1)	0.220***	(0.072)	0.121	(0.096)	0.127**	(0.059)	0.072	(0.089)
Mother 3 plus grades attained (=1)	-0.187	(0.14)	0.006	(0.10)	-0.104	(0.12)	0.068	(0.11)
Household Characteristics: Demographics								
Father not living in same household (=1)	-0.175	(0.10)	0.046	(0.064)	-0.180**	(0.074)	-0.005	(0.063)
Mother not living in same household (=1)	-0.060	(0.054)	-0.055	(0.036)	-0.067	(0.044)	-0.011	(0.034)
Child of household head (=1)	0.097	(0.084)	-0.011	(0.039)	0.156**	(0.067)	-0.028	(0.041)
Number of children of household head	0.967**	(0.36)	0.090	(0.25)	0.652	(0.42)	-0.070	(0.26)
Female household head (=1)	-0.110*	(0.055)	-0.010	(0.030)	-0.109**	(0.046)	-0.041	(0.036)
Age of household head	1.437	(2.70)	2.465	(1.53)	0.208	(1.88)	2.132*	(1.21)
Number of household members	0.533	(0.55)	0.318	(0.32)	0.002	(0.48)	0.294	(0.31)
Nuclear household (=1)	0.037	(0.099)	-0.058	(0.057)	0.121*	(0.062)	-0.097	(0.068)
Multigenerational household (=1)	0.093	(0.100)	0.054	(0.064)	-0.012	(0.070)	0.107*	(0.062)
Other household structure (=1)	-0.130	(0.093)	0.005	(0.045)	-0.109	(0.069)	-0.010	(0.037)

Number of children ages 0-8	0.008	(0.32)	-0.166	(0.12)	-0.056	(0.22)	-0.170	(0.15)
Number of children ages 9 to 12	-0.100	(0.20)	0.051	(0.21)	-0.054	(0.18)	0.075	(0.16)
<i>Household Characteristics: Economic Activities&Assets</i>								
Household head main occupation is agriculture (=1)	0.011	(0.071)	0.039	(0.042)	0.003	(0.075)	0.043	(0.041)
Size of landholdings ('000 sq meters)	5.517	(4.85)	1.123	(4.88)	-0.695	(5.33)	-1.696	(3.52)
Log of size of landholdings	2.758**	(1.22)	1.023	(0.74)	1.672**	(0.76)	0.570	(0.60)
Number of parcels of land	0.353**	(0.14)	0.125	(0.087)	0.192*	(0.11)	0.103	(0.073)
Log predicted expenditures (pc)	-0.047	(0.064)	0.053	(0.046)	-0.044	(0.052)	0.0902*	(0.045)
Wealth index - housing characteristics	-0.422	(0.44)	0.244	(0.31)	-0.511	(0.33)	0.203	(0.31)
Wealth index - productive assets	0.366	(0.30)	0.133	(0.20)	0.339	(0.21)	0.152	(0.14)
Wealth index - other assets	0.182	(0.16)	0.248	(0.17)	0.186	(0.15)	0.475**	(0.20)
<i>Village Characteristics</i>								
Village affected by hurricane Mitch (=1)	-0.049	(0.11)	0.072	(0.052)	0.015	(0.082)	0.062	(0.053)
Altitude of village	16.200	(38.0)	48.57*	(27.9)	2.014	(27.2)	57.15**	(26.7)
Village in coffee producing area (=1)	-0.090	(0.066)	-0.097	(0.061)	-0.145***	(0.049)	-0.074	(0.044)
Distance to night light (meters)	-5607**	(2107)	-2227	(1655)	-5869***	(1938)	-2520	(1596)
Live in Tuma region (=1)	-0.294***	(0.077)	-0.216**	(0.086)	-0.334***	(0.075)	-0.248***	(0.078)
Live in Madriz region (=1)	0.127*	(0.072)	0.150*	(0.078)	0.121*	(0.061)	0.148*	(0.080)
<i>Social Capital</i>								
Family network size (individuals)	42.70***	(10.9)	25.50***	(8.91)	41.38***	(9.01)	36.66**	(15.6)
Population size village	30.330	(132)	31.290	(34.2)	-6.363	(105)	30.200	(31.6)
<i>Proxy's of Permanent Residence in Village</i>								
Own house (=1)	0.161**	(0.077)	0.164*	(0.093)	0.105**	(0.045)	0.133*	(0.077)
House is obtained in exchange for service/labor (=1)	-0.059	(0.070)	-0.155*	(0.083)	-0.077	(0.045)	-0.129*	(0.072)
Address in hacienda (=1)	0.001	(0.033)	-0.142	(0.096)	-0.018	(0.032)	-0.129*	(0.070)
Address in hacienda & house rented (=1)	-0.009	(0.019)	-0.141*	(0.068)	-0.013	(0.029)	-0.109**	(0.050)
<i>Variables Based on Short-term Evaluation Surveys Indicating Very Early Attrition</i>								
Locality average attrition prior to program start	0.012	(0.009)	-0.022	(0.020)	0.002	(0.010)	-0.028	(0.021)
No individual in targeted age cohort had attrited before program start	-0.143	(0.11)	0.199**	(0.077)	-0.029	(0.11)	0.204**	(0.073)

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. Girls ages 9-12 at the start of the program in November 2000.

TABLE F4. LINEAR PROBABILITY ESTIMATES FOR PROBABILITY OF BEING FOUND DURING INTENSIVE TRACKING PHASE

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

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

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and in parentheses. For each model, column (a) reports coefficient on variable alone, and column (b) the coefficient on the variable interacted with the early treatment dummy. Girls ages 9-12 at the start of the program in November 2000.

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

At some point during the period 2000-05, the three-year CCT operated in *all* rural areas of the six municipalities where the evaluation took place. This included the 42 rural localities randomized into early and late treatment as well as the 17 rural localities not initially targeted because they were less poor in 1995 according to the marginality index used for geographic targeting. In these additional 17 localities, the program began in 2001 and was offered to 80 percent of the population based on a household-level proxy means targeting model. Consequently, by 2005 the three-year program had been implemented (at different times and to modestly different degrees) in all 59 rural localities in the six municipalities, covering over 90 percent of the rural population in them. (See appendix C for further details.) Given this high coverage, it is possible to use national census data to provide evidence on absolute program effects, albeit for a somewhat limited set of educational and demographic outcomes available in the census.

Specifically, we use the 1995 and 2005 censuses and a non-experimental double difference approach to estimate absolute program effects five years after the CCT began in the early treatment group. Together, the two 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.³⁹ The censuses provide information on current municipality of residence (and whether rural or urban), as well as municipality at birth and municipality of residence five years prior to the census administration date.⁴⁰ We assign all individuals to the municipality where they lived 5 years prior to each census (about 3 percent of the 9-12 year olds moved in that period), and assume the type of prior residence (rural or urban) is the same as the current residence. We calculate double difference impacts using relevant age cohorts (calculating ages on November 1 as done for the main analyses, in 1990 and in 2000) and comparing outcomes in rural areas of the six program municipalities to outcomes in rural areas of the six neighboring municipalities where the non-experimental comparison group was selected in 2002. The 9–12 age cohort in 2000 is the same cohort examined in the main experimental analyses. In addition, we provide estimates for the 7-8 age cohort and 13-year olds, covering the entire age range of children potentially eligible for the education component of the CCT. We estimate

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

Where Y_{imt} is the educational outcome for child i in municipality m measured in census year t , $T_{m,t-5}$ is an indicator for whether the child resided in a treatment municipality five years prior to the census year, and C_t an indicator for the 2005 census. δ_3 yields the double difference estimate five years after the program began on Y , which includes grades attained, enrollment and literacy.⁴¹ Standard errors are robust to heteroskedasticity.

³⁹ It is not possible to link individuals across the two census rounds.

⁴⁰ While the data also include more detailed location information, changes in the definition and boundaries of census areas between 1995 and 2005 make it impossible to match them across time. Municipality boundaries, however, remained constant.

⁴¹ δ_3 is the average impact for the three different groups—early treatment, late treatment and, the other 17 localities—each of which by 2005 had received the three-year program at different times. It is not possible to isolate the three distinct treatment group areas within the census data (see previous footnote).

Because the CCT did not operate in urban areas, all estimates limit the sample to individuals living in rural areas. The main double difference estimation equation takes a first difference between outcomes measured in 2005 for those living in program and non-program comparison municipalities in 2000, and a second difference between outcomes measured in 1995 for those living in program and non-program municipalities in 1990, as indicated by municipality of residence five years prior. These main results are presented in Table 6 in the text, and reproduced below.

In Table G1 we present the double difference impacts after five years for the three age cohorts (7-8, 9-12, and 13). With the exception of ever having been married (virtually none of the 7-8 cohort are by 2005 given their relatively young ages), effects for the younger cohort are similar in magnitude and significance to the 9-12 year olds, consistent with the program having had a positive absolute effect on these girls. Evidence of effects for the 13-year olds is less strong, but not significantly different from the 9-12 year olds.

We next explore the sensitivity of the main double difference findings to different sets of comparison municipalities and definitions for treatment status (Table G2). First, we expand the comparison municipalities to include rural areas in all non-program municipalities in the central regions of Nicaragua where the program was located (panel B). Second, we examine whether results differ when we instead use current residence (panel C) or, separately, municipality of birth (panel D) to determine program eligibility. While there are some differences in point estimates, the different approaches all suggest significant absolute impacts of the program on educational (+), civil status (-), and fertility (-) outcomes after five years.

Last, we provide evidence in support of the identifying assumption by estimating program “effects” on outcomes for a different cohort unlikely to have been affected by the intervention—household heads in households with a child in the 9–12 cohort (using one observation per household).⁴² The same empirical specification suggests there are no effects on their educational and demographic characteristics (Table G3). Moreover, point estimates are all close to zero, providing support for common trends.

⁴² We analyze common trends using household heads rather than an older age cohort of children because older children may still have been influenced by the program, including through migration patterns that are difficult to disentangle using the national census data.

TABLE G1: 2005 ABSOLUTE IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY, BY AGE
COHORT

	Grades Attained	Completed Grade 4 =1	Enrolled =1	Read and Write =1	Ever married =1	Has had live birth=1
	(1)	(2)	(3)	(4)	(5)	(6)
δ3 for 7-8 age cohort	0.418*** (0.081)	0.099*** (0.019)	0.028 (0.018)	0.084*** (0.017)	0.000 (0.002)	-
δ3 for 9-12 age cohort (as in Table 6)	0.536*** (0.084)	0.105*** (0.015)	0.028* (0.015)	0.066*** (0.013)	-0.022** (0.011)	-0.023** (0.010)
δ3 for 13 age cohort	0.320 (0.212)	0.079** (0.032)	-0.020 (0.026)	0.036 (0.029)	0.023 (0.033)	0.023 (0.033)
N for 7-8 age cohort	10,260	10,260	10,271	10,260	10,271	-
N for 9-12 age cohort	17,061	17,061	17,075	17,061	17,075	14,104
N for 13 age cohort	3,728	3,728	3,736	3,728	3,736	3,408
N Total	31,049	31,049	31,082	31,049	31,082	17,512
Mean comparison group 7-8	3.462	0.525	0.776	0.825	0.005	-
Mean comparison group 9-12	4.542	0.642	0.502	0.830	0.162	0.085
Mean comparison group 13	4.754	0.620	0.260	0.785	0.464	0.350
P-value (7-8 vs 9-12)	0.312	0.788	0.996	0.387	0.053	-
P-value (9-12 vs 13)	0.343	0.450	0.114	0.349	0.191	0.190

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use national census data to compare rural areas of program municipalities to rural areas of the six comparison group municipalities. Birth information unavailable for 7-8 year old cohort (too young at time of census) and unreported (missing) for 7-17 percent of observations in other age cohorts. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include girls at the indicated ages in November of 1990 and 2000, respectively.

TABLE G2: 2005 ABSOLUTE IMPACTS ON EDUCATION, CIVIL STATUS AND FERTILITY, 9-12 AGE COHORT

	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)
<i>Panel A: Treatment Municipality (5 Years Prior) vs 6 Municipality Comparison Group</i>						
Treatment municipality * 2005 (δ_3) (as in Table 6)	0.536*** (0.084)	0.105*** (0.015)	0.028* (0.015)	0.066*** (0.013)	-0.022** (0.011)	-0.023** (0.010)
N	17,061	17,061	17,075	17,061	17,075	14,104
Mean comparison group 2005	4.542	0.642	0.502	0.830	0.162	0.085
<i>Panel B: Treatment Municipality (5 Years Prior) vs Central Regions Comparison Group</i>						
Treatment municipality * 2005 (δ_3)	0.319*** (0.058)	0.069*** (0.011)	0.046*** (0.011)	0.052*** (0.009)	-0.017** (0.008)	-0.003 (0.008)
N	78,811	78,811	78,907	78,806	78,907	65,194
Mean comparison group 2005	4.540	0.655	0.465	0.838	0.184	0.089
<i>Panel C: Treatment Municipality (Current) vs 6 Municipality Comparison Group</i>						
Treatment municipality * 2005 (δ_3)	0.505*** (0.084)	0.098*** (0.015)	0.023 (0.015)	0.064** (0.013)	-0.029** (0.011)	-0.023** (0.011)
N	17,038	17,038	17,053	17,040	17,053	14,118
Mean comparison group 2005	4.554	0.645	0.500	0.832	0.171	0.088
<i>Panel D: Treatment Municipality (of Birth) vs 6 Municipality Comparison Group</i>						
Treatment municipality * 2005 (δ_3)	0.500*** (0.085)	0.096*** (0.015)	0.024 (0.015)	0.059*** (0.013)	-0.020* (0.011)	-0.025** (0.011)
N	16,764	16,764	16,779	16,767	16,779	13,915
Mean comparison group 2005	4.493	0.636	0.497	0.826	0.165	0.086

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use census data to compare rural areas of program municipalities to rural areas of other municipalities. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include girls ages 9-12 in November of 1990 and 2000, respectively.

TABLE G3: 2005 ABSOLUTE IMPACTS ON HOUSEHOLD HEAD CHARACTERISTICS

	Grades Attained	Completed Grade 4 =1	Read and Write =1	Female Head =1	Age of Head in Years
	(1)	(2)	(3)	(4)	(5)
Treatment municipality * 2005 (δ_3)	-0.020 (0.079)	0.014 (0.013)	0.005 (0.016)	0.007 (0.013)	0.342 (0.417)
N	15,192	15,192	15,286	15,292	15,292
Mean comparison group 2005	1.673	0.226	0.507	0.221	46.901

Notes: *** p<0.01, ** p<0.05, * p<0.10. 2005 absolute effects use census data to compare rural areas of program municipalities to rural areas of other municipalities. The mean of the comparison group is for the six comparison group municipalities. Sample includes household heads in households with a child in the 9-12 age cohort. Heteroskedasticity-robust standard errors are given in parentheses. All variables measured using the 1995 and 2005 population censuses and include heads from households with a boy or girl age 9-12 in November of 1990 and 2000, respectively.

APPENDIX H: NON-EXPERIMENTAL ABSOLUTE PROGRAM EFFECTS —MATCHING

To estimate the absolute effects of the program, we compare 2010 outcomes for girls ages 9–12 at baseline in the early and late treatment groups to the 2010 outcomes for the same cohort of girls in a non-experimental comparison group. In 2002, prior to the phase-in of the late treatment group, 21 non-experimental comparison localities from neighboring rural municipalities were added to enhance the potential for an evaluation of the longer-term effects of the program. The principal criteria for selection included: 1) the same marginality index score cut-offs from the Nicaraguan national census used in the selection of the original 42 localities; 2) minimal ongoing or planned development interventions related to the CCT's objectives; and 3) coverage of the geographic regions of the original municipalities.⁴³ The comparison group was surveyed in 2002, 2004, and 2010 using the same survey instruments as the experimental groups.

As the *ex ante* match of the comparison area to the program areas on locality-level characteristics may not be sufficient to balance household and individual characteristics, we estimate the absolute effects using five nearest neighbors matching (NN5), one or two nearest neighbors matching (NN1, NN2), and kernel and local linear matching. We draw on the relatively rich data available and include household and individual characteristics in the propensity score. The matching aims to balance these observable characteristics. To satisfy the unconfoundedness assumption we must assume balance of the unobservables (Rosenbaum and Rubin 1983), as required with any non-experimental method.

The nearest neighbor matching estimators are bias adjusted (Abadie and Imbens 2006; Imbens 2015) and standard errors are clustered at the locality level using the analytic asymptotic variance estimator developed by Abadie and Imbens (2008, 2011) that accounts for the fact that the propensity score is estimated. For kernel and local linear estimates, the standard errors are bootstrapped and the bandwidth is set to be small (0.06) to limit the bias (Todd 2007). We estimate average treatment effects on the treated (ATT) that matches all girls who are 9–12 at the start of the program in the early treatment group with same aged girls in the comparison group. For more direct comparisons with the ITT experimental differential results, the treatment group includes all girls who were eligible for treatment, regardless of whether they had taken up the treatment (i.e., it includes the non-compliers).

Estimation of Propensity Score—To estimate the propensity score, we combine data from the 2000 program census, with data from the 2002 household survey for the non-experimental comparison group. There are two important caveats. First, the baseline data used to determine the propensity score for the early and late treatment versus the comparison groups are from different years: 2000 for the early and late treatment groups and 2002 for the comparison group. We do not use 2002 data for the treatment groups because the early group had already received two

⁴³ More specifically, the comparison sample was drawn from rural municipalities adjacent to or neighboring the six original municipalities. Six comparison municipalities without any major planned development initiatives but with similar levels of poverty and density of schools and health clinics were selected to capture the geographic diversity of the original municipalities. After excluding a small number of localities, the same marginality index used to select the original 42 localities (Arcia 1999; Maluccio 2009), and based on the 1995 Nicaraguan National Census, was calculated for each remaining rural locality. From this exercise, 22 localities with marginality scores in the range targeted by the CCT were identified; one locality that was further way, and thus less likely to be similar, was dropped, leaving 21 comparison localities. A random sample of households was drawn in each. For additional details, see IFPRI (2005).

years of CCT benefits by 2002. We argue the difference in the timing of the surveys is not likely to be a major source of bias as the value of the variables used in the propensity score are unlikely to have changed much between 2000 and 2002 (e.g., mother’s age, mother’s and household head’s years of education, head’s gender, distance to the municipality center).⁴⁴

Second, the data come from different types of survey instruments; census and household surveys.⁴⁵ We use the 2000 program census data, rather than the 2000 baseline household survey, in order to include the oversample group in the estimate of the propensity score. The inclusion of the oversample is important for comparability with the differential experimental estimates and also increases the precision of the propensity score estimate. The 2000 program census has a more limited set of variables though all questions in the census and survey instrument are similar for the variables included in the propensity score.

The logit model used to estimate the propensity score is presented in Appendix Table H1. We estimate the propensity score using data on girls who are 9–12 at baseline from *both* early and late treatment groups and the comparison group. We use all available variables that are similar between the 2000 program census and 2002 household survey and important predictors of either treatment status or the outcomes of interest. We did not include variables whose values are likely to have changed between 2000 and 2002, binary variables which did not have sufficient variation, and information about fathers that was incomplete (e.g., father’s age at baseline was missing for more than 20 percent of the sample because it was only asked if the father was a resident of the same household).^{46,47} Because of the two-year gap in measurement, it is not possible to consistently measure baseline education variables at the same point in time. As a result, we do not include grades attained or enrollment in the main propensity score, but do include year of birth fixed effects, which are correlated with the education variables.

Propensity Score Balance—We follow Dehejia and Wahba (1999) to determine if the propensity score is balanced across the non-experimental groups and use initial estimates as guides to include interactions or polynomials of variables in the propensity score. We divide the common support into five blocks and test that the propensity score, and each of the variables in the propensity score, are balanced within each block using a t-test.

Appendix Figure H1 presents the distributions of the estimated individual-level propensity score model. Observations above the x-axis are from the treatment groups, and those below from the comparison group. In contrast to what we might see had the groups been randomly allocated, the overlap, while substantial, is imperfect with the treatment group skewed to the right and the comparison group to the left. Matching estimators address and correct for this difference in the distributions. To improve the overlap in the covariate distribution, we restrict the analysis to the common support between the experimental treatment and comparison groups. Often a “min-max”

⁴⁴ When appropriate, comparison group data from 2002 was adjusted in order to be consistent with the data for the early treatment group coming from 2000. For example, age of mother and the child were calculated for the same year, 2000.

⁴⁵ An important exception is the locality level marginality index, which is based on the 1995 national census for both the early treatment and the comparison group.

⁴⁶ We use mother’s education from the 2010 survey because it is more complete. In the 2010 survey mother’s education was collected for all household members, while in the earlier surveys it was collected for all individuals on the household roster, so mother’s education was only available if the mother was living in the same household.

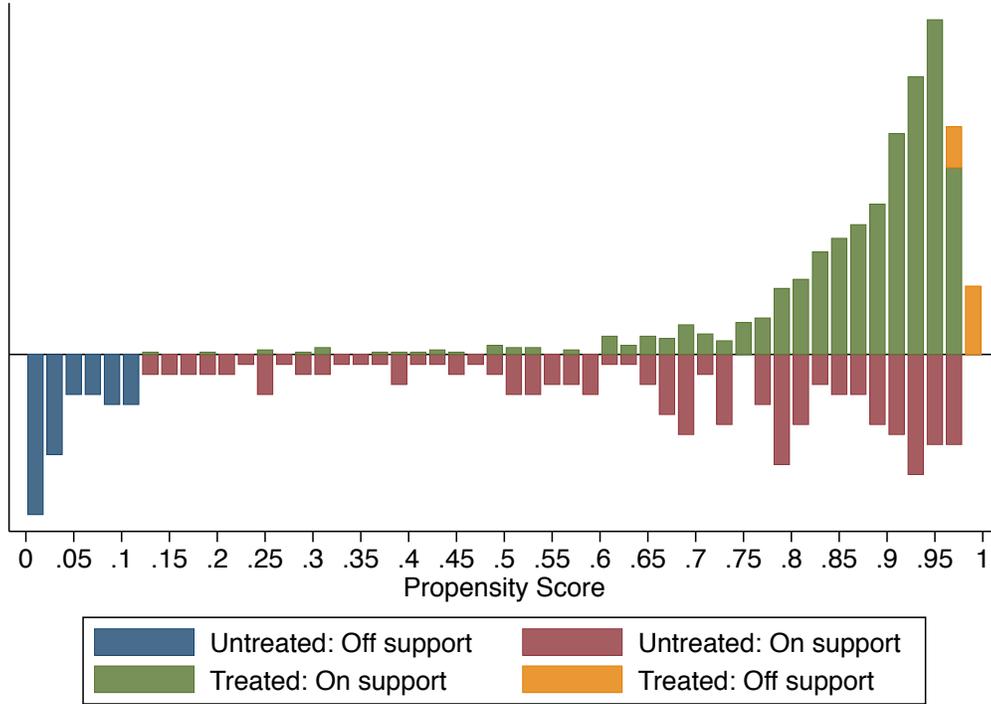
⁴⁷ We further do not include the household asset index because not all of the variables are available, however we include the variables that are the same between the two surveys and likely to be correlated with the outcomes.

common support is used (trim all observations that have a propensity score lower than the minimum of the treatment group distribution, as well as, observations whose propensity score value is greater than the maximum value of the comparison group). However, due to the sparseness of treatment observations on the far left of the distribution, this common support does not leave the propensity score balanced by block. Instead, we estimate results with the common support defined by trimming all observations that have a propensity score lower than the first percentiles (propensity score equal to 0.37), rather than the minimum value. We refer to this common support as the 1 percent trim. As a robustness check, we also examine the trimming at the second percentile (propensity score equal to 0.496), or 2 percent trim.

In Table H2 we test the balance of the matching using the 1 percent trim. Columns 1–5 present the p-value on the t-test of the difference in the propensity score and each of the variables used to make the propensity score between the treatment and comparison groups by block. They indicate that the p-score is balanced between the treatment and comparison groups in all blocks, though the difference in the p-score in block 1 is marginally significant at the 10 percent level. There are 6 variables that are statistically different within any block at the 5 percent level or lower (less than 8 percent of the block-variable combinations). We further test whether the treatment and comparison groups are balanced using the NN5 matching estimator on the variables in the propensity score. Table H2 columns 6 and 7 show that the baseline variables used in the propensity score are balanced between the treatment and comparison groups; the estimates are small and none are statistically significantly different at 10 percent level or below.

Absolute Program Effects—The absolute effects using a variety of matching estimators and trims are reported in Table H3. Results from Table 11 in the text are reproduced in panels A and B and panels C-F present further sensitivity analyses. First, given the sparse distribution of treatment observations in the left tail (Figure H1), we estimate results with a more stringent common support defined by trimming all observations that have a propensity score lower than the second percentile of the treatment group distribution in panel C, rather than the 1 percent trim used in panels A and B. Second, we consider alternative matching estimators (NN1, NN2, and local linear matching) in panels D-F. Overall, results are similar in magnitude regardless of which common support or matching estimator is used, though significance varies slightly by estimator as is expected (in particular, with higher standard errors for NN1).

FIGURE H1—INDIVIDUAL-LEVEL PROPENSITY SCORE DISTRIBUTION,
GIRLS 9–12 IN 2000



Notes: The treated include the early and late treatment groups together and the untreated the comparison group.

TABLE H1— LOGIT RESULTS FOR PROPENSITY SCORE MATCHING AT INDIVIDUAL LEVEL,
GIRLS 9–12 IN 2000

	Logit (1)	OLS (2)
Age 10 in 2000 (=1)	-0.652** (0.278)	-0.079*** (0.030)
Age 11 in 2000 (=1)	0.121 (0.272)	0.013 (0.027)
Age 12 in 2000 (=1)	-0.186 (0.302)	-0.031 (0.031)
1995 marginality index	636.259*** (80.638)	90.775*** (7.553)
1995 marginality index squared	0.467* (0.255)	0.052* (0.027)
Mom no education (=1)	18.102* (9.513)	2.323** (1.002)
Mom less than 3 years of education (=1)	-0.023* (0.014)	-0.002 (0.001)
Mothers age in 2000	-0.170 (0.308)	-0.020 (0.032)
Household head male (=1)	0.012 (0.060)	-0.000 (0.007)
Household head years of education	-0.097 (0.290)	-0.007 (0.031)
Household head has no education (=1)	0.103 (0.197)	0.034*** (0.012)
Family size	2.408*** (0.685)	0.280*** (0.073)
Family size Squared	-0.646*** (0.227)	-0.076*** (0.026)
Share of household members age 0-13	0.120 (0.260)	0.016 (0.027)
Has electric light (=1)	-0.350** (0.179)	-0.027 (0.018)
Has work animals (=1)	0.005 (0.012)	-0.001 (0.001)
Log (km to the municipality capital)	-70.403*** (9.051)	-10.037*** (0.850)
Mother no ed. * 1995 marginality index	-4.085* (2.137)	-0.525** (0.224)
R squared		
Observations	1,090	1,090

Notes: *** p<0.01, ** p<0.05, * p<0.10. The dependent variable is the treatment variable for matching, which is 1 if in the early or late treatment and 0 if in the comparison group. All variables measured using the 2000 program census.

TABLE H2: BASELINE BALANCE — BY BLOCK AND NN5 MATCHING

	P-Value of Difference in Mean By Block					NN5	
	1	2	3	4	5	Diff. in mean	P-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Propensity score	0.095	0.348	0.492	0.682	0.169		
Age 9 (=1)	0.367	0.155	0.968	0.992	0.381	0.003	0.938
Age 10 (=1)	0.909	0.865	0.279	0.700	0.876	-0.045	0.497
Age 11 (=1)	0.952	0.717	0.918	0.568	0.209	0.046	0.686
Age 12 (=1)	0.577	0.568	0.000	0.217	0.356	-0.004	0.946
Marginality Index	0.916	0.075	0.345	0.291	0.146	-0.085	0.959
Mother no grades attained (=1)	0.797	0.384	0.205	0.819	0.984	0.010	0.763
Mother 3 plus grades attained (=1)	0.677	0.475	0.398	0.901	0.109	0.039	0.297
Mother's age	0.323	0.923	0.072	0.409	0.927	0.031	0.986
Household head male (=1)	0.945	0.869	0.346	0.223	0.569	0.007	0.847
Household head years of education	0.882	0.978	0.187	0.000	0.436	0.237	0.320
Household head no grades attained (=1)	0.978	0.998	0.184	0.000	0.745	-0.051	0.430
Family size	0.257	0.397	0.396	0.750	0.763	0.083	0.894
Share of household members age 0-13	0.124	0.850	0.041	0.113	0.786	-0.010	0.444
Household has electric light (=1)	0.786	0.800	0.410	0.000	0.002	0.015	0.866
Household had work animals (=1)	0.749	0.174	0.462	0.282	0.527	0.003	0.957
Log distance to the municipality capital (km)	0.939	0.487	0.221	0.340	0.779	-0.149	0.446

Notes: *** p<0.01, ** p<0.05, * p<0.10. Compares baseline values from early and late treatment groups together (2000) to the comparison group (2002) where the common support uses a 1 percent trim. ATT biased adjusted estimator (Abadie and Imbens 2011) using five nearest neighbors. Standard errors on the differences are clustered at the locality level. All variables measured using the 2000 program census.

TABLE H3: 2010 ABSOLUTE IMPACT —ALTERNATIVE SPECIFICATION

	Economic Participation Family Z-Score	Earnings Family Z-Score		Fertility	Education		Learning Family Z-Score	Socio-emotional Family Z-Score
	(1)	Rank (2)	Absolute (5% Trim) (3)	(4)	Grades Attained (5)	Family Z-Score (6)	(7)	(8)
<i>Panel A: NN5 — 1 Percent Trim</i>								
ATT	0.040 (0.218)	0.184 (0.158)	0.119 (0.194)	-0.055 (0.191)	1.137 (0.708)	0.319* (0.177)	0.257* (0.140)	0.144*** (0.041)
N	969	969	960	886	969	969	905	899
<i>Panel B: Kernel Matching —1 Percent Trim</i>								
ATT	0.013 (0.159)	0.143 (0.100)	0.085 (0.125)	-0.039 (0.133)	1.227* (0.688)	0.336** (0.165)	0.284* (0.167)	0.156*** (0.059)
N	969	969	960	886	969	969	905	899
<i>Panel C: NN5 — 2 Percent Trim</i>								
ATT	0.039 (0.221)	0.185 (0.160)	0.116 (0.197)	-0.057 (0.193)	1.167 (0.714)	0.324* (0.178)	0.261* (0.141)	0.144*** (0.040)
N	950	950	941	870	950	950	889	883
<i>Panel D: NN1 —1 Percent Trim</i>								
ATT	-0.018 (0.259)	0.120 (0.194)	0.068 (0.234)	-0.107 (0.216)	1.217 (0.907)	0.342* (0.208)	0.335 (0.246)	0.126** (0.057)
N	969	969	960	886	969	969	905	899
<i>Panel E: NN2 —1 Percent Trim</i>								
ATT	0.016 (0.203)	0.145 (0.151)	0.116 (0.201)	-0.036 (0.236)	1.386** (0.697)	0.382** (0.174)	0.309* (0.180)	0.109*** (0.038)
N	969	969	960	886	969	969	905	899
<i>Panel F: Local Linear Regression —1 Percent Trim</i>								
ATT	0.007 (0.142)	0.131 (0.090)	0.094 (0.118)	-0.043 (0.122)	1.329* (0.716)	0.353** (0.169)	0.341** (0.172)	0.160*** (0.061)
N	969	969	960	886	969	969	905	899

Notes: *** p<0.01, ** p<0.05, * p<0.10. Standard errors are clustered at the locality level and are given in parentheses. Absolute effects compare early and late treatment together to comparison group in 2010. Mean of grades attained in the comparison group is 5.3 in 2010 for panel A. Z-scores are calculated as the mean and divided by the standard. Variables measured using the 2010 household survey, with the exception of the fertility family, which is based on variables from the 2010 individual survey.

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

The medical and nutrition literatures provide the scientific basis for understanding the role of nutritional status and nutritional shocks on the timing of the onset of puberty. A series of studies show that poor childhood nutrition is associated with delayed puberty for girls and better childhood health and nutrition associated with earlier menarche (Garn, 1987; Cooper et al 1996).⁴⁸ Originally hypothesized to be directly related to the attainment of a critical weight (Frisch and Revelle, 1971), more recent work indicates menarche is related to a minimum body fat mass (Blum et al, 1997). Nutritional status in childhood is believed to affect menarche through the leptin hormone (which helps regulate the body's energy balance⁴⁹), with some uncertainty as to whether leptin plays a permissive versus a triggering role (Shalitlin and Philip, 2003; INSERM, 2007). Leptin levels fluctuate with nutritional intakes, pointing to the mechanisms through which nutritional shocks can translate in relatively swift delays and accelerations in the age of menarche.⁵⁰

Nutrition may affect the onset of puberty differentially at different stages of childhood. A well known phenomenon of early puberty among migrant and adopted children (Mul et al, 2002; Parent et al., 2003) is often explained by an interaction of prenatal undernutrition with an enriched (later) childhood context (Gluckman and Hanson, 2006). Koziel and Jankowska (2002) show that girls with high BMI at age 14 are more likely to have had early menarche when they also had low birth weight, but not otherwise. Sloboda et al (2007) find that both low birth weight and high BMI at age 8 predict early age at menarche. To the best of our knowledge, however, the medical literature does not (yet) provide insights on the specific ages at which positive or negative nutrition shocks matter most. A review by Kaplowitz (2008), for example, recommends the need to study relationship between nutrition and puberty from ages as early as 6-7 up to 13-14. Cross-sectional and longitudinal evidence shows rises in leptin concentration between ages 7 and 15, paralleled by an increase in body fat during female puberty (in contrast with decreasing body fat during male puberty)—as well as strong correlations between leptin, body fat mass, and age of menarche (Blum et al, 1997; Garcia-Mayor et al, 1997; Ahmed et al, 1999).

Overall, numerous studies document the close correlation between body fat mass and the onset of puberty, even if it is not necessarily clear whether excess weight induces early sexual maturity, or whether early sexual maturity triggers excess weight (INSERM, 2007).⁵¹ The linkage between body fat and the reproductive system for girls has been interpreted as mechanisms for assuring that pregnancy will not occur unless there are adequate fat stores to sustain mother and fetus—with mammalian females being able to turn off their reproductive systems when the food supply is inadequate (Kaplowitz, 2008). Based on the evidence in the medical and nutrition literatures, we hence hypothesize that the differential timing of nutrition shocks in early versus late treatment groups could have led to differences in the age of menarche as well as adult BMI.

⁴⁸ Age of menarche declined from 17 to 14 in the United States and several Western European countries between the mid 19th and 20th centuries, a strong secular trend believed to be directly related to improvements in nutrition and health (INSERM, 2007).

⁴⁹ When fat stores are low, decreased leptin leads to increased appetite, helping to restore body fat and body weight.

⁵⁰ Various lab experiments with rodents and monkeys show that injection of leptin led to sudden changes in the age of menarche and reproductive outcomes (Chehab et al 1996, 1997; Cheung et al, 2001; Wilson et al 2003).

⁵¹ More generally, compensatory postnatal growth during childhood combined with poor prenatal growth is often argued to lead to adverse health outcomes (Hales et al, 2003).

Additional Appendix References

- Abadie, A., and G. Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects," *Econometrica* 74(1): 235–267.
- Abadie, A. and G. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76(6): 1537–1557.
- Abadie, A., and G. Imbens. 2011. "Bias-Corrected Matching Estimators for Average Treatment Effects," *Journal of Business & Economic Statistics* 29(1): 1–11.
- Adhvaryu, A., Molina, T., Nyshadham, A., and Tamayo, J. 2018. "Helping Children Catch Up: Early Life Shocks and the *PROGRESA* Experiment." Mimeo, University of Michigan.
- Ahmed ML, Ong KK, Morrell DJ, et al. 1999. "Longitudinal study of leptin concentrations during puberty: sex differences and relationship to changes in body composition." *J Clin Endocrinol Metab.* 84(3):899–905
- 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.
- Arcia, G. 1999. "*Proyecto de Red de Protección Social: Focalización de la fase piloto*" Report to the Inter-American Development Bank, Washington DC.
- Attanasio, O., S. Cattan, E. Fitzsimon, C. Meghir, and M. Codina. 2015. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." *IFS Working Paper* 15/06.
- Baird, S., J.H. Hicks, M. Kremer, and E. Miguel. 2016. "Worms at Work: Long-run Impacts of a Child Health Investment." *Quarterly Journal of Economics* 131 (4): 1637–1680.
- Blattman, C., N. Fiala, S. Martinez. 2014. "Generating Skilled Self-employment in Developing Countries: Experimental Evidence from Uganda." *Quarterly Journal of Economics*: 697-752.
- Chehab FF, Lim ME, Lu R. 1996. "Correction of the sterility defect in homozygous obese female mice by treatment with human recombinant leptin." *Nat Genet.* 12(3):318–320
- Chehab FF, Mounzih K, Lu R, Lim ME. 1997. "Early onset of reproductive function in normal female mice treated with leptin." *Science.* 275(5296):88–90
- Cheung CC, Thornton JE, Nurani SD, Clifton DK, Steiner RA. "A reassessment of leptin's role in triggering the onset of puberty in the rat and mouse." *Neuroendocrinology.* 2001;74(1): 12–21
- Cunha, F., J.J. Heckman, and S.M. Schennach. 2010. "Estimating the Technology of Cognitive and Non-Cognitive Skill Formation." *Econometrica* 78(3): 883–931.
- Dehejia, R., and S. Wahba. 1999. "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs". *Journal of the American Statistical Association* 94 (448): 1053–1062.
- 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.
- 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* 374:1997–2005.
- Filmer D, and L. Pritchett. 2001. “Estimating Wealth Effects Without Expenditure Data – or Tears: An Application to Educational Enrollments in States of India.” *Demography* 38(1): 115–132.
- Frisch RE and Revelle R. 1971. “Height and Weight at Menarche and a Hypothesis of Menarche.” *Arch Dis Child*, 46(249):695–701
- Garcia-Mayor RV, Andrade A, Rios M, Lage M, Dieguez C, Casanueva FF. “Serum Leptin Levels in Normal Children: Relationship to age, gender, body mass index, pituitary-gonadal hormones, and pubertal stage.” *J Clin Endocrinol Metab.* 1997; 82(9):2849–2855.
- Hales C.N. and S.E. Ozanne. 2003 "The Dangerous Road of Catch-up Growth." *J Physiol (Lond)* 547:5–10.
- Imbens, G. 2015. “Matching Methods in Practice: Three Examples.” *Journal of Human Resources*: 50(2): 373-419.
- IFPRI. 2005. *Sistema de Evaluación de la Red de Protección Social (RPS) – Mi Familia, Nicaragua: Evaluación de Impacto 2000–04*, Report submitted to the *Red de Protección Social*. International Food Policy Research Institute, Washington, DC. Photocopy.
- Kaplowitz, P.B, 2008. “Link Between Body Fat and the Timing of Puberty”, *Pediatrics*: 121; S208.
- Koziel S, and EA. Jankowska, 2002. “Effect of low versus normal birthweight on menarche in 14-year-old Polish girls.” *J Paediatr Child Health* 38:268–27.
- Kugler, A.D., and Rojas, I. 2018. “Do CCTs Improve Employment and Earnings in the Very Long-Term? Evidence from Mexico”. NBER Working Paper No.24248.
- Maluccio, J. A. 2009. “Household Targeting in Practice: The Nicaraguan Red de Protección Social.” *Journal of International Development* 21(1): 1–23.
- Orr, L. J. Feins, R. Jacob, E. Beecroft, L. Sanbonmatsu, L.F. Katz, J.B. Liebman, and J.R. Kling. 2003. *Moving to Opportunity: Interim Impacts Evaluation*. Unpublished.
- Paxson, C., and N. Schady. 2010. “Does Money Matter? The Effects of Cash Transfers on Child Health and Development in Rural Ecuador.” *Economic Development and Cultural Change* 59(1): 187–230.
- Radloff, Lenore S. 1977. “The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.” *Applied Psychological Measurement* 1 (3): 385–401.
- Regalia, F., and L. Castro. 2007. “Performance-Based Incentives for Health: Demand and Supply-Side Incentives in the Nicaraguan *Red de Protección Social*.” *Center for Global Development Working Paper* No. 119. Washington, DC, United States: Center for Global Development.

- Rosenbaum, P. and Rubin, D. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70: 41–50.
- Shalitin S, Phillip M. Role of obesity and leptin in the pubertal process and pubertal growth: a review. *Int J Obes (Lond)*. 2003; 27(8):869–874.
- Sloboda, DM, Hart R, Doherty DA, Pennell CE, Hickey M. 2007. "Age at menarche: influences of prenatal and postnatal growth", *Journal of Clinical Endocrinology and Metabolism* 92(1):46-50.
- Todd, P. 2007. "Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated." In *The Handbook of Development Economics* edited by. T. Paul Schultz and John A. Strauss, Vol. 4: 3848–3891.
- Wilson ME, Fisher J, Chikazawa K, Yoda R., Legendre A., Mook D. Gould KG. 2003. "Leptin administration increases nocturnal concentrations of luteinizing hormone and growth hormone in juvenile female rhesus monkeys." *J Clin Endocrinol Metab*. 88(10):4874–4883.
- World Bank. 2001. "Nicaragua Poverty Assessment: Challenges and Opportunities for Poverty Reduction," Report No. 20488-NI, The World Bank, Washington, D.C.
- World Bank. 2003. "Nicaragua Poverty assessment: Raising Welfare and Reducing Vulnerability," Report No. 26128-NI. Washington DC: The World Bank.