IDB
Inter-American
Development Bank

Long-Term Impacts of Conditional Cash Transfers in Latin America

Review of the Evidence

Teresa Molina-Millan
Tania Barham
Karen Macours
John A. Maluccio
Marco Stampini

Social Protection and
Health Division

TECHNICAL
NOTE Nº
IDB-TN-923

January 2016

Long-Term Impacts of Conditional Cash Transfers in Latin America

Review of the Evidence

Teresa Molina-Millan
Tania Barham
Karen Macours
John A. Maluccio
Marco Stampini

January 2016

IDB
Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library
Long-term impacts of conditional cash transfers in Latin America: review of the evidence / Teresa Molina-Millan, Tania Barham, Karen Macours, John A. Maluccio, Marco Stampini.
p. cm. — (IDB Technical Note ; 923)
Includes bibliographical references.
1. Transfer payments—Latin America. 2. Transfer payments—Caribbean Area. 3. Income maintenance programs—Latin America. 4. Income maintenance programs—Caribbean Area. I. Molina-Millan, Teresa. II. Barham, Tania. III. Macours, Karen. IV. Maluccio, John A. V. Stampini, Marco. VI. Inter-American Development Bank. Social Protection and Health Division. VII. Series.
IDB-TN-923

scl-sph@iadb.org

http://www.iadb.org/SocialProtection

Long-Term Impacts of Conditional Cash Transfers in Latin America: Review of the Evidence[1]

Teresa Molina-Millan, Tania Barham, Karen Macours, John A. Maluccio, Marco Stampini

Abstract

We critically review the literature on the long-term impacts of Conditional Cash Transfer (CCT) programs in Latin America. We define long-term impacts as those that both: 1) are related to the accumulation of human capital, and; 2) are observed after beneficiary children have reached a later stage of the life-cycle. We focus on two transitions. The first is that of children exposed to CCTs in utero or early childhood (under age 6), who have then transitioned to school ages. The second is that of children exposed to CCTs during school-age, who have then transitioned to early adulthood. Evidence is largely mixed. The experimental literature finds consistent positive effects on schooling, as well as some positive impacts on cognitive skills and learning, socioemotional skills and off-farm employment and income. However, many other estimates are not statistically different from zero and it is often not possible to discern whether this is due to lack of impact or to methodological shortcomings. Non-experimental evidence also is mixed. Developing further opportunities for research with a clean identification strategy for the measurement of long-term impacts should be high on the agenda. As original beneficiaries continue to age, this should also be increasingly possible.

JEL code: I38.

Keywords: Conditional Cash Transfers (CCTs), long-term impacts, Latin America, Progresa, Oportunidades, Prospera, Familias en Acción, Red de Protección Social.

[1] Teresa Molina-Millan is Assistant Researcher at Nova School of Business and Economics (teresamolin@gmail.com); Tania Barham is Associate Professor of Economics and Faculty at the Institute of Behavioral Science, University of Colorado (tania.barham@colorado.edu); Karen Macours is Associate Professor at the Paris School of Economics and INRA Researcher (karen.macours@psemail.edu); John A. Maluccio is Associate Professor of Economics at Middlebury College (maluccio@middlebury.edu); and Marco Stampini is Social Protection Lead Specialist in the Social Protection and Health division of the Inter-American Development Bank (IDB) (mstampini@iadb.org). This report has been prepared with funds from the IDB Economic and Sector Work "CCT Long-Term Impacts: Literature Review and Research Opportunities" (RG-K1421). We thank Ferdinando Regalia, Pablo Ibarrarán, Caridad Araujo, Nadin Medellin, Pedro Cueva and the participants in an IDB seminar for useful comments and suggestions. All remaining errors are our own. The content and findings of this paper reflect the opinions of the authors and not those of the IDB, its Board of Directors, or the countries they represent.

1. Introduction

Conditional Cash Transfer (CCT) programs, started in the late 1990s in Latin America, have become the anti-poverty program of choice in many developing countries in the region and elsewhere (Fiszbein and Schady 2009; Stampini and Tornarolli 2012, Paes-Sousa et al. 2013). Their objectives, including short-term poverty reduction via transfers targeted to the poor, and long-term poverty reduction through enhanced investment in human capital, have widespread policy appeal. The most common designs for CCTs broadly follow the original design of *Progresa*, the Mexican program begun in 1997. Targeted to the poor, the principal components include regular cash transfers to women (conditional on regular visits to health providers for younger children and school enrollment and regular attendance for school-age children) and social marketing to encourage investment in nutrition, health and education.

Numerous CCT evaluations, many based on rigorous experimental designs, consistently show short-term impacts. These include poverty alleviation, improved health outcomes and increased school attainment (Annex 1).[2] In contrast, only a few studies focus on whether short-term gains eventually translate into sustained *long-term* benefits. For example, does exposure to CCTs in early childhood lead to improvements in school-age outcomes? Do the increased investments in human capital improve labor market or other (young) adult outcomes? And, perhaps most importantly, do CCTs ultimately improve the welfare of the next generation? After nearly two decades of experience with these programs, there is a growing need, as well as an increasing potential, for establishing whether such long-term gains have been realized.

In this note, we critically review the existing evidence on whether, and to what extent, CCT programs have achieved their long-term objectives. To the best of our knowledge, there is no research that has examined whether individuals who benefitted from the interventions as children or teenagers remained poor as adults, or that has analyzed directly the welfare effects on the next generation.[3] This is in large part because it is still too soon to investigate such outcomes, as the earliest programs only began in the late 1990s.[4] Consequently, the bulk of the

[2] See also McEwan (2013) and Murnane and Ganimian (2014) for reviews that consider CCTs among other programs.

[3] There is some research examining longer-run poverty dynamics for CCT beneficiary households. For example, Gertler, Martinez and Rubio-Codina (2012) find that original beneficiary households in Mexico's *Progresa* made investments that led to continued consumption gains more than five years after the start of the program.

[4] Chile's *Subsidio Unitario Familiar*, implemented since 1981, is sometimes considered the first CCT. However, this program did not include penalties for noncompliance with program conditions, nor did it have an ex-ante rigorous impact evaluation design (Fiszbein and Schady 2009).

research we review focuses on whether CCTs have led to a sustainable accumulation of human capital. This is a key and arguably necessary component of CCT long-term objectives, especially to the extent that some elements of human capital are enduring, such as completed grades of schooling.

Rather than relying exclusively on time since first exposure, we define long-term impacts as those that both: 1) are related to the accumulation of human capital, and; 2) are observed after beneficiary children have reached a later stage of the life-cycle. We focus on two transitions. The first is that of children exposed to CCTs in utero or early childhood (under age 6), who have then transitioned to school ages. The second is that of children exposed to CCTs during school-age, who have then transitioned to early adulthood, where we take age 18 as an approximate cut-off for adulthood. The duration of CCT exposure or the length of time since CCT exposure were not explicit criteria for selecting studies for this review. However, they are also important factors for assessing impacts, and we document how they vary across studies. We restrict the review to research on Latin American programs, which have the longest history and consequently offer more opportunities for analyzing long-term impacts. Moreover, these programs share many key design features, which makes comparison more meaningful.

The bulk of our review focuses on three countries for which both experimental and non-experimental evidence on long-term effects is available: Mexico, Nicaragua and Colombia. Mexico's *Progresa*[5] and Nicaragua's *Red de Protección Social* (*RPS*) embedded experimental impact evaluations in their rollout, permitting experimental assessment of the impacts resulting from differential program exposure. In both cases, there was also survey work done on similar excluded populations permitting assessment of absolute program effects using non-experimental (matching) techniques. Colombia's *Familias en Acción* had no experimental evaluation. We review a long-term non-experimental evaluation of the overall program, and non-experimental research that relies on secondary data on learning and tertiary-level educational outcomes. Experimental evidence from Colombia comes from a randomized pilot program in urban Bogotá.

The remainder of the paper is organized as follows. In Section 2, we review the research that examines how participation in CCT programs for school-aged individuals (under age 18

[5] While over time Mexico's CCT program has been renamed *Oportunidades* and then *Prospera*, for simplicity and because we are describing long-term impacts related to the initial stages of implementation, we refer to it throughout the paper as *Progresa*.

years) translates in impacts on learning, labor market outcomes and marriage markets. Given the timeframe of the existing studies, evidence is necessarily limited to impacts during early adulthood, and hence reflects at least in part the trade-off between schooling and early work experience.[6] In Section 3, we review the research that examines how exposure in utero and early childhood translates into better cognitive, socioemotional and health outcomes during school age. Section 4 concludes, arguing the case for further research. An electronic Annex contains a comprehensive list of the articles reviewed in this paper, and summarizes findings on outcomes beyond those discussed in the main text of the paper.

2. Exposure to a CCT during School Ages and Outcomes in Early Adulthood

Virtually every CCT program with a rigorous evaluation design has been shown to have positive short-term impacts on school enrollment and attendance for children subject to schooling-related conditionalities, though the magnitudes of those impacts vary with program characteristics and type of target population (Fiszbein and Schady 2009).[7] In many cases, positive results are also found for school progression. While these short-term impacts are encouraging, they fall short of providing evidence on the more lasting changes that are the ultimate objective of CCTs. Does increased schooling in the short run lead to increased final educational attainment? And, perhaps more importantly, does increased schooling lead to better learning and improved labor market outcomes? A small number of studies, based on programs in Mexico, Nicaragua and Colombia, attempt to answer these questions.[8]

For each country, we first describe the experimental evidence, as this typically presents fewer concerns regarding internal validity. We then review the non-experimental evidence,

[6] If CCTs increase the years of schooling, their beneficiaries will likely have less work experience than an otherwise similar cohort without program exposure. This may reduce the net returns from CCTs measured during early adulthood, particularly if the returns to work experience are diminishing (i.e., if the returns to the first few years of work experience, that only those who have not continued to study have, is particularly high).

[7] Fiszbein and Schady (2009) review evidence from: Chile (*Chile Solidario*), Colombia (*Familias en Acción*), Ecuador (*Bono de Desarrollo Humano*), Honduras (*Programa de Asignación Familiar-II*), Jamaica (Programme of Advancement through Health and Education), Mexico (*Progresa*), Nicaragua (*Atención a Crisis* and *Red de Protección Social*), Bangladesh (Female Secondary School Assistance Program), Cambodia (Japan Fund for Poverty Reduction and Cambodia Education Sector Support Project), Pakistan (Punjab Education Sector Reform Program) and Turkey (Social Risk Mitigation Project). See also McEwan (2013) and Murnane and Ganimian (2014).

[8] Beyond these three countries, some evidence is also available for Honduras. Rackstraw (2014) provides a long-term analysis of the effects of *Programa de Asignaciones Familiares-PRAF-II*, which began in 2000. The author compiles secondary labor force survey data through 2012. She uses information on the municipality of birth to assign individuals' intent to treat status, based on *PRAF-II* municipality-level randomization. She then examines impacts for different age groups, including those exposed as young children (0-2 years) and those exposed at school ages. The paper provides an example of the possibility of using the original *PRAF-II* evaluation design, combined with secondary sources, to measure long-term impacts.

whose validity relies on much stronger assumptions. We end the discussion of each country case with a critical assessment of the strengths and weaknesses of the existing evidence.

2.1 Mexico: Experimental Evidence

Much of the experimental evidence on CCT long-term impacts for those exposed as school-age children comes from Mexico's *Progresa*. Behrman, Parker and Todd (2009a, 2011) examine whether differential exposure to the program significantly impacted schooling and learning, labor market outcomes, migration and marriage. They use the 1997 Survey of Household Socio-Economic Conditions (ENCASEH) along with Mexico's rural households evaluation survey panel (ENCEL) through 2003. Their sample includes individuals between 9 and 15 years of age at the start of the program, covering the typical ages when students in Mexico transition from primary to secondary education.[9] Previous research revealed that this cohort had experienced the largest gains in schooling in the short term (Behrman, Sengupta and Todd 2005). To assess long-term differential effects, Behrman, Parker and Todd (2009a, 2011) examine this same cohort in 2003 (i.e., at ages 15-21) and exploit the initial experimental evaluation design, in which 506 eligible rural communities were randomly assigned to treatment (320) and control (186) groups. Eligible households in original treatment communities started receiving cash transfers in 1998, while those in the original control communities started receiving them approximately 18 months later. This difference in the length of exposure, between randomly allocated "early-" and "late-treatment" groups is the cornerstone of their long-term experimental evaluation, and provides differential, rather than absolute, estimated effects.

In 2003, approximately 6 years after the program started in the early-treatment group, an 18-month difference in the length of exposure led to a significant difference of 0.2 grades completed for both males and females in this cohort; i.e., grades completed were 0.2 higher in the early- versus the late-treatment groups. Larger differential effects, of the order of 0.5 grades, were observed for those who were entering their last year of primary school around the time of first exposure to the program. The study next assessed whether the increase in grades attained translated into more learning, by examining the impact of the differential exposure on three achievement tests covering reading, writing and mathematics skills. These tests were administered to all respondents in their homes, regardless of completed grades of school or

[9] Although not all individuals 9-15 cross our age 18 "threshold" for early adulthood during the evaluation period considered, some reported results are not split more finely by age, so we describe results for all ages in the broader cohort they examine.

enrollment status. The study finds no significant differential impacts on any of the achievement tests, including from analyses exploring possible heterogeneous effects by age, gender or baseline schooling levels. These findings suggest that even though schooling differed on average between the groups, there was no corresponding differential impact of the Mexican CCT on learning (Behrman, Parker and Todd 2009a).

Examination of labor market outcomes found that longer exposure to the Mexican CCT significantly decreased male, but not female, labor force participation in the studied cohort. Male labor force participation declined by 2.7 percentage points (approximately 4 percent) among 15-21 year-olds. The result for men is consistent with their increase in completed grades and consequent delayed entry into the labor market. The lack of effect for women could reflect their much lower labor force participation. For example, in treatment communities in 2000, 65 percent of men worked whereas only 26 percent of women did (Behrman, Parker and Todd 2009a, 2011).

Behrman, Parker and Todd (2009a, 2011) also find a negative impact of the differential program exposure on male migration of 2 percentage points (approximately 6 percent), but no significant effect for women. Male migration in this context is generally associated with labor market decisions, hence the labor force participation and migration results appear to be consistent with one another.

Finally, Behrman, Parker and Todd (2009a) found no statistically significant differential impact on the probability of either gender being married in 2003.

2.2 Mexico: Non-Experimental Evidence

Behrman, Parker and Todd (2011) also present non-experimental evidence. This is based on the comparison of individuals aged 9-15 at the start of the program living in households from the original evaluation communities with the same age cohort living in households from other rural communities that were neither part of the original experimental design nor incorporated into the program before 2003. As this comparison introduces selection concerns, the authors employ matching methods to take into account differences in observed characteristics between samples. Individuals were matched on age, gender and a household propensity score based on both household- and community-level characteristics, with the latter being drawn from the 2000

Mexican National Census. Conditional on the assumption that the matching on observable characteristics also eliminates the selection bias related to unobservable characteristics,[10] this approach allows the authors to estimate the *absolute* impacts of approximately four and six years of program exposure (when the original control group and the original treatment group, respectively, are compared with the non-experimental comparison group). These estimates, therefore, reflect different underlying parameters than the differential effects estimated using the experimental variation.

The study estimates absolute impacts of between 0.5 and 1.0 additional grades completed for all but the oldest women (those aged 19-21 in 2003). Impacts for men are modestly larger than for women across all ages in the cohort, notable in part because program transfer sizes, by design, were larger for women. Last, program effects increase with the length of exposure (six years versus four years) to the program (Behrman, Parker and Todd 2011).

For young men aged 15-16 in 2003, there was a large reduction (14 percentage points or approximately 30 percent) in labor force participation after 6 years, consistent with their being more likely to be attending school. However, no significant effect was found for other age groups. Examining different types of work, the study found a large decrease (9 percentage points or approximately 25 percent) in participation in agricultural work for the older (19-21) men, suggesting a shift away from agricultural work. For women, in contrast, there was a large (6.4 percentage points or approximately 20 percent) and significant increase in the proportion working among the oldest (19-21) group, the very same group who did not experience an increase in schooling.

Rodriguez-Oreggia and Freije (2012) use a different non-experimental approach and more recent data from a subsequent wave of Mexico's rural households evaluation survey panel (ENCEL 2007). Notably, however, after ten years the sample is characterized by an attrition rate of 60 percent. The authors explore whether differences in the length of exposure are correlated with different labor market outcomes among beneficiaries who were 5-15 years old in 1998 and are therefore 14-24 years old in 2007. They categorize beneficiaries into three groups, based on the length of potential exposure to the treatment measured as the time since the program had been incorporated into one's community and roughly corresponding to dates when there was a

[10] This is strictly true only if observable and unobservable characteristics are perfectly correlated. The evaluation design is also valid if program impacts are uncorrelated with unobservable characteristics.

significant expansion of the program. They further split the sample into beneficiaries and non-beneficiaries where the latter is the group that was eligible for treatment but for some unknown reason did not receive the program. The validity of the results is conditional on this selection into beneficiary or non-beneficiary status being as good as random, after controlling for observables. Last, they further exclude all individuals who are still studying. For individuals no longer in school, the study finds no relationship between program exposure and labor force participation, and no consistent relationship with labor earnings.

Parker, Rubalcava and Teruel (2012) use a different panel survey, with much lower attrition, to evaluate the long-term differential impacts of *Progresa* on schooling, work participation, earnings and other job characteristics. They rely on three waves of the Mexican Family Life Survey panel (MXFLS 2002, 2005 and 2009). The identification strategy exploits the gradual geographic rollout of the program. The study focuses on children who were 10-14 years old in 1997, the year before the *Progresa* transfers began in the original treatment areas. It compares groups with long differences in exposure using the program rollout, and compares outcomes of young adults in communities that were selected to receive *Progresa* in the initial years of program operation with those in communities selected to receive the program in 2004 or later. They hence measure impacts for a difference in exposure of up to 7 years, compared with the 18-month differential that Behrman, Parker and Todd (2009a, 2011) exploit in their experimental results. The authors combine propensity score matching and difference-in-difference estimation to correct for selection bias. By 2005, a difference in program exposure of about 7 years significantly increased the probability of attending college by 4-5 percentage points (an increase of approximately 30-40 percent). It also significantly increased the probability of working. In contrast, no significant effects were found for hours worked or hourly labor earnings, which might reflect the fact that although early beneficiaries have more years of education they also likely have less experience in the labor market than late beneficiaries. A second reason for lack of results on hours worked and hourly labor earnings is that early beneficiaries may be more likely to be in school still. Lastly, as the authors point out, the lack of findings might also indicate that the returns to program-induced increases in schooling are not high enough to lead to changes in earnings for young adults.[11]

[11] We do not describe the 2009 results as the authors make clear that they are difficult to interpret because the surveyed migrants had not been incorporated in the analysis sample yet. These may be the individuals with the highest returns to education and consequently strongest program impact on earnings.

2.3 Mexico: Assessment of the Evidence[12]

One important caveat in interpreting both experimental and non-experimental evaluations that rely on the ongoing impact evaluation panel surveys is the high rate of sample attrition, mostly due to the fact that in Mexico migrants were not followed in these surveys. Forty percent of individuals between 15 and 21 years old interviewed in 1997 were not found in 2003, and only 64 percent of those interviewed in the 2003 survey were found in 2007. Consequently, for example, the 2007 sample used by Rodriguez-Oreggia and Freije (2012) comprises less than 40 percent of the original. This sample is highly selected; moreover, the rate of attrition is different for the initial treatment and control groups. Selectivity poses a substantial threat to the internal and external validity of the estimates based on the 2003 and 2007 ENCEL data.

To correct for the potential selection bias using the 2003 survey, Behrman, Parker and Todd (2009a, 2009b, 2011) follow two strategies. First, they rely on proxy information reported by the household of origin for individual migrants. Second, they apply a density re-weighting method to correct for sample selection.

For outcomes for which proxy information is available from other household members (e.g., reported grades completed), attrition is reduced to around 20 percent in the 2003 sample. Such proxy reports are unavailable when no one in the original household is found, however, and therefore do not correct for household-level attrition, e.g., when all household members have migrated and there are no longer any household members living in the community of origin. In addition, they are unavailable for important measures that require direct person-to-person interview of the respondent, such as the administration of achievement tests. Finally, they rely on the assumption that current household members accurately report education and labor market outcomes of former household members (those who were in the original sample and left), or that any potential misreporting is random or at least balanced across treatment groups.

[12] In this subsection, we critically assess the evidence described earlier for Mexico. Many of the concerns outlined here are pertinent to some of the studies on other countries reviewed later in the paper. In several instances, we describe them in detail here and make shorter reference to them in the parallel assessments of the evidence from Nicaragua and Colombia, though not because they are necessarily less important there (but instead for brevity).

The density re-weighting method weights the sample of individuals interviewed in 2003 in order to replicate the baseline distribution of household and individual characteristics. The key assumption behind this methodology is that attrition is random conditional on observables within each treatment group. The validity of the estimates will hence be jeopardized if there is important selection based on unobservables that differs between treatment and control groups. Molina-Millan and Macours (2015) show that this assumption might not be valid, for example, for attrition associated with work related migration among young adults.

A number of additional concerns can be raised for the non-experimental results in Behrman, Parker and Todd (2011). Pre-intervention baseline data (from 1997) for the new comparison group was not available and was collected retrospectively in 2003. This might have introduced measurement error due to recall bias. Additionally, *Progresa* was initially targeted geographically to marginal areas; consequently, non-*Progresa* communities in the ENCEL survey may have been *a priori* different, in which case balancing at the community level may be difficult. This is particularly relevant as matching cannot be combined with a difference-in-difference strategy for most of the outcomes of interest. Indeed while a difference-in-difference would normally eliminate time-invariant unobservable characteristics, baseline values of labor market outcomes or marriage have no meaningful variation due to the ages being studied (e.g., virtually all are unmarried at baseline), and hence considering the differences does not cancel out the relevant fixed effects. In addition to these concerns for the internal validity of the results, the findings of the above evaluations are not generalizable to the complete population covered by the program. They hold only for non-migrating beneficiaries (except perhaps for outcomes where there is proxy information on migrants). As returns to education might to a certain extent be realized through migration, estimates using only the non-migrating beneficiaries might provide a very incomplete picture.

The results in Rodriguez-Oreggia and Freije (2012) on the other hand, basically provide a comparison between those selected into the program at various time points during its rollout, and those that were apparently eligible but for some unknown reason were not selected for (or chose not to participate in) the program. It is unlikely that the latter group would be comparable to the former, even after controlling for observables. The authors' decision to include only individuals who are no longer studying introduces additional non-random selection in the sample, as it excludes those that might have benefitted the most from the program and are therefore still studying. In other words, it focuses on youth for whom the impact of the program could be lower than average, only including individuals who have dropped out or who did not

pursue tertiary education. Finally, the sample is strongly selected because of the high rates of attrition and incomplete information. Overall, the internal and external validity of this study appear weak.

Finally, for all of the experimental and non-experimental evidence based on the ENCEL data, limited external validity is also a source of concern. This is due to the fact that the sample for the original experimental evaluation was drawn from the poorest rural communities (with supply of health and education services), while now many of the program beneficiaries live in less poor rural areas and in urban areas.

The MXFLS data used by Parker, Rubalcava and Teruel (2012) have lower levels of attrition than the ENCEL due to intensive migrant tracking. In addition, the long differential in exposure increases the power of the evaluation. However, these positive features are at least partially offset by the much smaller sample size (approximately 700 observations) as compared to the large ENCEL surveys. And, while the MXFLS is nationally representative, the external validity of the study is limited as the analysis is restricted to non-indigenous youth in rural communities with overall low levels of community poverty. This choice is made to increase internal validity as it allows identification of more comparable groups for 1998 and 2004. Another potential concern regarding this study is the reliance on recall data for 1997 to obtain difference-in-difference estimates, which possibly introduces recall bias. For example, the authors rely on recall data on labor force participation of cohorts aged 22-26 in 1997.

A final consideration for the whole set of studies regards the difficulty in studying young adults who are in important life cycle transitions. For example, the observed delayed entry in the labor market for males implies that for this age group the evaluation can reveal only initial (or partial) information about the ultimate program effects on occupation or income. The obvious solution, of course outside the scope of these studies, is to continue following these individuals further into the future until all of them have left school and entered the labor market, while making sure to keep attrition as low as possible.

2.4 Nicaragua: Experimental Evidence

Modeled after *Progresa*, *RPS* started in 2000, and had a short-term experimental evaluation built into its initial stages. This evaluation took place in 42 localities in six rural municipalities with initial poverty rates around 80 percent. Unlike Mexico's *Progresa* and Colombia's *Familias en Acción*, *RPS* was not a program with national coverage. Localities were randomly assigned

to early- and late-treatment groups in a public lottery. Early-treatment localities became eligible for transfers in November 2000 and were eligible for three years, receiving their last transfers in late 2003. Late-treatment localities were phased in at the beginning of 2003 and were also eligible for three years of transfers. Households in the early-treatment group did not receive any transfers after 2003, and had no conditionalities after that date. However, they continued to be eligible to use the *RPS*-provided health supply services. The program ended in late 2005. This difference in the timing of exposure, between randomly allocated "early-" and "late-treatment" groups is the cornerstone of the long-term experimental evaluation in Nicaragua, and provides differential, rather than absolute, estimated effects. In contrast to Mexico, it compares groups that randomly received the program for a fixed 3-year period at different points in time. In Mexico, the relative difference in length of exposure diminishes with time since exposure in both groups continues to increase (as the program is ongoing). This is not the case in Nicaragua.

Between November 2009 and November 2011, i.e., approximately 10 years after the start of the program for the early-treatment group, a long-term follow-up survey was conducted (Barham, Macours and Maluccio, 2013a, 2013b). All original households were tracked. When a household was found (and interviewed), but a former member 12 or under in 2000 was no longer present, the individual was tracked and his or her (new) household added to the sample. Relative to other long-term evaluations, attrition was low in Nicaragua. Households and individuals in the original sample were tracked beyond their original communities throughout Nicaragua and in Costa Rica, leading to a household attrition rate below 8 percent. Attrition was higher for individual-level outcomes, especially among the more mobile young adults who were not always possible to interview in person. For example, for males 9-12 years old in 2000, attrition in 2010 was 6 percent for schooling and occupation information (which could be collected by proxy), 12 percent for earnings data, and 19 percent for test scores.

The follow-up survey, together with pre-intervention data, is used to estimate the differential impacts of *RPS* on educational attainment, learning and labor market outcomes for males who were 9-12 years old in 2000 (and therefore 18-21 at follow-up). Due to the random difference in the timing of the interventions for the early- and late-treatment groups, focusing on this specific age cohort allows estimating the long-term effects of benefiting from a CCT in a period of the life cycle that is considered critical for educational investments (the age at which the probability of dropping out of school is high) versus three years later (when dropout is more likely to have already taken place).

Barham, Macours and Maluccio (2016) show that the short-term program effect of a half grade increase in completed schooling for males was largely sustained five years after the end of the program and into early adulthood. In contrast to the findings on the long-term evaluation of *Progresa*, the differential increase in grades attained was accompanied by gains in learning. Males in the early-treatment group experienced an average improvement of about 0.2 standard deviations on standardized tests in mathematics and Spanish (relative to the late-treatment group). This finding is consistent across the range of achievement tests. The study also finds positive differential impacts on socioemotional outcomes, such as optimism and positive self-evaluation. Last, it finds differential effects on labor market outcomes, with the young men in the early-treatment group being more likely to work off-farm, migrating temporarily to do so. This results in an increase of 10-30 percent in monthly off-farm income. Overall, this study shows that the *RPS* produced large long-term differential impacts on earnings for men, consistent with increased human capital leading to better labor market outcomes.

Even with substantial resources dedicated to tracking respondents in this ten-year panel, keeping attrition to a minimum, the remaining attrition is unlikely to be random; indeed Molina-Millan and Macours (2015) demonstrate that there is remaining attrition bias in the standard intent-to-treat estimates, and suggest a correction that uses information from the intensive tracking carried out during the field work to re-weight and correct for sample selection. Substantive results are unchanged after these corrections are made.

2.5 Nicaragua: Non-Experimental Evidence

Barham, Macours and Maluccio (2016) also explore the absolute effects of exposure to three years of *RPS* on outcomes measured 10 years after the start of the program in the early treatment group. The non-experimental control group is drawn from a sample of individuals living in 21 localities in adjacent municipalities, which were selected using the same marginality index used for selecting localities in the experimental evaluation. This sample was first surveyed in 2002, i.e., two years after the start of the program, as part of the short-term program evaluation.

Individuals from the early- and, separately, late-treatment groups are matched to individuals in the described non-experimental control sample, based on a set of individual, household, and community characteristics measured in 2000 and 2002. Then, outcomes in 2010 are compared to measure program impacts. The key assumption underlying this strategy

is that, given the selection of similar and neighboring localities, the matching on observables also controls for all other differences in unobservables.

The non-experimental results show positive absolute impacts on schooling, learning, labor market outcomes and earnings for young men 9-12 in 2000, in line with the experimental results but larger in magnitude. For women in the same age group there were no significant experimental differential effects on grades attained, the standardized tests, or income. The non-experimental results, however, show absolute learning effects for these young women that are similar in magnitude across the early- and late-treatment groups. This suggests that positive and equal absolute impacts on learning may underlie the lack of significant experimental results (which capture the differential effect) for females. This is further supported by the finding that there was a large experimental differential impact on grades attained for an older cohort of girls, aged 13-14 in 2000. The program benefited these older girls in the early treatment group and therefore also may have benefited the younger girls in the late treatment group, as they received it when they were 12-14.

2.6 Nicaragua: Assessment of the Evidence

Despite attrition rates that are much lower than in other long-term studies, attrition bias remains a source of concern. The validity of the estimates still depends on the quality of data provided by proxy informants, or on the validity of assumptions made to re-weight the data to correct for sample selection. While estimates are robust to alternative assumptions, suggesting that they are reliable, determining the direction of any remaining bias is difficult.

The fact that *RPS* targeted areas in which pre-treatment levels of schooling were very low implies that results may have limited external validity for many other settings in Latin America. In other words, it is unclear whether the same results on learning and earnings could be obtained in other countries in the region.

As in the studies on Mexico, beneficiaries were only observed as young adults, when some were still studying (despite the average lower levels of education) and many were still living with their parents. Hence, only additional rounds of data collection will be able to gauge the full long-term impact of the *RPS* in terms of returns to human capital and, more broadly, welfare outcomes. Moreover, the experimental results on learning and labor markets are only significant for the boys, and while results for an older cohort of girls indicate there is a large

differential impact on their grades attained, achievement tests were not administered to them so their learning results could not be examined.

For the non-experimental results, the same caveats apply as for the non-experimental matching results for studies on Mexico, although the details vary.

2.7 Colombia: Experimental Evidence

Barrera-Osorio, Linden and Saavedra (2015) provide experimental evidence on the long-term impacts of a one-year CCT pilot, *Subsidios*, implemented in 2005 in Bogotá.[13] The program consisted of a conditional education transfer for secondary school students (aged 12-18), and included ages at which the probability of dropout is high in Colombia.

The study tests three alternative program designs. The first is referred to as the “basic treatment”, in which transfers were made bi-monthly, conditional on meeting a specified attendance target. The second is a "savings treatment", in which one third of the transfer payments was delayed until enrollment in the following school year. The third is a "tertiary treatment", implemented only for students in upper secondary school (grades 9-11), in which one third of the transfer payments was delayed until after graduation from secondary school and: 1) upon enrollment in a tertiary institution if the individual enrolled in one, or; 2) one year later, if the individual did not enroll. All those in the tertiary treatment who graduated from secondary school (including if this happened in the years following the end of the pilot) became eligible for this final transfer.

The basic and savings treatments are compared relative to one randomized control group. Separately, the tertiary treatment is compared to a different randomized control group (comprising only students in upper secondary school at baseline).[14] The three treatment arms were randomly assigned at the individual level. The control groups never received the intervention, allowing the authors to evaluate the long-term absolute effect of this program.

[13] This is the only CCT program included in our review that operated exclusively in urban areas. It is described further in Barrera et al. (2011).

[14] The savings treatment implied a reduction of the monthly transfer amount, which could lead to children attending school less if families faced short-term liquidity constraints. In contrast, if long-term savings constraints were more significant than immediate liquidity constraints, the savings treatment could generate higher re-enrollment rates than the basic treatment, as the families received the money when the next year’s schooling expenses began, without affecting current attendance rates. The tertiary treatment generated incentives for graduation, which could result in higher re-enrollment and graduation rates and higher levels of enrollment in tertiary institutions, unless yearly saving constraints prior to reaching tertiary levels were the most binding constraint. Attendance under the tertiary treatment also could increase if participants viewed school attendance as instrumental for graduation.

The authors merge program participation data with national administrative records on upper secondary school graduation exams and enrollment in tertiary institutions for the 8 years following the 2005 intervention. These merges were based on student ID number, name and date of birth. The study shows that the percentage that could be merged corresponds to expected rates of graduation from upper secondary school and enrollment in tertiary institutions for the study population, and that the probability of a successful merge was not related to baseline characteristics, nor did it differ between basic treatment and control groups.

Results for students in grades 9 to 11 (i.e., in upper secondary education) at the time of program exposure show that only the savings treatment significantly increased the probability of taking the exit exam. However, the authors cannot reject the hypothesis that the different experimental treatments had equal impacts. Both the savings and tertiary treatments led to higher enrollment in tertiary institutions (by 21 and 35 percent, respectively) suggesting that savings constraints may have been a barrier for enrollment in tertiary education. Furthermore, the savings treatment encouraged enrollment in universities, while the tertiary treatment led to enrollment in lower quality tertiary institutions. This suggests that the incentives for tertiary enrollment encouraged students to enroll more indiscriminately. No significant treatment effects were found for students in lower secondary education.

These results broadly confirm the short-term findings one year after the intervention, based on a combination of administrative and survey data, for older children making transition out of upper secondary school (Barrera-Osorio et al. 2011). Using self-reported information, the authors found a significant increase in the probability of enrolling in a tertiary institution among those assigned to the savings or tertiary treatments in grade 11, but no significant effects on graduation rates for any of the treatment arms. The program had a significant negative impact on labor participation for those in the tertiary treatment, consistent with the large positive effect on continued schooling one year after the intervention found for the group of students in grade 11.[15]

[15] While the savings treatment increased tertiary enrollment by 9.4 percentage points, the tertiary treatment increased it by 48.9 percentage points.

2.8 Colombia: Non-Experimental Evidence

Colombia's national CCT program, *Familias en Acción*, did not have an experimental impact evaluation but was the subject of two long-term non-experimental evaluations.

Baez and Camacho (2011) investigate the effects of up to nine years of participation in *Familias en Acción* using household survey data, registration records from SISBEN[16] and administrative data on the results of the secondary school graduation test, the same data source used by Barrera-Osorio, Linden and Saavedra (2015). The study uses both difference-in-difference matching and Regression Discontinuity Design (RDD). It focuses on two different samples of children who had the potential to complete grade 11 by 2009, constructed from the 2002 baseline program evaluation sample (for the matching analysis) and from merging the program administrative data with the SISBEN census (for the RDD). These two samples were merged with the national secondary school graduation test scores based on person identifiers. The results show that beneficiary children were 4-8 percentage points more likely than non-beneficiary children to complete secondary school. However, the authors found no evidence of differential exit test score performance, conditional on completion.

García et al. (2012) conduct a non-experimental evaluation of long-term effects on education and labor market outcomes. The identification strategy relies on a simple-difference (with baseline controls) or difference-in-difference estimation and compares children from eligible households from municipalities covered by the program in 2002 with children from potentially eligible households from comparable areas that were only targeted in 2007. These control municipalities were selected through propensity score matching. The authors use this methodology to investigate differential exposure effects, using data from 2002 (i.e. pre-intervention), and a follow-up survey conducted between November 2011 and February 2012. They measure these differential impacts on a set of education and labor market outcomes for which the impact of the program can be thought of as cumulative.

When looking at young adults aged 18-26 in 2012, the study finds that the 2 to 5 years additional exposure to *Familias en Acción* increased school attainment by 0.6 grades in rural areas. Also for rural areas, it finds a positive and significant impact on the probability of graduating from upper secondary school, and a somewhat puzzling negative impact on the

[16] The SISBEN is Colombia's proxy means score. Eligibility for *Familias en Acción* is based on having a SISBEN score below a given threshold.

probability of enrolling in tertiary education. The only significant impact on labor market outcomes found by the study is an increase of 2.5 percentage points in the probability of formal employment among women in rural areas. In urban areas, impact estimates for this age group are not significantly different from zero.

The study only reports estimates of impact on cognitive skills, based on a mathematics ability test and the Raven's Progressive Matrices test, for adolescents aged 12-17 in 2012. It shows that the differential exposure to *Familias en Acción* increased mathematics scores by 1.07 standard deviations, which is large compared to other studies on learning outcomes.[17] As this result is for the cohort that was 2-7 years old at baseline, it reflects the effect of both the nutritional and educational components of the CCT program. Given that the program effect on school attainment for this cohort was 0.3 years of schooling, the large estimate on mathematics test scores may be driven by the early childhood nutritional components of the program, described in Section 3.3.

2.9 Colombia: Assessment of the Evidence

Overall, Barrera-Osorio, Linden and Saavedra (2015) provide a good example of the possibilities and the limits of using secondary administrative data to follow up on an earlier experimental evaluation. If data can be merged (which notably requires excellent identifiers at baseline), it is possible to conduct a study with low budget and high internal validity and statistical power. On the other hand, the set of outcomes that can be studied using such secondary data, and hence the possibility to understand the different parts of the impact pathway, are naturally more limited.

The results in Baez and Camacho (2011) have a similar limitation. In addition, the finding on test scores is not generalizable to the complete population covered by *Familias en Acción* as the test was only administered to children who stayed in school and progressed until grade 11, which may introduce sample selection bias. Further selection may be due to the low rate of success of the merge between program administrative records and school test data, ranging between 18 percent (for propensity score matching) and 24 percent (for RDD). Finally, the study cannot distinguish the effect for different ages from the effect due to length of exposure. For example, enrolled beneficiaries joining the program when they are older, have fewer years of

[17] Results for the Raven (+0.16 standard deviations) were not statistically significant.

school to complete than younger beneficiaries, thus they are more likely to be observed finishing secondary school. If this is the case, shorter exposure to the program could be incorrectly associated with higher secondary school completion rates.

An important drawback of the use of the RDD approach in Colombia is that the SISBEN is not only used to determine eligibility for *Familias en Acción*, but also for several other government interventions (Velez et al. 1999). Hence the estimates are potentially confounding the impacts of different programs.

The main drawback of propensity score matching (simple difference or difference-in-difference) estimates in García et al. (2012) is similar to that described for the non-experimental studies for Mexico and Nicaragua. The study relies on the arguably strong assumption that selection into the program is only related to observable and time-invariant unobservable characteristics. As the studied outcomes are typically only meaningfully observed for adults (secondary school completion) or likely should be interpreted differently for adults compared to children (e.g., employment), controlling for baseline outcomes might not help control for unobservable confounders. In addition, the baseline survey used in the evaluation of *Familias en Acción* was implemented after the program had already been announced, and as such might capture some anticipation effects.

3. Exposure to a CCT during Early Childhood and Outcomes during School Age

In this section we review the research that examines whether and how exposure to CCT programs in utero and under age 6 translates into better cognitive, socio-emotional, and health outcomes during primary and secondary school. The evidence base for this cohort is more limited, with notably fewer studies. As above, much of the long-term evidence comes from Mexico, Nicaragua and Colombia.

3.1 Mexico

Fernald, Gertler and Neufield (2009) investigate the effect of *Oportunidades* on child growth, cognition, language, and behaviors, 10 years after the start of the program. They exploit the 18-month differential exposure between the experimental treatment and control localities described above for the school-age cohorts examined by Behrman, Parker and Todd (2009a).

Outcomes are measured for individuals who were in utero or at most 13 months old when the program started, and therefore 8-10 years old at the time of the follow-up survey.

The study finds a significant reduction in behavioral problems. No significant experimental results are found for any measure of child growth, cognition or language when using the basic ITT estimates (comparison of mean outcomes in treatment and control). The authors also present an alternative estimate, however, which takes into account the potential cumulative cash transfers received since household enrollment up to 2007. Cumulative transfers differ across households for two reasons: 1) the experimental variation in timing of entry into the program, and; 2) different household composition and grade achievements of potential eligible children at baseline (with transfers tied to gender, age and grade). Incorporating this latter source of non-experimental variation, the authors find that higher cash transfers are significantly and positively associated with height-for-age z-scores and higher verbal and cognitive test scores. They also find a negative association between accumulated cash transfers and the number of reported behavioral problems, consistent with the findings from the experimental evaluation.

Because the accumulated cash transfers depend on household structure as well as the randomized assignment, the interpretation and internal validity of these results have been called into question (Attanasio, Meghir and Schady 2010). Given the lack of evidence when only using the randomized assignment, the results must be driven by differences in baseline household demographics. Yet baseline household demographics are not randomly assigned, and might well affect cognitive and anthropometric outcomes in their own right.

Behrman, Parker and Todd (2009b) examine children aged 0-8 at the start of the program, and thus 6-14 in 2003, following the same approach as in their experimental studies on older children discussed in Section 2.1. Attrition for these age groups is lower than for the older children, at 23 percent, and similarly addressed through re-weighting. The authors find a slight differential reduction (of 0.05 years) in the age of entry into primary school for girls 7-8 years old in 2003, but no significant effects for the older ages or for boys. They also look at program impacts on grade progression, which reflects both enrollment and continuation. The results show that the 18 months differential exposure to *Progresa* had no significant effect on grade progression for children aged 9-11 in 2003.[18] In contrast, the difference-in-difference

[18] Results are not available for the youngest age group.

matching estimates that compare the original treatment group receiving six years of benefits to the 2003 non-experimental comparison group show positive and significant absolute effects in progression rates of about 15 percent for boys and 7 percent for girls.

Similarly, the differential results do not show robust evidence for grade completion, but the matching difference-in-difference estimates suggest that the program increased completed grades of schooling. While effects for children aged 6-8 are not significant, girls aged 9-11 accumulated about 0.3 grades and boys about 0.4 grades more than non-beneficiary peers. Overall, the findings indicate limited experimental differential effects, but stronger results for the non-experimental absolute effects. The assumptions underlying both approaches are similar to those discussed above, and hence the same caveats for interpretation apply, though not necessarily to the same degree; for example, there is much lower attrition for this age group and therefore possibly less potential attrition bias.

3.2 Nicaragua

Using the randomized rollout of Nicaragua's *RPS*, Barham, Macours and Maluccio (2013a) analyze the impact for boys exposed in utero and during the first two years of life, as compared to boys exposed outside of this potentially critical 1000-day window. A set of seven age-appropriate cognitive tests were administered 10 years after the start of the program, to a cohort of children that was born in the first 6 months of the program. The tests measured processing speed, memory, receptive vocabulary and executive functioning. In addition, height and weight also were measured. As children were tracked to new locations in case of migration, attrition was limited (6 percent). They were tested and measured in their homes, regardless of schooling status. The analysis focused on boys to maximize power, since boys are more vulnerable in early life than girls, particularly during the pre-natal period (Eriksson et al*.* 2010).

Ten years after the start of *RPS*, the differential timing of exposure to the 3-year program resulted in cognitive outcomes that are on average 0.15 standard deviations higher for the early-treatment group.[19] These results are largely consistent across a variety of cognitive outcomes. At the same time, the analysis showed no significant differential impacts on anthropometric measures, which contrasts with the short-term absolute effects found by Maluccio and Flores

[19] The average impact was obtained following Kling, Liebman and Katz (2007) family of outcomes approach.

(2005). Together, the results suggest complete catch-up for boys in the late treatment communities for physical, but not for cognitive, outcomes.

While the experimental results require relatively few assumptions, they still leave one to wonder whether (relative to untreated peers) there might have been persistent absolute impacts for outcomes other than cognition for boys (where the significant differential effects strongly suggest positive absolute effects for the early-treatment group). Indeed, the insignificant differential experimental results on anthropometrics are consistent with several patterns of possible effects over time that result in there having been the same absolute long-term effect in both early- and late-treatment groups, and therefore no differential effect for these outcomes. For example, both treatment groups may have experienced (similarly sized) large improvements, or alternatively, the early-treatment group may have experienced a strong short-term gain that faded out in the long-term such that it was the same as any long-term gain experienced in the late-treatment group. A pattern of positive absolute impacts cancelling each other out in the differential seems likely for the anthropometric outcomes where there is evidence of short-term gains. Finally, parallel analyses on girls in the same cohort are not reported.

3.3 Colombia

García *et al.* (2012) report difference-in-difference non-experimental evidence of the effect of five years of differential exposure to *Familias en Acción* on nutrition and health outcomes for children aged 0-6 at baseline (in 2002). This is complemented by RDD estimates for children aged 3-11 years old at baseline. RDD exploits variation in assignment to treatment arising from the discontinuous rule that determines eligibility to the program. Households that score just below and just above the SISBEN eligibility threshold are statistically comparable except for their potential participation in *Familias en Acción* (or other programs that may use the same threshold).

Using the difference-in-difference approach, the study finds positive and significant impacts of exposure during the first five years of life on anthropometric measures. In particular, the height-for-age z-scores increased by 0.21 standard deviations in rural areas for kids 0 to 2-3 years old in 2002, and by 0.16 standard deviations in rural and urban areas combined. The treated children in this cohort are compared to children who only became eligible when they were 5 to 7-8 years old. This positive effect is accompanied by a reduction in stunting of about 6

percentage points. The authors do not observe improvements in weight-for-age indicators, but do find an increase in the percentage of overweight children of 5.6 percentage points, which they link to poor eating habits (García et al. 2012). As with the non-experimental estimates for the older cohort in Colombia reviewed above, the strong assumptions required for identification form the principal caveat to this research.

The difference-in-difference estimates described above in section 2.8 indicate that there were large impacts on the mathematics ability test for adolescents aged 12-17 in 2012 (i.e., for children aged 2-7 at baseline) and that the nutritional components of the program during childhood may be the underlying driver of those effects. The RDD results for children aged 3-11 in 2012 are consistent with that possibility. They show modest evidence of impacts on cognition around the threshold, with an increase in the TVIP (Picture Peabody Vocabulary Test of receptive vocabulary) score of 0.09 standard deviations for those children. As mentioned earlier, however, a drawback of the RDD approach in Colombia is that the SISBEN is used to determine eligibility for several social programs (Velez et al. 1999), hence the estimates potentially confound the impacts of different programs and do not isolate the impact of *Familias en Acción*.

4. Conclusions

In large part because of their twin objectives (short-term poverty reduction via transfers targeted to the poor and long-term poverty reduction through enhanced investment in human capital) CCTs have widespread policy appeal. Numerous evaluations, many based on rigorous experimental designs, leave little doubt that such programs have been effective in the *short term*. For a variety of reasons, however, the evidence base is much less developed as to whether these short-term gains eventually translate into sustained long-term benefits. Even if it is not yet possible to comprehensively assess all long-term implications (for example, whether they succeed in breaking the intergenerational transmission of poverty), after nearly two decades of experience with CCT programs, evidence on important long-term impacts has begun to accumulate.

In this review, we defined long-term impacts as those that materialize across stages of the life-cycle. We considered two life-cycle transitions. The first is from early childhood to childhood/adolescence; the focus in this case is on health, schooling, cognitive and socio-emotional outcomes of children who benefited from CCTs during early childhood. The second transition is from childhood/adolescence to early adulthood; the focus in this case is on

schooling-related, family and labor market outcomes of young adults who benefited from CCTs during school ages, in particular at ages at which they were at high risk of dropping out of school.

For both transitions, we reviewed and highlighted the strengths and limitations of the available experimental and non-experimental evidence. The literature employing non-experimental methods does not require that CCT programs embedded an experimental impact evaluation in their initial design or rollout. The credibility of such non-experimental results, however, is severely hindered by the difficulties inherent in constructing a valid comparison group, particularly when there are important unobservables that cannot be controlled for but that influence the outcomes of interest. In contrast, the literature based on experimental methods is more likely to yield internally valid results, but is often limited by the fact that few programs were set up for rigorous long-term evaluation of their overall impacts. Most initially randomized control groups subsequently received the program. Consequently, long-term impact evaluations that exploit the experimental design often can only measure differential impacts (e.g., the impacts of longer exposure).

For both the experimental and non-experimental evidence, sample attrition (likely to be related to migration, itself a potential outcome of interest) is an important source of concern. In addition, in most cases the beneficiaries have yet to complete their transition to the labor market. Only one study (in Colombia), was able to follow individuals benefiting during school-ages into their mid-20s. Before this age, many individuals are still transitioning out of school, or have only recently started to work; hence the interpretation of labor market impacts is difficult (due to the tradeoff between additional schooling and shorter work experience).

Overall, we find that the existing evidence on CCTs long-term impacts in Latin America is mixed. The experimental literature provides consistent evidence of impacts on schooling (in Mexico, Nicaragua and Colombia), as well as some evidence of impacts on cognitive skills and learning (in Nicaragua), socioemotional skills (in Mexico) and off-farm employment and income (in Nicaragua). The effects on other outcomes are generally not statistically different from zero, though it is often difficult to discern whether this is due to lack of impact or other methodological concerns (e.g., lack of power or a short difference in exposure between original treatment and control groups). The non-experimental literature provides a similarly mixed picture.

Our interpretation of the existing evidence, therefore, is that “the jury is still out”. There are cases with notable long-term impacts but, with the possible exception of schooling, there is little consistent evidence across outcomes for all programs. Expanding the evidence base with additional long-term studies that convincingly address the highlighted challenges is paramount. This may include exploiting cases in which the modality of rollout, unexpected changes in eligibility criteria (e.g., in the age of eligibility or the specifics of school grades covered), retargeting exercises or other changes in program rules allow a clean identification strategy for the estimation of long-term impacts. Encouragingly, as initial beneficiaries finish the transition to adulthood, greater opportunity to examine the more “permanent” returns to human capital become available. Uncovering such opportunities, and developing strategies to account *ex ante* for selection and identification concerns, is crucial for providing more conclusive evidence on if, how and when CCTs are achieving their long-term objectives.

5. References

Attanasio, O., Meghir, C. and Schady, N. (2010). “Mexico’s Conditional Cash Transfer Programme: Comment” *The Lancet*; 375(9719):980.

Baez, J.E. and Camacho, A. (2011). “Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital: Evidence from Colombia.” IZA *Discussion Paper* No. 5751

Baird, S., McIntosh, C., and Özler, B. (2011). “Cash or Condition? Evidence from a Cash Transfer Experiment.” *The Quarterly Journal of Economics*, *126*(4), 1709–1753. doi:10.1093/qje/qjr032

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

Barham, T., Macours, K., and Maluccio, J.A. (2013b). “More Schooling and More Learning?: Effects of a Three-year Conditional Cash Transfer Program in Nicaragua after 10 years.” *IADB Working Paper Series* IDB-WP-432.

Barham, T., Macours, K., and Maluccio, J.A. (2016). “More Schooling, More Learning, More Earnings: Effects of a Three-Year Conditional Cash Transfer Program in Nicaragua after 10 Years.” Unpublished manuscript.

Barrera-Osorio, F., Bertrand, M., Linden, L.L. and Perez-Calle, F. (2011). “Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia.” *American Economic Journal: Applied Economics* 3(2), 167-95.

Barrera-Osorio, F., Linden, L.L., and Saavedra, J.E. (2015). “Long Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia.” Unpublished manuscript.

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

Behrman, J.R., Parker, S.W. and Todd, P.E. (2009b). “Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico.” *Economic Development and Cultural Change* 57(3): 439–477.

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

Behrman, J. R., Sengupta, P. and Todd, P. (2005). “Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico.” *Economic Development and Cultural Change* 54(1), 237-75.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P. and Pouliquen, V. (2015). “Turning a Shove into a Nudge? A “Labeled Cash Transfer” for Education.” *American Economic Journal: Economic Policy*, 7(3): 86-125.

Edmonds, E.V., and Schady, N. (2012). "Poverty Alleviation and Child Labor." *American Economic Journal: Economic Policy*, 4(4): 100-124.

Eriksson, J.G., Kajantie, E., Osmond, C., Thornburg, K., and Barker, D.J.P. (2010). "Boys live dangerously in the womb." *American Journal of Human Biology* 22, no. 3: 330-335.

Fernald, L.C., Gertler, P.J. and Neufeld, L.M. (2009). “10-year Effect of *Oportunidades*, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: a Longitudinal Follow-up Study.” *The Lancet* 374(9706), 1997-2005.

Fernald, L.C.H. and Hidrobo, M. (2011). “Effect of Ecuador’s Cash Transfer Program (Bono de Desarrollo Humano) on Child Development in Infants and Toddlers: a Randomized Effectiveness Trial.” *Social Science and Medicine (1982)*, *72*(9), 1437–46. doi:10.1016/j.socscimed.2011.03.005

Filmer, D. and Schady, N. (2014). “The Medium-Term Effects of Scholarships in a Low-Income Country.” *Journal of Human Resources*, *49*(3), 663–694. doi:10.1353/jhr.2014.0022

Fiszbein, A. and Schady, N.R. (2009). “Conditional Cash Transfers: Reducing Present and Future Poverty.” *World Bank Policy Research Report*. Washington, DC, United States: 24, World Bank.

Gaarder, M., Glassman, A. and Todd, J. (2010). “Conditional Cash Transfers and Health: Unpacking the Causal Chain.” *Journal of Development Effectiveness*, *2*(1), 6–50.

Galiani, S. and McEwan, P.J. (2013). “The Heterogeneous Impact of Conditional Cash Transfers.” *Journal of Public Economics*, *103*, 85–96. doi:10.1016/j.jpubeco.2013.04.004

García, A., Romero, O.L., Attanasio, O. and Pellerano, L. (2012). “Impactos de Largo Plazo del Programa Familias en Acción en Municipios de Menos de 100 mil Habitantes en los Aspectos Claves del Desarrollo del Capital Humano.” Technical report, Union Temporal Econometria S.A. SEI. con la asesoria del IFS.

Gertler, P., Martinez, S. and Rubio-Codina, M. (2012). "Investing Cash Transfers to Raise Long Term Living Standards." *American Economic Journal: Applied Economics* 4(1): 164–92

IDB (2014). “Conditional Cash Transfers Research Brief.” *Unpublished manuscript.* Washington, DC, United States: Inter-American Development Bank.

Kling, J.R., Liebman, J.B. and Katz, L.F. (2007). "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1): 83-119.

Levy, S. (2006). “Progress Against Poverty: Sustaining Mexico’s Progresa-Oportunidades Program.” Washington DC, United States: *Brookings Institution Press.*

Levy, Santiago, and Schady, N. (2013). “Latin America’s Social Policy Challenge: Education, Social Insurance, Redistribution,” *Journal of Economic Perspectives*, vol. 27(2): 193-218.

Macours, K., Schady, N. and Vakis, R. (2012). “Cash Transfers, Behavioral Changes, and Cognitive Development in Early Childhood: Evidence from a Randomized Experiment.” *American Economic Journal: Applied Economics* (4)2.

Maluccio, J.A., and Flores, R. (2005). “Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan *Red de Protección Social*.” Research Report No. 141, International Food Policy Research Institute.

McEwan, P. (2013). “Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments.” Unpublished manuscript, Wellesley College

Molina-Millan, T. and Macours, K. (2015). “Attrition in Randomized Control Trials: Regular versus Intense Tracking Protocols.” Unpublished manuscript.

Murnane, R.J. and Ganimian, A.J. (2014). “Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Evaluations”, *NBER Working Paper 20284*.

Paes-Sousa, R., Regalia, F. and Stampini, M. (2013). “Conditions for Success in Implementing CCT Programs: Lessons for Asia from Latin America and the Caribbean.” *IDB Policy Brief* n. 185. Washington D.C., USA: Inter-American Development Bank.

Parker, S., Rubalcava, L., Teruel, G. (2012). “Do Conditional Cash Transfer Programs Improve Work and Earnings among its Youth Beneficiaries? Evidence after a Decade of a Mexican Cash Transfer Program.” Unpublished manuscript.

Rackstraw, E. (2014). “A Decade Later: An Evaluation of the Longer-Term Impacts of a Honduran Conditional Cash Transfer.” *Wellesley College Honors Thesis Collection,* Paper 215.

Rasella, D., Aquino, R., Santos, C.A.T., Paes-Sousa, R., and Barreto, M.L. (2013). “Effect of a Conditional Cash Transfer Programme on Childhood Mortality: a Nationwide Analysis of Brazilian Municipalities.” *Lancet*, *382*(9886), 57–64. doi:10.1016/S0140-6736(13)60715-1

Rodriguez-Oreggia, E. and Freije, S. (2012). “Long Term Impact of a Cash Transfers Program on Labor Outcomes of the Rural Youth in Mexico.” *CID Working Paper* No 230. Harvard University.

Ruiz-Arranz, M., Davis, B., Handa, S., Stampini, M. and Winters, P. (2006). “Program Conditionality and Food Security: The impact of PROGRESA and PROCAMPO Transfers in Rural Mexico.” *Economia*, *7*(2), 249–278.

Saavedra, J. and García, S. (2012). “Impacts of Conditional Cash Transfer Programs on Educational Outcomes in Developing Countries.” *RAND Corporation. Working Papers*, (WR-921-1). Retrieved from http://www.rand.org/pubs/working_papers/WR921-1.html

Stampini, M. and Tornarolli, L. (2012). “The Growth of Conditional Cash Transfers in Latin America and the Caribbean: Did They Go Too Far?” *IDB Policy Brief* n. 185. Washington D.C., USA: Inter-American Development Bank.

Velez, C. E., Castano, E., and Deutsch, R. (1999). "An Economic Interpretation of Targeting Systems for Social Programs: the case of Colombia’s SISBEN." World Bank. *Working Papers.*

Annex 1 – Short term impacts of CCTs on the children in beneficiary households

Consumption and Poverty. Various impact evaluations, summarized in Fiszbein and Schady (2009), show that CCTs have unambiguously met their primary short-run objective of increasing consumption and reducing poverty. Decomposition exercises, like those carried out by Stampini and Tornarollli (2012) and Levy and Schady (2013) also suggest that CCTs have been important in reducing poverty and inequality in the region in the last decade. CCT have not only increased consumption, they have also improved its composition. For example, Ruiz-Arranz *et al.* (2006) show that CCTs have increased the quantity, quality and variety of food intake, leading to consumption of more nutritious and expensive goods such as meat and vegetables.

Education. Another consistent finding across rigorous impact evaluations is that CCTs have substantially decreased child labor (Galiani and McEwan, 2013; Levy, 2006; Edmonds and Schady, 2012) and increased school enrollment and attendance (with rates that vary from 0.5 percentage points (pp) in Jamaica to 12.8 pp in Nicaragua). This is –together with higher use of health services- the key behavioral outcome intended by the theory of change of CCTs. CCTs also increased school attainment. For example, in Mexico after 3-5 years of participation in *Oportunidades*, the beneficiaries accumulated between ½ and 1 year of additional schooling.

The evidence on learning achievement is mixed (Fiszbein and Schady, 2009; García *et al.*, 2012; Saavedra and García, 2012). Barham *et al.* (2013b) find that, in Nicaragua, receiving the CCT for three years had significant impacts on years of schooling and on mathematics and language learning for young men 10 years after participating in the program. Learning increased by ¼ of a standard deviation, which loosely corresponds to half a year of learning. On the other hand, Behrman, Parker and Todd (2009b) find that higher enrollment levels have not resulted in better performance on achievement tests in Mexico. Evidence from outside the region is also mixed. Baird *et al.* (2011) report positive impacts on learning for a pilot CCT in Malawi, while Filmer and Schady (2014) and Benhassine *et al.* (2015) find no effect of a CCT on learning outcomes in Cambodia and Morocco, respectively.

From the perspective of CCTs, increasing schooling is the relevant indicator and main responsibility. At the same time, from a social perspective, learning is the purpose of schooling and the possibility that some children that go to school –whether they are CCT beneficiaries or not- may not learn is a reason for concern. The mixed findings stress the need to seriously address the issue of quantity and quality of education services, and how this interacts with the demand-side subsidies provided by CCTs.

Health. CCT programs have consistently shown positive effects on the use of preventive health services. Estimated impacts range between 6.3 pp in Nicaragua and 33 pp in Colombia. Some evaluations have also found that CCTs contributed to improvements in child height among some population groups, although the evidence is mixed (Fiszbein and Schady, 2009). There is some evidence that program beneficiaries have better health status and reduced morbidity (Gaarder *et al.*, 2010). Rasella *et al.* (2013) report that *Bolsa Família* reduced infant mortality caused by conditions associated with poverty, such as malnutrition and diarrhea. As is the case for education, health outcomes depend largely on the quality of health services, an issue that lies outside of CCTs' direct responsibility.

Child development. Rigorous evaluations suggest that CCT programs have positive impacts on child development. Paxson and Schady (2010) study the *Bono de Desarrollo Humano* (BDH) in Ecuador, and find significant effects on somewhat older children, although only for those in the poorest wealth quintile. Fernald and Hidrobo (2011) also study the BDH and find that children randomized to receive in Ecuador had higher scores on a test of the number of words children can say. Macours *et al.* (2012) find that the *Atención a Crisis* program in Nicaragua had an effect of about 0.12 standard deviations on the family of cognitive and behavioral outcomes they analyze.

Source: IDB (2014, p. 2-4).