IDB WORKING PAPER SERIES Nº IDB-WP-732

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

Inter-American Development Bank
Social Protection and Health Division

October 2016

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

October 2016

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 Working Paper Series ; 732)
Includes bibliographic references.
1. Transfer payments-Latin America. 2. Income maintenance programs-Latin America. 3. Economic assistance, Domestic-Latin America. 4. Poverty-Government policy-Latin America. 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-WP-732

http://www.iadb.org

http://www.iadb.org/SocialProtection

scl-sph@iadb.org

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 review the literature on the long-term impacts of Conditional Cash Transfer (CCT) programs in Latin America. Long-term impacts are defined 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 life-cycle transitions. The first is children exposed to CCTs in utero or early childhood, who have then transitioned to school ages. The second is children exposed to CCTs during school ages, who have then transitioned to early adulthood. The evidence is inconclusive. The experimental literature finds consistent positive long-term 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 in the evaluation studies. Non-experimental evidence also is mixed. Developing further opportunities for analyses with rigorous identification strategies for the measurement of long-term impacts should be high on the research agenda. As original beneficiaries continue to age, this should also be increasingly possible.

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

JEL code: I38.

[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 work was supported by funds from the IDB Economic and Sector Work "CCT Long-Term Impacts: Literature Review and Research Opportunities" (RG-K1421). We thank Ferdinando Regalia, Hector Salazar, Norbert Schady, Pablo Ibarrarán, Caridad Araujo, Nadin Medellin, Pedro Cueva, an anonymous reviewer and the participants in an IDB seminar for useful comments and suggestions. All remaining errors are our own. This working paper was previously published as (and replaces) IDB technical note number 923. 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.

Resumen

Revisamos la literatura sobre los impactos de largo plazo de los Programas de Transferencias Monetarias Condicionadas (PTMC) en América Latina. Definimos los impactos de largo plazo como aquellos que: (1) están relacionados con la acumulación de capital humano; y (2) se observan en una etapa posterior del ciclo de vida. Nos enfocamos en dos transiciones del ciclo de vida. La primera es aquella de los niños expuestos a los PTMC en el útero o en la primera infancia, quienes luego han transitado a la edad escolar. La segunda es aquella de los niños expuestos a los PTMC durante la edad escolar, quienes luego han transitado a la edad adulta. La evidencia no es concluyente. La literatura experimental encuentra efectos consistentemente positivos en escolaridad, así como algunos efectos positivos en el aprendizaje, las habilidades cognitivas, las habilidades socioemocionales y el ingreso proveniente de actividades no agrícolas. Sin embargo, otras estimaciones no son estadísticamente significativas; frecuentemente, no es posible discernir si esto se debe a la ausencia de impacto o a las debilidades metodológicas de los estudios. La evidencia no experimental también es mixta. El desarrollo de nuevas investigaciones para la medición de los impactos de largo plazo de los PTMC, basadas en estrategias de identificación rigurosas, debería ser una prioridad en la agenda de investigación. A medida que los beneficiarios originales crecen, implementar dichas investigaciones debería ser más sencillo.

Palabras clave: Programas de Transferencias Monetarias Condicionadas (PTMC), impactos de largo plazo, América Latina, *Progresa, Oportunidades, Prospera, Familias en Acción, Red de Protección Social*.

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; Levy and Schady 2013; Paes-Sousa, Regalia and Stampini 2013). Their objectives, including short-term poverty reduction via transfers, and long-term poverty reduction through enhanced investment in human capital, have broad policy appeal. The most common designs for CCTs generally follow the original design of *Progresa*, the Mexican program begun in 1997 (Levy 2006). Targeted to the poor, the principal program components include regular cash transfers to women (conditional on scheduled visits to health care providers for young children and on school enrollment and regular school attendance for school-age children) and social marketing to encourage investment in nutrition, health and education.

Numerous evaluations of CCTs, many based on rigorous experimental designs, consistently show positive short-term impacts. These include poverty alleviation, improved health and nutrition, particularly for young children, and increased school attainment for older children (Annex 1).[2] In contrast, few studies investigate whether these 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.

We critically review the existing evidence on whether, and to what extent, CCT programs have begun to achieve 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 effects on the welfare of the next generation.[3] This is likely because it is still too soon to investigate such next

[2] Other recent work examining short-term educational outcomes of CCTs (among other programs) include reviews by Murnane and Ganimian (2014) and Glewwe and Muralidharan (2015) and meta-analyses by Saavedra and García (2012), Baird et al. (2013) and McEwan (2015).

[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, beyond those associated with ongoing CCT benefits.

generation outcomes, as even the earliest programs only began in the late 1990s.[4] Consequently, the bulk of the 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 their long-term objectives. Moreover, some elements of human capital, such as completed grades of schooling, are enduring and therefore can themselves be considered long-term outcomes.

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 life-cycle transitions. The first is children exposed to CCTs in utero or early childhood (under age 6), who have transitioned to school ages. The second is children exposed to CCTs during school ages, who have transitioned to early adulthood, using age 18 as an approximate cut-off for adulthood. Short-term evidence makes clear that CCTs can lead to gains in nutrition and health for young children in the first group and to gains in schooling for the second. The duration of CCT exposure or the length of time since CCT exposure are not explicit criteria for selecting studies for our review. They are, however, important considerations for assessing impacts, and we document how they vary across studies. We restrict our review to research on Latin American programs, which have the longest history and continuous application, and consequently offer greater scope for analyzing long-term impacts. Moreover, the programs, while not identical, share many key design features, facilitating more general conclusions.

The bulk of evidence comes from three countries, for which both experimental and non-experimental evaluations on long-term effects are 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 differential impacts resulting from different program exposures over time. In both country cases, there were also surveys administered to similar, excluded populations, permitting assessment of absolute program effects using non-experimental techniques. Colombia's national CCT, *Familias en Acción* had no experimental evaluation. For that CCT, we review a long-term non-experimental

[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 a rigorous ex-ante impact evaluation design (Fiszbein and Schady 2009).

[5] While Mexico's CCT program was 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*.

evaluation of the overall program, as well as non-experimental research that relies on secondary data on learning and tertiary-level educational outcomes. Experimental evidence for Colombia comes from the randomized evaluation of a distinct pilot CCT program in 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 years) translates into impacts on schooling, 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, schooling and health outcomes during school ages. Section 4 concludes.

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

Nearly every Latin American CCT program that underwent a rigorous evaluation 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]; Murnane and Ganimian 2014; Glewwe and Muralidharan 2015). In many cases, positive results are also found for school progression. While these short-term impacts are encouraging, they fall short of providing definitive 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, provide evidence on these questions. For each country, we first describe the experimental

[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, are relatively 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).

evidence, as this typically presents fewer concerns regarding internal validity though it does not provide estimates of absolute program impacts for all countries. We then review the non-experimental evidence where validity relies on much stronger assumptions but for which estimates of absolute impacts exist. We conclude 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*. In particular, two studies by 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 baseline Survey of Household Socio-Economic Conditions (ENCASEH) along with several rounds of the associated rural household panel evaluation survey (ENCEL) through 2003. Their sample includes individuals between 9 and 15 years of age at the start of the program, encompassing the ages when students in Mexico typically transition from primary to secondary school.[8] Previous research revealed that this cohort had experienced the largest gains in schooling in the short term (Schultz 1997; 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 (Behrman, Parker and Todd 2009a, 2011).

In 2003, approximately six years after the program started in the early-treatment group, an 18-month difference in the length of program 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, on the order of 0.5 grades, were observed for those who were entering their last year of primary school around

[8] 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.

the time of their first exposure to the program. The authors 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 enrollment status. The study found 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, 2011).

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 2003, 65 percent of men in this cohort reported working compared with only 26 percent of women (Behrman, Parker and Todd 2009a, 2011).

Behrman, Parker and Todd (2009a, 2011) also found a negative impact of the differential program exposure on male migration of 2.0 percentage points (approximately 6 percent), but no significant effect for women. Male migration in this context is typically positively associated with entry into the labor market, 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, based on the comparison of individuals aged 9-15 at the start of the program living in households from the original early- and late-treatment 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 important selection concerns, the authors employ difference-in-difference 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,[9] 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 of absolute impacts, therefore, reflect different underlying parameters than the differential effects estimated using the experimental variation and consequently the two are not directly comparable.

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 because program transfer sizes, by design, were larger for women. Last, program effects increase with the length of exposure (four versus six 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 effects were 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 oldest men (19-21 in 2003), suggesting a shift away from agricultural work. For women, in contrast, there was a large increase (6.4 percentage points or approximately 20 percent) in the proportion working among the oldest (19-21 in 2003), the very same group who did not experience an increase in schooling.

Parker, Rubalcava and Teruel (2012) use a different panel survey with much lower attrition, the Mexican Family Life Survey panel (MXFLS 2002, 2005 and 2009), to evaluate longer-term differential impacts of *Progresa* on schooling, labor force participation and labor income. Their identification strategy exploits the gradual, albeit non-random, geographic rollout of the program. The study focuses on all children (i.e., combining boys and girls) who were 10-

[9] 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.

14 years old in 1997, the year before the *Progresa* transfers began in the original treatment areas. It compares outcomes for individuals in communities selected to receive *Progresa* in the initial years of program operation (1997 or 1998) with outcomes for individuals in communities selected to receive the program only in 2004 or later. They hence measure impacts for a difference in exposure of about 7 years, longer than the non-experimental estimates of the absolute program impacts, and substantially longer than the 18-month experimental differentials, that Behrman, Parker and Todd (2011) report. Parker, Rubalcava and Teruel (2012) employ difference-in-difference matching estimation to correct for selection bias, similar to the approach taken by Behrman, Parker and Todd (2011) although the set of matching variables is more limited. By 2005, when the individuals were 18-22 years old, a difference in program exposure of 7 years significantly increased grades of completed schooling by 0.5 years (or about 5 percent) and the probability of attending university by around 5 percentage points (an increase of approximately 40 percent). It also significantly increased labor force participation by about 8 percentage points (or about 15 percent), although these results are less precise. In contrast, no significant effects were found for hours worked or hourly labor earnings, possibly reflecting the fact that although early beneficiaries had more years of education they also likely have had less experience in the labor market than late beneficiaries. Impacts on hours worked and hourly labor earnings also may have been muted if in addition to working, early beneficiaries were more likely to be in school still, as suggested by the results on university attendance. Lastly, as the authors make clear, it is possible that the returns to program-induced increases in schooling are not large enough to lead to increased earnings for young adults.[10]

2.3 Mexico: Assessment of the Evidence

One important consideration for interpreting both experimental and non-experimental evaluations that rely on the Mexican evaluation panel surveys (ENCASEH and ENCEL) is the high rate of sample attrition, mainly a result of the survey protocol in which migrants were not followed to new locations. Forty percent of the individuals 9-15 when first interviewed in 1997, the primary sample for most analyses, were not found in 2003. Consequently, selectivity poses

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

a substantial threat to both the internal and external validity of the estimates based on the 2003 ENCEL data.[11]

To correct for the potential selection bias when using the 2003 data, Behrman, Parker and Todd (2009a, 2009b, 2011) follow two strategies. First, for some outcomes they are able to reduce the number of observations with missing information by using proxy information reported by remaining members in the household of origin for individual migrants who were not themselves directly interviewed. Second, for all analyses 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 or basic labor market outcomes), attrition is effectively reduced to around 20 percent of the 2003 sample. Such proxy reports are unavailable when no one in the original household is interviewed, however, and therefore do not correct for household-level attrition, e.g., when all household members have migrated from the community of origin. Moreover, they are unavailable for measures that require direct person-to-person interview of the respondent, such as the administration of achievement tests to assess learning. Finally, use of proxy information relies on the assumption that current household members accurately report the 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.

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 underlying this methodology is that conditional on observables within each group (early-treatment, late-treatment, and the non-experimental comparison), attrition is random. The internal validity of both the experimental and non-experimental estimates are jeopardized if there is important attrition-related selection based on unobservables that differs between treatment and control groups. In Nicaragua, Molina-Millan and Macours (2016) show that this assumption might not be valid, for example, for attrition associated with work-related migration among young adults. The external validity of the estimates is also jeopardized if there is important attrition-related selection overall.

[11] We do not explicitly review an unpublished study (Rodriguez-Oreggia and Freije 2012) that uses a subsequent round of ENCEL data in 2007. Since migrants are not followed, the sample is even more highly selected (with more than 60 percent attrition over baseline) and characterized by differential attrition for the early-treatment, late-treatment and non-experimental comparison groups, so that both the internal and external validity of the study appear to be weak.

A number of additional concerns can be raised for the non-experimental results presented in Behrman, Parker and Todd (2011). Pre-intervention baseline data (i.e., from 1997) for the non-experimental comparison group was not available and was collected retrospectively in 2003. This might introduce measurement error due to recall bias. Additionally, initially *Progresa* was geographically targeted to marginal areas; consequently, non-*Progresa* communities in the ENCEL survey may have been *a priori* different, in which case balance at the community level may be difficult to achieve. In addition to these concerns for the internal validity of the results, the findings of the above evaluations are not generalizable to the entire population covered by the program. They hold only for non-migrating beneficiaries, with the exception of outcomes for which there is proxy information. As returns to the program could be in part realized through migration, estimates using only the non-migrating beneficiaries might provide an incomplete picture.

Finally, for all of the experimental and non-experimental evidence based on the ENCEL data, external validity is also a concern. This is because the estimates are necessarily based on the sample for the original experimental evaluation, which was drawn from the poorest rural communities (with supply of health and education services). Subsequent program beneficiaries live in less poor rural areas and in urban areas, and therefore long-term program impacts for them may be different.

The MXFLS data used by Parker, Rubalcava and Teruel (2012) have lower levels of attrition than the ENCEL due to the difference in survey protocol with intensive migrant tracking. In addition, the long differential in exposure increases the power of the evaluation. However, these strengths are at least partially offset by the much smaller sample sizes (approximately 700 observations, even after combining men and women) as compared to the much larger ENCEL surveys that allow sample sizes 10 times that. 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 early- and late-treatment groups. Another potential concern with the study is its use of 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. Last, for some outcomes, such as labor market activities where there are strong gender differences, program impacts possibly differ by gender.

A final consideration for the whole set of studies pertains to the difficulty of studying young adults who are still undergoing important life cycle transitions. For example, the observed delayed entry in the labor market for males means 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 specific 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. Multiple measurements, so that time paths of effects could be traced out, would be even more valuable.

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. The 21 early-treatment localities became eligible for transfers in November 2000 and were eligible for three years, receiving their last transfers in late 2003. The 21 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. As in Mexico, 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 (Barham, Macours and Maluccio 2016). In contrast to Mexico, however, 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 randomly generated exposure diminishes with time since exposure in both early- and late-treatment 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 evaluation survey was conducted (Barham, Macours and Maluccio, 2013, 2016). All original households were tracked. When a household was found (and interviewed), but a former member 12 or under in 2000 was no longer resident, 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 into 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 (but for whom some proxy information could be collected, as in Mexico). For example, for males 9-12 years old in 2000, attrition in 2010 was 6 percent for schooling and basic occupation information, 12 percent for earnings data, and 19 percent for test scores.

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.

The 2010 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 (at which point many would have already dropped out). All estimates are weighted for attrition as described in Molina-Millan and Macours (2016).

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 from the long-term evaluation in Mexico, 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). The effect size was similar across the range of achievement tests. The study also found positive differential impacts of about 0.2 standard deviations 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.[12]

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, and 7 years after in the late-treatment group. As for Mexico, these estimates reflect different underlying parameters than the differential effects estimated using the experimental variation. 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, including the marginality index used for selecting localities. 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 generally larger in magnitude. For example, the absolute effect on grades attained is more than a full year of school in the early-treatment group. For women in the same age group, although there were no significant experimental differential effects on grades attained, the standardized tests or income, the non-experimental results demonstrate significant absolute learning effects of about 0.3 standard deviations that are similar in magnitude across the early-

[12] Parallel analyses for the same age cohort of women, however, showed no significant differential effects, possibly because the program had similar effects on this cohort of girls in the early- and late-treatment groups (as girls tended to drop out of school at older ages). See section 2.6.

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.

2.6 Nicaragua: Assessment of the Evidence

Many of the concerns detailed above in the assessment of the evidence for Mexico are also pertinent to Nicaragua (and Colombia below). For brevity we make shorter reference to them here, though they are not necessarily less important.

Despite attrition rates that are much lower than in other similarly long-term studies, attrition bias remains a source of concern, especially for certain age groups. The validity of the estimates still depends on the validity of assumptions made to re-weight the data to correct for sample selection and the quality of data provided by proxy informants. While Barham, Macours, and Maluccio (2016) indicate that estimates are robust to alternative assumptions, suggesting that they are reliable, it is not possible to determine the direction of any remaining bias.

The fact that *RPS* was not a national program and 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 would be seen in other, less poor, regions or countries.

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 with additional rounds of data collection will analysts be able to gauge the full long-term impact of *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 authors tested three alternative program designs. The first is referred to as the "basic treatment", in which transfers were made every other month, 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 then paid either: 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 (comprising secondary school students aged 12-18). Separately, the tertiary treatment is compared to a different randomized control group (comprising only students in upper secondary school at baseline, to match the treatment group for this intervention).[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.

The authors merge program participation data with national administrative records on upper secondary school graduation exams and enrollment in tertiary institutions for the eight years following the 2005 intervention. These merges were based on student ID number, full

[13] This is the only CCT program included in our review that operated exclusively in urban areas and that had no conditionalities related to younger children. 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 (Barrera-Osorio, Linden and Saavedra 2015).

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 secondary school exit exam, by 2.8 percentage points (about 4 percent). However, it is not possible to 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 about 10 and 20 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 first exposed in lower secondary education (grades 6 to 8).

These results broadly confirm short-term impacts estimated one year after the intervention for individuals in grade 11 at the start of the program, and thus on the cusp of the post-secondary school transition (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 (9.4 percentage points) or tertiary (48.9 percentage points) treatments in grade 11, but no significant effects on graduation rates for any of the treatment arms. For these same individuals, the program had a (corresponding) significant negative impact on labor force participation (reducing the probability that the primary activity was work and reducing hours worked and earnings from work in the last week) for those in the tertiary treatment.

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.

García et al. (2012) conduct a comprehensive 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 based on similarities with the treatment in terms of region, level of urbanization, number of eligible households, a quality of life index, and health and school infrastructure. The authors 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.

Examining young adults aged 18-26 in 2012, the study found 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 found a positive significant impact on the probability of graduating from upper secondary school, alongside a negative impact on the probability of enrolling in tertiary education, which is somewhat puzzling. The only significant impact on labor market outcomes found by the study was 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 on these same outcomes are not significantly different from zero.

The study further reports estimates of impacts 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 quite large compared to other studies on learning outcomes.[15] As this result is for the cohort that was 2-7 years old at baseline, it reflects in part the effect of exposure to the CCT program during early childhood.

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

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

[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.

merged with the national secondary school graduation test scores based on national ID number, full name and date of birth. 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 secondary school exit test score performance, conditional on completion.

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 successfully (which notably requires excellent unique identifiers at baseline), it is possible to conduct a study with low budget and high internal validity and statistical power. Match rates from merging the different data sources were relatively high (e.g., 70 percent or above for upper secondary school exit exams) and largely unrelated to treatment arm. 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) are subject to similar limitations as Barrera-Osorio, Linden and Saavedra (2015). In the former, however, there may be substantially more selection bias as match rates of the merge between program administrative records and school test data are less than 25 percent. 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 further sample selection bias. 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. Finally, as they acknowledge, the study design 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 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.

The concerns regarding simple difference or difference-in-difference estimates in García et al. (2012) are similar to those 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 (in the case of difference-in-differences) 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 adequately 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 reflect program-related changes in behavior, or 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 somewhat more limited. As above, we draw on long-term evidence 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 early- and late- experimental treatment localities described above for the school-age cohorts examined by Behrman, Parker and Todd (2009a). Outcomes are measured in 2007 for individuals who were in utero or under 13 months of age when the program started, and therefore 8-10 years old at the time of their follow-up survey.

The study found a significant reduction in behavioral problems. No significant experimental results are found for any measure of child growth, cognition or language when using the main experimental approach (comparison of mean outcomes in early- and late-treatments). The authors also present an alternative estimate, however, which takes into account the potential cumulative cash transfers received between initial household enrollment

and 2007. Potential 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, since transfer amounts are tied to gender, age and grade-level. Incorporating this latter source of non-experimental variation, the authors report a negative association between accumulated cash transfers and the number of reported behavioral problems, consistent with the findings from the experimental evaluation. In addition, they find that higher cash transfers are significantly and positively associated with height-for-age z-scores and higher verbal and cognitive test scores.

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, using first-difference and difference-in-difference estimators exploiting the original experiment, and also difference-in-difference non-experimental matching methodologies to estimate absolute program effects as in their work with the older cohorts. Because of their ages at baseline, however, all but the oldest of these children do not have meaningful baseline outcomes regarding schooling. They modify the difference-in-difference estimator to instead control for the outcomes of children who were 6-14 at baseline. Attrition for the targeted age group is lower than for the older cohort, at 20 percent for eligible children, 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 examine program impacts on grade progression, which reflects both enrollment and continuation. The results show that the 18-month differential exposure to *Progresa* had no significant effect on grade progression for children aged 9-11 in 2003.[17] In contrast, the difference-in-difference 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.

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

Similarly, the experimental 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 (2013) 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.[18]

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

[18] Results for girls are not reported.

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

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.

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, using as the comparison sample a different set of households who, in 1999, were ineligible for the program. 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*.

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 reflected in 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 these results.

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 exposed to the program in early childhood (aged 2-7 at baseline) . 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 CCTs succeed in breaking the intergenerational transmission of poverty), after nearly two decades of experience with these programs, some 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 might be 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 absolute 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 full transition to the labor market. Only one study (of the *Subsidios* program 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 complicated by the inherent tradeoff between additional schooling and shorter work experience.

Overall, we find that the existing evidence on CCTs long-term impacts in Latin America is inconclusive. 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, along with greater concerns about internal validity.

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 credible 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 rigorous identification strategy for the estimation of long-term impacts. Encouragingly, as initial beneficiaries now make their transition to adulthood, more opportunities 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. (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). “Medium 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.

Glewwe, P., and K. Muralidharan, K. (2015). Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications. Research on Improving Systems of Education Working paper, RISE-WP-15/001.

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. (2015). “Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments.” *Review of Educational Research*, vol. 85(3): 353-94.

Molina-Millan, T. and Macours, K. (2016). “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.E. 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

Schultz, T.P. (1997). “School Subsidies for the Poor: Evaluating the Mexican Progresa poverty program.” *Journal of Development Economics,* 74: 199-250.

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.* (2016) 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).