Long-Term Impacts of Conditional Cash Transfers: Review of the Evidence[i]

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

July 2018

Abstract

Conditional Cash Transfer (CCT) programs, started in the late 1990s in Latin America, have become the antipoverty program of choice in many developing countries in the region and beyond. We review the literature on their long-term impacts related to the accumulation of human capital and observed after children have reached a later stage of their life cycle, focusing on two life-cycle transitions. The first includes children exposed to CCTs in utero or early childhood who have reached school ages. The second includes children exposed to CCTs during school ages who have reached young adulthood. Most studies find positive long-term effects on schooling, but fewer find positive impacts on cognitive skills, learning or socioemotional skills. Impacts on employment and earnings are mixed, possibly because former beneficiaries were often still too young. A number of studies find estimates that are not statistically different from zero, but for which it is often not possible to be confident that this is due to an actual lack of impact rather than to the methodological challenges facing all long-term evaluations. 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 age, this should also be increasingly possible, and indeed important before concluding whether or not CCTs lead to sustainable poverty reduction.

JEL code: I38.

Keywords: Conditional Cash Transfers (CCTs), long-term impacts, *PROGRESA*.

1. Introduction

Conditional Cash Transfer (CCT) programs started in the late 1990s in Latin America and have become the antipoverty program of choice in many developing countries in the region and beyond.[ii] Their objectives, including short-term poverty reduction via cash transfers and long-term poverty reduction through enhanced investment in human capital, have broad policy appeal. Most CCTs follow the general design of *PROGRESA*, the Mexican CCT program begun in 1997 (Levy 2006). Targeted to the poor, their principal program components include regular (every month or two) cash transfers to women—conditional on scheduled visits to healthcare 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 experimental designs, find positive short-term impacts. These include current poverty alleviation, some improved nutrition and health for young children and increased school attainment for older children (Fiszbein and Schady 2009).[iii] In contrast, fewer studies investigate whether these short-term gains eventually translate into sustained *longer-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 eventual labor market or other adult outcomes? And, ultimately, do CCTs improve the welfare of the next generation? The body of evidence demonstrating short-term impacts on human capital accumulation points to the potential for such long-term benefits. However, since CCTs typically operate in contexts with multiple labor market frictions, as well as other market imperfections, the magnitude of any such long-term impacts is an empirical question. After two decades of experience with these programs, there is a growing need, alongside increasing possibilities, to address these questions.

In this paper, we critically review the existing evidence on whether, and to what extent, CCT programs have begun to achieve their long-term objectives. To our knowledge, there is no research examining whether households formed by individuals who benefitted from the interventions as children or teenagers sustainably escaped poverty.[iv] This is likely because it is still too soon to investigate such next generation outcomes as even the earliest programs only began in the late 1990s. Consequently, the bulk of the research we review focuses on whether CCTs have led to an increase in human capital and how this influences labor market outcomes. Some elements of human capital, such as completed grades of schooling, are enduring and therefore can themselves be considered long-term outcomes. Improved labor market outcomes are a key, and arguably necessary, aspect of ultimate program success.

We consider long-term impacts related to the accumulation of human capital and observed after beneficiary children have reached a later stage of their life cycle, focusing on two life-cycle transitions. The first includes children exposed to CCTs in utero or early childhood (under age 6) who have transitioned to school ages. The second includes children exposed to CCTs during school ages who have transitioned to young adulthood, treating 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 older children in the second. The duration of CCT exposure or the length of time since CCT exposure are not explicit criteria for including studies in the review. They are, however, important considerations for assessing impacts, and we document how they differ across studies.

The evidence comes from both experimental and non-experimental studies. The former use the experimental variation created by random assignment or rollout of a CCT for its evaluation. Often, however, it was planned that the experimental control group itself would receive the program a few years or so after program start. Studies investigating long-term impacts of those programs exploit the randomized design together with differences in timing and the age-specific conditionality of components of the program to estimate *differential* program impacts of “early” versus “late” treatment. Non-experimental studies use a variety of methods and focus on CCTs for which the allocation of the program between beneficiaries and non-beneficiaries was determined following a nonrandom assignment rule, for example targeting poorer areas or households first. In many cases, these enable estimates of *absolute* program impacts, and in others of exposure differentials. Differential and absolute impact estimates reflect different underlying parameters and the two are not directly comparable. All studies considered were designed to provide estimated magnitudes and probability statements about the program effects.

Reflecting the pioneering role policymakers in the region had in their adoption and scale-up, the bulk of the research comes from Latin America where CCT programs have the longest history. For comparability, we focus on programs sharing key design features with *PROGRESA*, including conditionality pertaining to the health of young children and education of school-age children. To ensure coverage outside of Latin America, we also include CCT programs that differ from the benchmark in that they include only an education component. These programs target school-age children and often focus on improving the transition from primary to secondary, or from secondary to tertiary education, as well as entry into the labor market. Programs covered are listed in Table 1, which summarizes key design similarities alongside some important differences including scale and research designs utilized for their evaluations. In the appendix, we provide a more detailed description of search methods and inclusion criteria for studies in this review.

The remainder of the paper is organized as follows. In Section 2, we review research that examines how exposure to CCT programs for school-age individuals (under age 18 years) translates into impacts on schooling, learning, labor market outcomes, migration and marriage markets. Given the timeframes covered by the existing studies, most evidence is limited to impacts during young adulthood and hence reflects at least in part a mechanical relationship between additional schooling and later labor market entry and family formation. We present the evidence for health and education CCTs modeled after *PROGRESA* first, and then separately the evidence for education-only CCTs. In Section 3, we turn to the younger cohort describing the research that examines how exposure in utero or early childhood translates into better anthropometric, cognitive, learning or education outcomes at later school ages. Section 4 concludes.

2. Exposure to CCTs during School Ages and Outcomes in Young Adulthood

2.1. Health and Education CCTs

Most CCT programs incorporating both health and education that have undergone a rigorous evaluation have been shown to have positive short-term absolute impacts on school enrollment and attendance for children subject to schooling-related conditionalities, though the magnitudes of those impacts vary with characteristics of the program and target population (Fiszbein and Schady 2009; Murnane and Ganimian 2014; Glewwe and Muralidharan 2015). In many cases, there are also positive effects on school progression. While these short-term impacts are indicative, they only permit projections of future long-term gains and therefore fall short of providing definitive evidence on the more lasting changes that are the ultimate objectives of CCTs. Does increased schooling in the short run lead to increased final schooling attainment? Does it also lead to better learning or improved labor market outcomes?

Several studies based on programs in Latin America provide evidence on these questions. We begin with the research on the large national programs in Mexico and Colombia, and then turn to the regional programs in Nicaragua and Honduras. Last, we present the case of Ecuador, another national program where conditionalities were part of the initial program design but never monitored. For each case, we first describe any available experimental evidence, as this typically presents fewer concerns regarding internal validity though in some cases it only provides estimates of differential program impacts. We then review the non-experimental evidence where validity relies on much stronger assumptions but for which more estimates of absolute impacts are possible. We conclude the discussion of each country case with a critical assessment of the evidence.

2.1.1 Mexico

The Mexican CCT program *PROGESA* started in the central region of Mexico in 1997, but was scaled up to a large national program shortly after. Its well-known experimental evaluation focused on 506 rural localities in seven states; subsequent non-experimental evaluations exploit the national rollout.[v]

Experimental Evidence

The experimental design of the initial evaluation of *PROGRESA* and the incorporation of the control group enables examination of long-term assessment of an 18-month differential. Behrman, Parker and Todd (2009a, 2011) examine whether differential exposure to the program during school-age years significantly affected schooling and learning, labor market outcomes, migration and marriage. They use a 1997 baseline survey along with follow-up rounds of the linked household panel evaluation survey through 2003. They focus on individuals ages 9-15 years at the start of the program, encompassing the ages when students in Mexico typically transition from primary to secondary school. To assess long-term differential impacts, Behrman, Parker and Todd (2009a, 2011) examine this cohort in 2003 (i.e., at ages 15-21) and exploit the initial experimental evaluation design, in which the 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 experimental evaluation, which provides differential, rather than absolute, program impacts.

In 2003, approximately six years after the program began in the early-treatment group, there was a differential impact of 0.2 grades completed for both men and women in the cohort; i.e., grades attained were 0.2 higher in the early- versus the late-treatment groups. Larger differential effects, of about 0.5 grades attained, were observed for those entering their last year of primary school around the time of their first exposure to the program. The authors also consider whether the increased grades attained translated into more learning, examining impacts on three achievement tests covering reading, writing and mathematics skills. Achievement tests were administered to all respondents in their homes, regardless of enrollment status or grades attained. Despite differences in grades attained, however, the study finds no significant differential impact on learning.

In the labor market, differential exposure to *PROGRESA* significantly decreased male, but not female, labor force participation. Male labor force participation declined by 2.7 percentage points (p.p.) (approximately 4 percent) for the 15-21 year-olds, consistent with delayed entry into the labor market associated with their increase in grades attained. The insignificant effect for women could reflect in part 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. There was also a negative impact of differential program exposure on male migration of 2.0 p.p. (on a base of approximately 6 percent who had left their communities by 2003), but no significant effect for women. Male migration in this context is typically positively associated with entry into the labor market and therefore effects on it may explain in part the reduction in labor force participation. Finally, Behrman, Parker and Todd (2009a) find no statistically significant differential impact on the probability that either men or women were married in 2003.

Building on this empirical strategy, Adhvaryu et al. (2018) analyze the heterogeneity of differential program impacts between individuals who had experienced negative rainfall shocks during their first year of life (before the program started), and all others. They examine education outcomes for all children 12-18 years old in the 2003 survey and labor market outcomes for 18-year olds, pooling boys and girls in a single-difference framework. For educational outcomes, they largely replicate the Behrman, Parker and Todd (2009a) difference-in-difference findings, and find larger differential impacts for those exposed to negative rainfall shocks in the first year of life. For the 18-year olds, they find similar heterogeneity, including an approximately 8 p.p. increase (per year of differential program exposure) in labor force participation and in the probability of holding a stable (i.e., non-laborer) job for those with the unfavorable early childhood conditions, but no effects for those without early childhood rainfall shocks.

Non-Experimental Evidence

Behrman, Parker and Todd (2011) present non-experimental evidence for the 15-21 cohort in 2003 comparing those from the original evaluation communities with the same age cohort living in other rural communities that were not part of the original experiment and had not yet been incorporated into the program. Because this comparison is no longer based on the experimental design, it introduces selection concerns and the authors employ difference-in-difference matching methods to take into account differences in observed characteristics between samples. Conditional on the assumption that the matching on observable characteristics also eliminates the selection bias related to unobservable characteristics, this approach allows estimation of the absolute impacts of approximately four and six years of program exposure (when the early- and late-treatment groups, respectively, are compared with the non-experimental comparison group). These estimates of absolute impacts, therefore, are not directly comparable to the differential effects estimated using the experimental variation reported above.

The study finds absolute impacts of between 0.5 and 1.0 additional grades attained for all but the oldest women (those ages 19-21 in 2003). Impacts for men are modestly larger across all ages in the cohort, notable because program transfer sizes, by design, were larger for women. Program effects increase with the length of exposure (four versus six years) to the program. As these estimates are more than double the experimental differential results, they underscore the important distinction between differential and absolute program impacts, and suggest caution against considering the former as measures of the latter.

For young men ages 15-16 in 2003, there was a large reduction (14.0 p.p. or approximately 30 percent) in labor force participation, consistent with that age group still being more likely to be in school. However, there were no significant effects for other age groups. Examining different types of work, the study finds a large decrease (9.0 p.p. or approximately 25 percent) in participation in agricultural work for the oldest men (19-21 in 2003). For women, in contrast, there was a large increase (6.4 p.p. points or approximately 20 percent) in the proportion working among the oldest (19-21 in 2003), the same group who did not experience an increase in grades attained.

In separate work, Parker, Rubalcava and Teruel (2012) use the smaller but nationally representative Mexican Family Life Survey panel (MxFLS 2002 and 2005) to evaluate longer-term differential impacts of *PROGRESA*. Their identification strategy exploits the gradual, albeit non-random, national geographic rollout of the program. The study combines boys and girls 10-

14 years old in 1997, the year before the program began in the early-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 in 2004 or later, using difference-in-difference matching estimations to correct for selection bias similar to Behrman, Parker and Todd (2011).

By 2005, when the individuals were 18-22 years old, a difference in program exposure of 7 years significantly increased grades attained by about 0.5 years and the probability of attending university by about 5 p.p. (an increase of approximately 40 percent). It also significantly increased labor force participation by about 8 p.p. (or approximately 15 percent). In contrast, no significant effects were observed for hours worked or hourly labor earnings, possibly reflecting the fact that although they had higher grades attained, early-treatment beneficiaries also likely had less experience in the labor market. Impacts on hours worked and hourly labor earnings also may have been muted if in addition to working, early-treatment beneficiaries were more likely to be in school still, as suggested by the results on university attendance. Lastly, as the authors suggest, returns to program-induced increases in schooling may not be large enough to lead to increased earnings for young adults.

Parker and Vogl (2018) estimate absolute impacts 13 years after the start of *PROGRESA*, using a related approach, but with much larger samples. They use the 10 percent sample of the Mexican 2010 population census and exploit the non-random municipal-level national rollout of the program, together with cohort-level variation in exposure at critical ages. The main analysis uses difference-in-difference and compares the cohort of children ages 7-11 at the start of the program in 1997, who fully benefitted from the educational component of the program, with an older cohort of 15-19 year-olds living in the same municipalities, who were in general too old to have benefitted directly. The analysis also accounts for possible convergence in outcomes related to baseline characteristics, by controlling for the interaction between the cohort and the program intensity in 2005 (an endogenous proxy for poverty). With the census data, they can assign individuals living throughout Mexico to their municipality of residence in 2005, allowing to account for recent internal migration.

Estimated absolute program impacts, measured in 2010 when the cohort of interest (7-11 in 1997) is 20-24 years old are large and show an increase of 1.4 additional grades attained for both men and women, with accompanying higher levels of secondary- but not university-level schooling. Women also exhibit a 7-11 p.p. (or an approximately one-third) increase in labor force participation, and an approximately 50 percent increase in earnings (from a low base). Part of this may reflect similarly sized increases in migration for women. For men, they find a shift away from agricultural work, more hours worked weekly and about a 9 p.p. (approximately 60 percent) increase in the probability of working in the formal sector, but no significant increase in monthly earnings. There are no effects on migration for men nor on marriage for either men or women.

Assessment of the Evidence

We now critically assess the evidence for the Mexican case of *PROGRESA*. The discussion is somewhat longer for Mexico than for the other cases that follow below, but not because the evidence is less strong. Rather, as it is the first case described we provide a more thorough explanation of a number of concerns repeated for many of the other evaluations. Additionally, there is substantially more research available for Mexico.

One important consideration for interpreting both experimental and non-experimental evaluations that rely on the Mexican evaluation panel surveys is the high rate of sample attrition for these cohorts, mainly a result of the survey protocol in which migrants were not followed to new locations. About 40 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 a substantial threat to both the internal and external validity of the estimates based on the 2003 evaluation survey data.[vi]

To correct for the potential selection bias when using the 2003 data, Behrman, Parker and Todd (2009a, 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 not themselves directly interviewed. Second, for all analyses they apply a density reweighting method to correct for sample selection. Adhvaryu et al. (2018), however, do not reweight for attrition.

For outcomes on which proxy information is available from other household members (e.g., reported grades attained or basic labor market outcomes), attrition is effectively halved to around 20 percent. Such proxy reports are unavailable when no one in the original household is interviewed, however, and therefore do not correct for household-level attrition, when all household members have migrated from the community of origin. Proxy reports are also unavailable for measures that require direct person-to-person interview of the respondent, such as the administration of achievement tests. 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 either random or at least balanced across treatment groups.

The density reweighting method weights the sample of individuals interviewed in 2003 in order to replicate the baseline distribution of individual and household characteristics. The key assumption underlying this methodology is that conditional on observables within each group (early-treatment, late-treatment, and the non-experimental comparison groups), attrition is random (Fitzgerald, Gottschalk and Moffitt 1998). The internal validity of both the experimental and non-experimental estimates are jeopardized if there is sizable attrition-related selection based on unobservables that differs between treatment and control groups. In Nicaragua, Molina Millán and Macours (2017) show that this assumption might not be valid, e.g., for attrition associated with work-related migration among young adults for whom heterogeneous treatment effects may be large.

A different concern for the non-experimental results presented in Behrman, Parker and Todd (2011) is that pre-intervention baseline data were unavailable for the non-experimental comparison group. The difference-in-difference estimates therefore rely on six-year retrospective data collected in 2003, possibly introducing measurement error due to recall bias. Moreover, *PROGRESA* was geographically targeted to marginal areas first; consequently, non-program communities in the survey may be different *a priori*, in which case balance between program and non-program areas 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 immediately generalizable to the entire population covered by the program. With the exception of outcomes for which there is proxy information, they only hold for non-migrating beneficiaries. As returns to the program could be in part realized through migration, estimates using only the non-migrating beneficiaries might provide an incomplete picture. Attrition hence likely also implies that the estimates are not representative of the target population. A second external validity concern about the evidence based on the evaluation surveys is that the sample for the original experimental evaluation was drawn from relatively poor rural communities. Long-term program impacts in urban areas, for instance, may be very different.

The MxFLS data used by Parker, Rubalcava and Teruel (2012) have lower levels of attrition than the evaluation surveys due to the difference in survey protocol with intensive migrant tracking (albeit from an initial sample frame five years after *PROGRESA* began). In addition, the longer differential in exposure, while not random, increases the power of the evaluation. However, these strengths are at least partially offset by smaller sample sizes that are about one tenth those of the large evaluation surveys. And while the full MxFLS is nationally representative, the authors restrict the analysis to non-indigenous youth in rural communities with low levels of poverty. Communities incorporated into *PROGRESA* in 2004 and after are less poor. To strengthen comparability, the poorest communities (i.e., those incorporated at the very beginning of *PROGRESA*) are excluded from the analysis. This restriction has the benefit of increasing internal validity at the cost of reducing external validity. The study also uses five-year retrospective data on some outcomes for the difference-in-difference calculations, possibly introducing recall biases.

The 2010 Mexican census data used by Parker and Vogl (2018) by definition also have relatively low levels of attrition, and technically only the international migrants and deceased are missing. As the authors recognize, this is a larger potential concern for the men, who both migrate more and have higher mortality. But the Mexican census does not allow assignment of individuals to the location where they lived at the time of initial program exposure, and this may be more problematic for the older comparison cohort (15-19 in 1997) than for the younger cohort (7-11 in 1997). The study uses the spatial variation (driven by rollout and take-up) in program intensity for municipalities incorporated between 1997-1999 and 2001-2005 to identify the program effect, and controls for program intensity in 2005. Without understanding the drivers underlying the variation in program intensity, however, possible program placement selection concerns remain.[vii]

2.1.2 Colombia

Colombia's national CCT program, *Familias en Acción*, began in 2002 in municipalities that were not departmental capitals and had fewer than 100,000 residents, eventually expanding to larger municipalities. It did not have an experimental impact evaluation but did have comprehensive short- and long-term evaluations (Attanasio et al. 2005; García et al.

2012). Households were eligible if they had a proxy means test score (assigned by the national social programs targeting system, known as SISBEN) below a given threshold.

Non-Experimental Evidence

García et al. (2012) report on the non-experimental evaluation of differential long-term impacts on schooling and labor market outcomes. The identification strategy uses single-difference (with baseline controls) or difference-in-difference estimation and compares children from eligible households in municipalities covered by the program in 2002 with children from potentially eligible households from comparable areas that were not targeted by the program until 2007, when it expanded to larger municipalities. The comparison municipalities were selected based on similarities with the treatment municipalities in terms of region, degree of urbanization, number of eligible households, a quality of life index and health and school infrastructure. The authors investigate differential exposure impacts, using pre-program data from 2002 and a follow-up panel survey conducted near the beginning of 2012.

Examining children initially exposed between ages 8-16 and thus young adults ages 18-26 in 2012, they find that 2 to 5 additional years of exposure to *Familias en Acción* increased school attainment by 0.6 grades in rural areas. Also for rural areas, there was a positive and 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 reported by the study was an increase of 2.5 p.p. in the probability of formal employment among women in rural areas. In urban areas for this age group, estimates of all of the schooling and labor market impacts are not significantly different from zero.

The study further reports estimates of differential impacts on learning and cognition, based on a mathematics test and the Raven's Progressive Matrices test, for adolescents ages 12-17 in 2012. It shows that the differential exposure to *Familias en Acción* had no significant impact on the Raven but increased mathematics scores by 1.07 standard deviations (SD), which is quite large compared to other studies on learning outcomes. 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, as well as during later school ages.

Baez and Camacho (2011) investigate the absolute impacts of up to nine years of potential exposure to *Familias en Acción* using household survey data, registration records from SISBEN and administrative data on the results of the secondary school graduation test. The study uses both difference-in-difference matching and a 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 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 p.p. more likely than non-beneficiary children to complete secondary school. However, the authors found no evidence of differential secondary school graduation test score performance, conditional on completion.

Duque, Rosales-Rueda and Sanchez (2018) similarly combine administrative data from the secondary school graduation test (up to 2014) and the SISBEN scores, as well as incorporating data on the universe of Colombia's public schools (up to 2015). The paper focuses on children born between 1988 and 2000, and aims to analyze whether absolute program impacts are higher for children exposed to adverse weather events (similar to what Adhvaryu et al. [2018] do for Mexico). The absolute impact of the CCT is identified at ages between ages 14 and 27 through a fuzzy RDD, after merging the different administrative datasets. The findings show positive impact of being eligible for *Familias en Acción* on school progression and secondary school completion (17 percent higher). In contrast with Baez and Camacho (2011), and possibly because additional years of data could be included and children were older, they also find a positive impact on the secondary school graduation test score of (0.13 SD higher). In contrast to the findings for Mexico, however, treatment effects do not differ depending on early life adverse events.

Assessment of the Evidence

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. 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 of the different geographic areas. 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. This is a general concern applicable to difference-in-difference estimates of other countries as well. 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. Attrition in the 10-year panel survey was more than a third and is related to migration; no adjustments for attrition are made.

The results in Baez and Camacho (2011) and Duque, Rosales-Rueda and Sanchez (2018) are subject to limitations common to using administrative data. If data sources can be fully combined and without error (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. In Baez and Camacho (2011) match rates of the merge between program administrative records and school test data are less than 25 percent, potentially indicating important selection bias. Duque, Rosales-Rueda and Sanchez (2018) have a match rate of about 50 percent, and show it is not correlated with CCT eligibility. They also show evidence of selective matching on other observables, but point estimates suggest selectivity overall is limited, reducing such concerns for this study. That said, the results for test scores are not generalizable to the complete population covered by *Familias en Acción* as the test was administered only to children who stayed in school and progressed until grade 11, which may introduce further sample selection bias.

The study by Baez and Camacho (2011) has an additional caveat as the design cannot distinguish the effect for different ages from the effect due to length of exposure, as they acknowledge. 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.

Finally, an important drawback of the use of the RDD approach used in Colombia is that SISBEN score 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 potentially confound the impacts of different programs.

2.1.3 Nicaragua

Modeled after *PROGRESA*, the Nicaraguan CCT program, the *Red de Protección Social* (*RPS*) started in 2000, and had a short-term experimental evaluation built into its initial stages. Unlike Mexico's *PROGRESA* and Colombia's *Familias en Acción*, *RPS* was a regional program, and the evaluation focused on six rural municipalities with initial poverty rates of around 80 percent. In 2002, a non-experimental comparison group was drawn from a sample of individuals living in adjacent municipalities.

Experimental Evidence

Forty-two localities in the six municipalities 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. The 21 late-treatment localities were phased in at the beginning of 2003 and also were eligible for three years. Households in the early-treatment group did not receive any transfers after 2003. 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 estimates of differential, rather than absolute, impacts (Barham, Macours and Maluccio 2017, 2018). In contrast to Mexico, however, it compares groups that randomly received the program for a fixed 3-year period at different points in time. A long-term follow-up panel evaluation survey was conducted approximately 10 years after the start of the program. All original households and all individuals 12 years or younger in 2000 were tracked throughout Nicaragua and international migrants tracked into Costa Rica. To show that the relatively limited number of randomized localities is not driving the results, the authors also estimate Fisher exact p-values using randomization inference.

Barham, Macours and Maluccio (2017) use the 2010 follow-up survey, together with pre-intervention data, 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 considered critical for educational investments (the age at which the probability of dropping out of school is high) versus three years later. All estimates are weighted for attrition.

Results show that a significant differential increase in grades attained five years after the end of the program is accompanied by gains in learning. Males in the early-treatment group experienced an average improvement of about 0.2 SD on mathematics and Spanish tests. The study also finds positive differential impacts of about 0.2 SD on socio-emotional outcomes, such as optimism and positive self-evaluation. Finally, there were also differential effects on labor market outcomes, with the young men 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 *RPS* produced large long-term differential impacts on earnings for men, consistent with increased human capital leading to better labor market outcomes.

Following a parallel approach, Barham, Macours and Maluccio (2018) show that girls randomly exposed to the CCT during early teenage years were also more likely to be economically active and had higher incomes than those exposed three years later. These results are similar to those for boys in the same age cohort. The pattern of findings, however, suggests the causal pathway explaining the differential impacts is different for girls. For girls, there were no differential effects on learning, which contrasts with the findings for the boys. The differential effect on grades attained is also small and not statistically significant. However, there were differential reproductive health impacts, with early-treatment girls starting sexual activity later than late-treatment girls, resulting in lower fertility. The timing of program exposure, in particular exposure to nutrition shocks and reproductive health information at the onset of puberty in the late-treatment group, may help explain the differential fertility results. Overall, the results suggest that, at least for girls, the long-term labor market impacts likely in part reflect changes in reproductive health outcomes rather than changes in education.

Non-Experimental Evidence

Barham, Macours and Maluccio (2017, 2018) 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 seven years after the start in the late-treatment group. The non-experimental comparison group is drawn from a sample of individuals living in 21 localities in adjacent municipalities, selected using the same marginality index used for selecting localities in the experimental evaluation and first surveyed in 2002, i.e., two years after the start of the program. Individuals from the experimental sample are matched to individuals in the non-experimental comparison sample. 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 matching results show positive absolute impacts on schooling and learning outcomes for both young men and women that are in line with the experimental effects but larger in magnitude. For example, the absolute effect on grades attained is more than one full year. For women the non-experimental results yield absolute learning effects of about 0.25 SD. This suggests that positive and equal absolute impacts on learning in both the early- and the late-treatment groups may underlie the lack of significant experimental differential impacts 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, age 13 in 2000, as well as by findings from alternative non-experimental estimates using national census data five years after the start of the program showing absolute gains in education. The matching estimates of absolute effects on earnings are positive but not significant for both young men and women.

Assessment of the Evidence

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

A strategy of tracking all migrants led to attrition rates that are lower than in some similarly long-term studies, but still between 6 and 22 percent depending on gender, the outcome of interest, and whether there is proxy information. Therefore, attrition bias remains a source of concern, especially as attrition may be related to migration for work. Molina Millán and Macours (2017) demonstrate that there is remaining attrition bias in the standard intent-to-treat estimates for boys, and develop a correction that uses information from the intensive tracking carried out during the field work to reweight and correct for sample selection. The validity of the estimates still depends on the validity of assumptions made to reweight the data to correct for sample selection and the quality of data provided by proxy informants. They also show that without intensive tracking, estimates on grades attained and labor force participation would have been overestimated, demonstrating that impact estimates based on individuals who remained in the origin communities are not necessarily downward biased, as is sometimes argued in other studies.

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 less marginal or more educated populations in Latin America, though arguably more comparable to other developing 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 follow-up will analysts be able to gauge the full long-term impact in terms of returns to human capital and, more broadly, welfare outcomes.

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

2.1.4. Honduras

Also modelled after *PROGRESA*, the Honduran CCT program *Programa de Asignación Familiar-II* (*PRAF-II*) began in 2000 in 70 western municipalities with the highest levels of child malnutrition. The program was a five-year CCT originally evaluated via municipal-level randomized assignment. In contrast to Mexico and Nicaragua, the CCT was never phased in to the experimental control group so studies on it provide absolute program impacts.

Experimental Evidence

Molina Millán et al. (2018) use the municipal-level randomized assignment of *PRAF-II*, combined with individual-level data from the national population census 13 years after the start of the program, to analyze long-term absolute impacts across a wide range of age groups. At the start of the program in 2000, 20 municipalities were assigned to the basic CCT and 20 to the control. The remaining 30 municipalities were assigned supply side interventions (with or without the CCT), but delays hampered the implementation of those components. The authors measure long-term impacts of exposure during school-going years for cohorts 6-13 years old at the time of the program start in 2000, who were thus 19-26 years old in 2013. Individuals can be assigned to their municipality of birth and because international migration is limited among the relevant cohorts, the study provides an example of an experimental long-term follow-up with secondary data that can largely circumvent the attrition concerns of many other studies. The main results compare young adults from the 20 municipalities exposed to the basic CCT with the 20 control municipalities. Given the relatively limited number of municipalities in the randomization, the authors also use randomization inference of the sharp null hypotheses and the Fisher exact p-values demonstrating robustness of the findings. Results are presented for non-indigenous and indigenous groups separately.

The authors find significant increases in grade attainment for non-indigenous men and women of about half a grade, and for older (24-26) indigenous women, but no significant results for indigenous men. They also find relatively large increases in secondary school completion rates and in the probability of enrolling in university for indigenous and non-indigenous men and women, increasing more than 50 percent (compared with low levels in the control). Moreover, the study shows an increase in international migration of 4 p.p. (compared to a base of 3 percent in the control) among non-indigenous men, and smaller, but significant effects for indigenous men and non-indigenous women.

As the authors note, the impact of the CCT on international migration (small in absolute levels but large in relative terms) may still lead to sample selection issues when analyzing the program effect on education and labor outcomes, even though all domestic migration is accounted for in the census. Given the low levels of international migration in Honduras, sample selection concerns are likely smaller, however, than for work using longitudinal surveys in which attrition is typically an order of magnitude larger.

Molina Millán et al. (2018) complement the findings from the census with a parallel analysis of the long-term impacts on labor force participation and earnings for the same cohorts, using repeated cross-sections of the national Multiple Purpose Household Survey from 2010-2016. Overall, the participation and earnings results for young women and men present a mixed picture regarding the potential labor market returns. For women from predominantly non-indigenous villages, there is a 12 p.p. (approximately 30 percent) reduction in labor force participation and, consistent with that, a decline in earnings (even if earnings per hour worked increase). For men, the results show no evidence of any effects in the labor market. The authors acknowledge that despite reweighting, these results should be interpreted cautiously as the household sample, by design, is not representative at the municipality level. More generally, these findings echo the challenges faced in several other countries regarding the examination of labor market returns when transitions to the labor market are not yet complete.

Ham and Michelson (2018) also use the randomized design of *PRAF-II* and national census data, focusing on the different treatment groups. They compare municipal-level averages (constructed from the population census) in a difference-in-difference framework for children ages 6-12 in 2001 (eight months after the start of the program) and 18-24 in 2013. The analysis controls for a number of time-varying controls, but does not account for population weights. For the broad age groups examined, their results point to increases in grades attained, secondary school completion and labor force participation, especially for the women who received the CCT combined with the supply side interventions. Because the paper uses average outcomes based on residence in 2013, it makes the strong assumption that migration between 2001 and 2013 (over 25 percent for this age cohort with less than 5 percent internal to the program municipalities) does not affect the internal validity, and cannot account for any returns that materialize through migration. The large set of time-varying controls (including programs introduced in the same municipalities post-2005) also raise concerns of endogeneity, overfitting and results being driven by modeling assumptions.

2.1.5 Ecuador

Araujo, Bosch and Schady (2018) use a regression discontinuity approach to analyze the 10-year absolute program impacts of the Bono de Desarollo Humano (BDH), the national cash transfer program in Ecuador that began in 2004 and at its peak covered 40 percent of households in the country. Unlike the other CCT programs considered in this review, although conditionalities were part of the initial program design, ultimately the transfers were not explicitly conditional on pre-specified behaviors like school enrollment. However, households were encouraged by the program to spend transfer income on children. The authors use information on school attainment and work status from a 2013/14 poverty census of program areas and compare children living in households who were just eligible versus those in households who were just ineligible for transfers. They focus on children 9-15 years old at program start (and therefore in late childhood so likely to undergo transition to secondary school from 2003 to 2009) and 19-25 in 2013/14. They find modest positive significant absolute impacts on secondary school completion (1-2 p.p. on a base of approximately 75 percent) that appear larger for women, but no significant impacts on enrollment or labor force participation for either gender.

For the analysis, they match an earlier poverty census in 2000 to the one implemented in 2013/14, using the national identification number of the adult woman in the household. As the authors acknowledge, this means the estimates are based on, and potentially only relevant for, the behavior of young adults who continued to live in the household they were in as children (and not for those who started their own household or moved into a different one). Moreover, if children leaving eligible households were different from those leaving ineligible households (e.g., as a result of transfer receipt), estimates could be biased. Finally, as for any RDD implemented within geographic areas, not only are the estimates most relevant for those households close to the eligibility cut-off threshold, but one needs to assume there were no substantial spillovers on ineligible households. These different caveats limit the generalizability of the findings.

2.2. Education-only CCTs

We next review long-term evidence of CCT programs that differ from the *PROGRESA* model in that they were targeted only to school-age children, and transfers conditioned only on schooling-related behaviors. This does not necessarily mean that long-term impacts will materialize only through an educational causal mechanism, however, as transfers may have affected beneficiaries through other channels as well. Notably, consideration of this type of CCT-program permits assessment of evidence outside of Latin America. We start with two national programs and then discuss two smaller pilots. As in section 2.1, we examine first the experimental and then the non-experimental evidence. The previous section detailed many of the principal types of criticisms of the evidence, and therefore in this section we make briefer reference to concerns that are repeated from above.

2.2.1 Cambodia

Filmer and Schady (2014) evaluate the Cambodian CESSP (Cambodia Education Sector Support Program) Scholarship Program that provided three-year scholarships upon graduation from elementary school. The program targeted 6th grade students (with median age 14) in 2005 in 100 lower secondary schools serving poor areas throughout the country. It offered the scholarships to the poorest students in those schools based on an initial composite “dropout-risk score” for grades 7, 8 and 9, conditional on enrollment and maintaining passing grades. A household survey was administered in 2010 to estimate the impacts five years after children started receiving the transfers, and thus two years after the transfers ended. The study uses a sharp RDD, with the dropout-risk score calculated at baseline as the running variable and cut-offs normalized by school, and includes school-level fixed effects. Attrition is relatively modest (14.1 percent) and, notably, is not discontinuous at the threshold and no adjustments for it are made.

The authors report substantial positive absolute program effects on grade attainment (0.6 years higher and similar for boys and girls). Consistent with these results, there are large positive effects on past school enrollment, retrospectively measured for the years preceding the household survey (8-20 p.p. increases for grades 7-10, an approximately one-third increase for the higher grades, i.e., in more recent years). However, there are few significant and much smaller RDD impacts on enrollment in the most recent year (corresponding to grade 11) and no significant impacts on test scores, the probability of paid or unpaid work, or monthly earnings.

This pattern of results suggests that at least for those on the margin of eligibility, the additional years of schooling induced by the scholarships did not translate into gains in learning or labor market productivity by age 19. Moreover, the program had no significant effects on subjective social status, mental health, marriage or fertility. Results are robust to a variety of RDD specifications.

As with other studies described above, the interpretation for the lack of labor market results is unclear, and the authors recognize it may be too early to observe the full returns to increased schooling. Possible learning or labor market returns also may be higher or lower for even poorer children, compared to those just at the threshold used for the RDD identification.

2.2.2 Pakistan

Alam, Baez and Del Carpio (2011) analyze the impacts of the Punjab Female School Stipend Program (FSSP) in Pakistan, a CCT program started in late 2003 to promote female middle school (grades 6-8) enrollment in public schools. The program first targeted 15 districts with the lowest literacy rates (below 40 percent) where girls enrolled in grades 6-8 were eligible, but has since expanded. The identification strategies for individual girls include a difference-in-difference estimation and a RDD model within a difference-in-difference framework, both estimating absolute program effects. Using cross-sectional household level data from 2003 and 2007/08, the approach compares outcomes between targeted and non-targeted districts, while the RDD exploits the literacy cut-off for program eligibility across districts. The analysis focuses on girls potentially exposed to the program for at least one year, based on age and grade level, between 2003 and 2008.

Results using both estimation approaches yield statistically significant increases of about 5 p.p. (about a 6 percent increase) for middle-school completion (grade 8) and transition to high school for the 15-16 year old cohort. There was also a reduction in labor force participation of 4-5 p.p. among girls ages 12-19. There was no evidence of FSSP effects on the probability of marriage.

The identification strategy has various limitations. Concerns regarding difference-in-differences estimates and RDD are similar to those discussed above for other non-experimental studies employing these methods. Moreover, in the case of the FSSP, the authors must rely on a proxy measure of the level of exposure to the program based on age, enrollment and grades attained from the cross-sectional household surveys, introducing potential measurement error with unknown bias into the estimates.

2.2.3 Malawi

Baird, McIntosh and Özler (2018) provide experimental evidence of the absolute program impacts of a two-year pilot program in the Zomba District of southern Malawi begun in 2008. The Schooling, Income and Health Risk study consisted of a cash transfer for never-married adolescent girls ages 13-22 and was stratified on baseline school enrollment. Two alternative designs were tested: a cash transfer conditional on school attendance and an unconditional cash transfer.[viii] Forty-six enumeration areas were randomized to the CCT and 88 to the control. A panel survey was conducted in 2012 to estimate the impacts on a broad range of outcomes more than two years after the program ended; attrition was 13-16 percent, and results are robust to inverse probability weighting for attrition.

For girls not enrolled in school at baseline, the experimental results show sustained absolute impacts on education—0.6 more grades attained—but no impact on a competencies test score. The girls were also 11 p.p. less likely to be married (compared to the control group mean of 81 percent), 4 p.p. less likely ever to have been pregnant (compared to 92 percent), and had significantly lower desired fertility than the control. For girls enrolled in school at baseline, the authors find few sustained effects. There were no effects on HIV infection, labor market or empowerment outcomes for either group of girls.[ix]

As the authors acknowledge, and similarly to other papers, the long-term labor market results are hard to interpret as the sample was of young adults and individuals in the treatment group were in school longer, and thus might have less work experience. Indeed, 28 percent were still in school at endline. Moreover, very few women in the sample have any work experience, suggesting that local labor market opportunities for young women are extremely limited, and hence, possibly, attrition could be correlated with work-related migration. In addition, in such settings, returns to education may operate through other channels, including the agricultural sector where income is harder to measure. That said, the paper stands out for the broad and comprehensive set of outcomes considered, and more will probably be learned from this innovative pilot program once longer-term follow-up data is available.

2.2.4 Colombia

The last education-only CCT we examine is another pilot program, *Subsidios*, implemented in two of the poorest localities in Bogotá in 2005 and evaluated using individual-level randomization (Barrera-Osorio, Linden and Saavedra 2017).[x] The program targeted conditional education transfers to secondary school students (age 14-16) from socio-economically disadvantaged backgrounds.

Three alternative program designs were tested. The first was the “basic treatment”, a fairly standard CCT in which transfers were made every other month, conditional on meeting a specified attendance target. The second was the "savings treatment", in which one third of the transfer payments was delayed until enrollment in the following school year. The third was a "tertiary-level 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 education institution if the individual enrolled in one, or; 2) one year later, if not. All those in the tertiary-level treatment who graduated from secondary school became eligible for this final transfer. The basic and savings treatments are compared to one randomized control group and the tertiary treatment to a different randomized control group (each only including students of the ages targeted by the respective treatments). The control groups never received the intervention, allowing the authors to evaluate the long-term absolute impacts with transfers continuing until the first of graduation from secondary school or dropout.

The authors combine program participation data with national administrative records on secondary school graduation exams (the same data source used by Baez and Camacho [2011]) after eight years, and administrative data on enrollment in tertiary institutions eight and twelve years after the start of the program in 2005. Matches were based on national student identification number, full name and date of birth. The percentages that matched correspond to expected graduation rates from secondary school and enrollment rates in tertiary institutions for the study population. The probability of an observation from the program participation data matching with national records was unrelated to baseline characteristics and did not differ between treatment and control groups.

There were no significant absolute long-term treatment impacts for students first exposed in lower secondary education (grades 6 to 8). Results for students in upper secondary (grades 9-11) at the time of first program exposure, however, indicate that only the savings treatment significantly increased the probability of taking the secondary school graduation exam, by 2.8 p.p. (about 3 percent), but differences between treatments are not significant. Both the savings and tertiary-level treatments led to higher ever enrollment in tertiary institutions after eight years (by about 10 and 20 percent for children in upper secondary at baseline, respectively) suggesting that savings constraints may have been a barrier for enrollment in tertiary education. Results after 12 years (hence at ages 26-28) confirm the positive and significant impacts for on-time tertiary enrollment, but no longer on ever having enrolled. This is consistent with higher tertiary enrollments rates in the control after 12 years, suggesting young adults are enrolling in tertiary education until late in their twenties. Tertiary graduation rates were about 2 p.p. higher (on a base of 10 percent) for the basic and savings treatments.

Barrera-Osorio, Linden and Saavedra (2017) is the only individually randomized study in this review, an important strength. In earlier work after just one year and using a more traditional survey, however, they report evidence of the possibility of negative intra-household spillover effects, a potential complication with such a design (Barrera et al. 2011). Overall, the long-term study provides a good example of the possibilities and the limits of using secondary administrative data to follow up on an earlier experimental evaluation. Match rates for the different data sources were relatively high (e.g., 70 percent for upper secondary school graduation exams) and largely unrelated to treatment arm. On the other hand, as with similar studies the set of outcomes that can be examined using such secondary data, and hence the possibility of understanding the different parts of the impact pathway, are naturally more limited. Regardless, the study demonstrates the advantage of having multiple rounds of longer-term follow-up data into mature adulthood in settings where final schooling outcomes only become clear after many years.

2.3. Summary of Long-term Impacts of Exposure to CCTs During School Ages

Table 2a summarizes the evidence on the long-term impacts of CCTs for children exposed in primary or secondary school. Taken together, the evidence indicates CCTs consistently help these children obtain higher grades, and often enable completion of higher levels of schooling. This conclusion holds for all countries considered, although the magnitude of the effects, as well as the particular levels of schooling affected, differ. Evidence on learning, available in only a subset of countries, is more mixed—with clear learning gains in Colombia (*Familias en Acción*) and Nicaragua, but no significant differential impacts in Mexico, nor long-term absolute effects in Cambodia or Malawi. Experimental evidence from Honduras and Colombia (*Subsidios*), and non-experimental evidence from Mexico further show substantial positive effects on secondary school completion and starting university, which could be indicative of learning gains.

These results indicate that CCT programs lead to improved human capital accumulation that goes beyond primary school education. Less conclusive is the extent to which these investments improve labor market and family-related outcomes or lead to higher lifetime earnings. Indeed some studies find negative effects on labor market participation, or on hours worked, but in most cases these appear to reflect higher continued school enrollment during young adulthood (and hence arguably could be interpreted as beneficial). Experimental evidence from Nicaragua shows positive differential impacts on labor earnings and on off-farm employment. Non-experimental evidence of impacts after 10 years or more in Mexico and Colombia (*Familias en Acción*) also suggests beneficial impacts on labor market outcomes. In contrast, experimental evidence from Mexico, Honduras and Malawi, and non-experimental evidence from Ecuador, Cambodia and Pakistan suggest no clear gains in the labor market.

An important caveat for the evidence of CCT long-term impacts on children during school-going ages is the high attrition rates in many studies. The difficulty of longitudinal tracking of individuals stems in part from the high mobility of these cohorts and complicates the estimation of unbiased treatment effects. The same caveat applies for most studies with administrative data, as they typically require making assumptions on migration patterns or other factors underlying incomplete matches across different data sources. Evidence on migration itself is quite limited, even if in many settings domestic migration is common and almost certainly related to economic opportunities. Evidence from Nicaragua shows positive effects on temporary migration related to labor market participation in off-farm activities, while experimental results for Mexico show negative effects on male migration. Evidence from Honduras further shows positive and relatively large effects on international migration, especially among young men. Finally, for a small subset of countries there is also evidence on impacts on marriage and fertility decisions for young women (Nicaragua and Malawi).

The limited learning and labor market returns found for some countries may well indicate that increased schooling alone will not be sufficient to sustainably improve livelihoods in contexts where many other factors are likely to constrain economic opportunity. That said, with the evidence at hand, it is arguably too soon to reach such a conclusion. Indeed, an important consideration pertains to the inherent difficulty of studying young adults who are still undergoing transition into the labor market and family formation. Many young adults in the reviewed studies are still pursuing schooling long after they turn 18, in part because of accumulated delays or interruptions in earlier schooling. In addition, because CCTs increase grades attained, beneficiaries will likely have less work experience than otherwise similar cohorts. This may reduce the net returns from CCTs measured during young adulthood, particularly if returns to work experience are diminishing (i.e., if returns to the first few years of work experience, that only those who have not continued to study have, are relatively high). The observed delayed entry in the labor market in some settings, therefore, implies that evaluations only reveal initial, or partial, information about the ultimate program effects on occupation or income. One possible resolution to this limitation, outside the scope of these specific studies, is to continue following the evaluation samples further into the future until all individuals have left school and fully entered the labor market, while at the same time making sure to keep attrition as low as possible. Multiple measurements, so that the time paths of program impacts can be traced out, would be even more informative.

3. Exposure to CCTs during Early Childhood and Outcomes at School Ages

3.1 Health and Education CCTs

Next we consider research on an earlier life-cycle transition, examining whether and how exposure to CCT programs in utero or under age 6 translates into better outcomes when children are ages 6-18. Evidence is available for all of the health and education CCTs described in section 2.1 (as well as the program in El Salvador) and for the most part the research designs are similar to those already described above. Despite this, the evidence base is thinner than for children exposed to CCTs at older ages, but possibly qualitatively stronger as some important concerns, such as attrition selection, are often more limited.

As with the older cohorts, rigorous evaluations on this younger age group have shown a number of positive short-term impacts, including on health care utilization and to a lesser extent, on anthropometrics (Fiszbein and Schady 2009). While these short-term impacts are indicative, as for the older children they fall short of providing definitive evidence on the more lasting changes that are the ultimate objective of CCTs. For example, does exposure to CCTs in early childhood lead to improvements in anthropometrics, cognition, learning or education at school ages? A complicating feature for analyses of this age group is that for ongoing programs the children remain potential current beneficiaries.

3.1.1 Mexico

Using the same data and similar approaches for their analyses of the older cohort, Behrman, Parker and Todd (2009b) examine children ages 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 *PROGRESA* experiment to obtain differential program impacts, and also difference-in-difference non-experimental matching methodologies to estimate absolute program effects. Because all but the oldest of these children do not have meaningful baseline schooling outcomes, they modify the difference-in-difference estimator to control for the outcomes of other children in the community who were 6-14 at baseline. Attrition for the targeted age group is lower than for the older cohort, at 20 percent, and similarly addressed through reweighting. 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. The results further show that the 18-month differential exposure to *PROGRESA* did not significantly affect on-time (for age) grade progression for children ages 9-11 in 2003. 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 indicate positive and significant absolute improvements in progression rates of about 15 percent for boys and 7 percent for girls.

The experimental differential results do not show robust evidence for grade completion, but the matching difference-in-difference estimates suggest girls ages 9-11 in 2003 accumulated about 0.3 grades and boys about 0.4 grades more than non-beneficiary peers (effects for children ages 6-8 are not significant). Overall for education, the findings indicate limited experimental differential effects, but more positive results for the non-experimental absolute effects. This pattern may reflect the relatively short (18-month) differential being exploited for these children currently still eligible after six years. The assumptions underlying both approaches are similar to those discussed earlier, 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.

Fernald, Gertler and Neufield (2009) investigate the effect of *PROGRESA* on anthropometrics, cognition, language and behavior, 10 years after the start of the program. Similar to Behrman, Parker and Todd (2009a, 2009b, 2011), they exploit the 18-month differential exposure between the experimental early- and late-treatment localities. 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 the follow-up survey. Attrition in the study is approximately 40 percent.

They find a significant differential reduction in behavioral problems, but no significant impacts on child growth, body mass index, cognition or language when using the standard single-difference experimental approach. They also present an alternative non-experimental estimator that uses 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) differences in household composition and grade achievements of eligible children at baseline, since transfer amounts are tied to gender, age and grade level. Actual cumulative cash transfers received differ further for a third reason: they depend directly on schooling decisions made in the household. The authors report a negative association between cumulative cash transfers received and the number of reported behavioral problems, consistent with the findings from the experimental evaluation. In addition, they find that higher cumulative cash transfers are significantly and positively associated with height-for-age z-scores and higher verbal and cognitive test scores. The results largely hold when actual cumulative transfers are endogenized using potential cash transfers (ignoring actual schooling decisions) as an instrumental variable (Fernald, Gertler and Neufeld 2010).

Because the cumulative cash transfers, even when instrumented, depend on household structure as well as the randomized assignment, the interpretation and internal validity of these results have been questioned (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. Those demographics, of course, are not randomly assigned and might well affect cognitive and anthropometric outcomes in their own right.

3.1.2 Colombia

García et al. (2012) report difference-in-difference non-experimental evidence of five years of differential exposure to *Familias en Acción* on nutrition and health outcomes for children ages 0-6 at baseline (in 2002) after 10 years. This is complemented by RDD estimates, exploiting variation in assignment to treatment arising from the discontinuous rule that determined eligibility for the program to obtain absolute program impact estimates. The RDD estimates compare children ages 3-11 years old at baseline (13-21 at follow-up) whose SISBEN score (in 1999) were just above and below the eligibility threshold.

Using the difference-in-difference approach, the study finds positive and significant impacts of differential exposure during the first five years of life on anthropometric measures. In particular, for children 0-3 years old in 2002, the height-for-age z-scores increased by 0.21 SD in rural areas and by 0.16 SD in rural and urban areas combined. The treated children in this cohort are compared to children who only became eligible when they were 5-8 years old. This positive impact corresponds to 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. 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 in section 2.1.2 indicate that there were no impacts on the Raven (a cognitive test), but large impacts on the mathematics test for adolescents ages 12-17 in 2012, i.e., for children exposed to the program in early childhood (ages 2-7 at baseline). The RDD results for children ages 3-11 in 2012 are consistent with that possibility. They show modest and marginally significant impacts on cognition around the threshold, with an increase in the Picture Peabody Vocabulary Test of receptive vocabulary (TVIP for its acronym in Spanish) score of 0.09 SD. As described earlier, however, a drawback of the RDD approach using the SISBEN is that it 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 necessarily isolate the impact of *Familias en Acción*. The RDD approach also does not allow separating the 10-year cumulative effect of the transfers from any short-term contemporaneous effect related to current conditionalities or transfers. Last, as for any RDD, impact estimates may only be relevant for households close to the threshold, as the authors acknowledge.

3.1.3 Nicaragua

Using the randomized rollout of Nicaragua's *RPS*, Barham, Macours and Maluccio (2013) analyze differential program impacts for boys exposed in utero and during the first two years of life, as compared to those exposed outside of this potentially critical 1000-day window. Ten years after the start of the CCT, cognitive tests were administered to a cohort of children born in the first 6 months of program operations. The tests measured processing speed, memory, receptive vocabulary (via the TVIP) and executive functioning. In addition, height and weight also were measured. All interviews and tests took place in their homes, regardless of schooling status, and attrition was 6 percent. Results for girls are not reported.

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 SD higher for the early-treatment group. At the same time, the analysis shows no significant differential impact on anthropometric measures, despite evidence of positive short-term absolute effects. Together, the results suggest complete catchup for boys in the late-treatment communities for physical, but not for cognitive, outcomes.

While the experimental results require relatively few assumptions, they are differential results and there might have been persistent absolute impacts for outcomes other than cognition for boys (where the significant differential effects 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. For example, both treatment groups may have experienced similarly sized improvements that canceled each other out in the differential, or alternatively, the early-treatment group may have experienced a large short-term gain that faded out in the long term such that it was equal to any long-term gains experienced in the late-treatment group. The authors show evidence of short-term gains in anthropometrics for both the early- and late-treatment groups suggesting that a pattern of positive absolute impacts cancelling each other out is most likely. Further exploration of the underlying mechanisms or other intermediate outcomes, as well as results for girls, is needed.

3.1.4. Honduras

Molina Millán et al. (2018) use the municipal-level randomized assignment of the 5-year *PRAF-II* program in Honduras discussed earlier, combined with individual-level data from the population census 13 years after the start of the program, to analyze long-term absolute impacts for the cohorts exposed to the nutrition and health components of the program during early childhood. As the program targeted the municipalities with the highest levels of malnutrition in the country, the outcomes on the cohorts exposed during early childhood, who hence benefitted at an age in which transfers and the nutritional components of the program may be the most important, are particularly relevant.

For the non-indigenous boys and girls, program exposure starting in utero or during the first 2 years of life leads to increases of about 0.4 grades attained 13 years later. They were also 6-8 p.p. more likely to have completed primary school than the control group (an approximately 25 percent increase for the youngest). While these results are significant and robust, with a five-year program exposure arguably they are of modest size given. Results for the indigenous are smaller and not significant.

The reliance on national census data, together with very low levels of international migration for the relevant age groups, mitigates concerns about attrition as with other studies using such data; however, the outcomes that can be studied for this age group are limited, and the long-term effects on anthropometrics, cognitive outcomes and learning are unknown.

3.1.5. Ecuador

Araujo, Bosch and Schady (2018) present experimental evidence of differential program impacts 10 years after BDH began, comparing children from the 51 parishes randomly selected to receive transfers starting in 2004 when they were under age 6 with those from the 26 parishes randomly selected to become eligible three years later. As indicated earlier, BDH provided cash transfers but conditionalities were never enforced. Achievement, cognitive and behavioral tests were administered to children in 2014 within the 10-year household panel survey carried out for the evaluation. The authors present results both for the full sample and for different wealth quartiles. While the program led to short-term gains in early childhood outcomes for the lowest quartile of the population in rural areas, no differences are found for language, math or an index of cognitive tests and behavioral outcomes after 10 years. Differential impacts on the full sample of children are not significant in the short or long run.

The attrition rate in the panel is 19 percent, but the endline sample is balanced on observables. Attrition does not appear to be driving the differences between the short- and long-term results as restricting the short-term sample to those found after 10 years gives broadly similar results as those for the full sample. Overall, the long-term results are consistent with a fading out of the short-term impacts, but could also indicate catchup among those that started later. This suggests that exposure to unconditional cash transfers during very early childhood compared to later on in childhood does not necessarily provide children with permanent long-term advantages. As conditionalities in BDH were never enforced despite the initial program design, however, we note that these findings are less directly comparable to the rest of the literature reviewed.

3.1.6. El Salvador

Sanchez Chico et al. (2018) use a non-experimental identification strategy akin to RDD to analyze the impact of the ongoing CCT in El Salvador, *Comunidades Solidarias Rurales*, on early school entry after six years of program exposure. The Salvadoran CCT was rolled out in the 100 poorest municipalities of the country between 2005 and 2010. Program eligibility was determined at the time of the initial program census, and households who conceived their first child after the program began in their municipality were not eligible for the program until recertification several years later. The authors compare firstborn eligible and ineligible children with eligibility based on when the initial census was conducted in their municipality. Using data from a new program census conducted in 2013, they estimate single-difference effects for the children born between 2006 and 2007 controlling for month-of-age and municipality (or finer) geographic level fixed-effects. Five-year old children with about six years of program exposure (including in utero) are 12 p.p. more likely to be attending preschool and 9 p.p. more likely to have completed at least one year of school (including preschool), both reflecting approximately 25 percent increases over the control mean. For 6 year olds, similar estimates show an increase of 7 p.p. (a near doubling) of the probability of having completed at least one year of primary school, but no effects on school attendance.

As sample sizes are relatively small, identification necessarily relies on parametric assumptions underlying the age controls, an important caveat of the study given the age-dependent nature of the outcomes examined. Placebo results using non-first-born children of similar ages do, however, provide some support for those assumptions. Identification also relies on assumptions of limited differential attrition, which is partly supported by evidence from the national population census, showing 6 percent outmigration for the relevant age group. Finally, as the authors acknowledge, the six-year cumulative effect of the transfers cannot be separated from any short-term contemporaneous effect related to current conditionalities or transfers, as is also the case in other studies of on-going programs using similar research designs. This complicates interpretation of the results.

3.2. Summary of Long-term Impacts of Exposure to CCTs During Early Childhood

Table 2b summarizes the evidence on the long-term impacts of CCTs for children exposed in utero or early childhood. Overall, the evidence base for exposure in early childhood is more limited than for exposure during school-going ages. Fully understanding the long-term effects of early childhood exposure to CCTs will also require much longer-term follow-up, as the returns to nutrition and health gains arguably may only fully materialize once these children have grown into adults. That said, the evidence from these studies still provides tentative lessons regarding impacts in the next phase of these children's lives at primary-school ages. They are tentative both because of the caveats described regarding the evidence as well as because of the fact that, as Section 2 demonstrates, much lies ahead in the development of these children.

Several experimental differential studies suggest fadeout of impacts and/or catchup of the control groups that received similar benefits a little bit later in life (cognitive outcomes in Mexico and Ecuador; anthropometrics in Mexico and Nicaragua). In those cases, it is an open question whether differential impacts will re-emerge later in life. Other experimental studies, where either the control group never received the program or differential exposure between the early- and late-treatment groups is large, show positive long-term effects on educational and cognitive outcomes 10 to 13 years later (Honduras, Nicaragua). Non-experimental estimates also show positive long-term effects on education (Mexico, Colombia, El Salvador). In part, such impacts may come from earlier enrollment in preschool or primary schools. Where programs are ongoing and differences in eligibility persist when children reach the age of school entry, it is not possible to disentangle whether the estimated impacts are driven by the cumulative exposure to the CCT since early childhood, transfers and/or the start of the schooling conditionality when children reach school age, or a combination of both (Mexico, Colombia, El Salvador). But evidence from programs that were no longer operating by the time children reached school-going ages (Honduras, Nicaragua), allow isolating the impact of early childhood exposure and also show long-term gains. Overall, this suggests that the early childhood components of CCTs are important to consider when evaluating their impacts on educational and other outcomes later in life. As for the older cohort, and maybe even more so, continuing to follow these younger cohorts, while paying attention to keeping attrition to a minimum, will be important.

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 assess all of the possible long-term implications of these programs (for example, whether CCTs succeed in breaking the intergenerational transmission of poverty), after two decades of experience, evidence on important long-term impacts has begun to accumulate.

In this review, we defined long-term impacts as those that materialize across two stages of the life-cycle. The first is from childhood/adolescence to young adulthood; the focus in this case is on educational, 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. The second transition 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.

For both transitions, we reviewed the evidence and highlighted various strengths and limitations of the available experimental and non-experimental studies. The research employing non-experimental methods allows studying the long-term impacts even when CCT programs did not embed an experimental impact evaluation in their initial design or rollout. The credibility of such non-experimental results, however, is hindered by the difficulties inherent in constructing a valid counterfactual, 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 because 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, and therefore may have limited statistical power. For both the experimental and non-experimental evidence, sample attrition (likely to be related to migration, itself an outcome of interest) is an important source of concern, particularly for the older cohorts.

The existing evidence on CCTs long-term impacts is clearer for some than for other outcomes, as summarized in sections 2.3 and 3.7. The experimental literature provides consistent evidence of impacts on schooling, as well as some evidence of impacts on cognitive skills and learning, socioemotional skills and improved labor market outcomes. Yet many studies also find a fair share of results that are not statistically different from zero. Unsurprisingly, it is often difficult to discern whether this is due to lack of impact or other methodological concerns. The non-experimental literature provides a similarly mixed picture, along with greater concerns about internal validity.

This review also highlights a different type of challenge affecting the interpretation of many recent studies. In most cases, individuals have yet to fully transition into the labor market, even when they are observed a number of years after the start of the intervention. Indeed, as several studies show, many young adults are still transitioning out of school until late in their twenties. As final educational outcomes can only be observed after this transition is finished, the interpretation of labor market impacts is complicated by the inherent tradeoff between additional schooling, labor market choices and shorter work experience. Additional interactions with fertility and marriage market decisions further suggest that assessment of the long-term returns to CCTs needs to account for different potential pathways—this may be particularly relevant for women.

The overall evidence to date for the younger cohorts suggests that the early childhood components of CCTs are important to consider when evaluating impacts on educational and other outcomes later in life. For this cohort, however, a better understanding of the potential for fadeout or increased impacts over time is needed. This includes understanding how early-life impacts interact with ongoing program interventions later on. Relatedly, it includes disentangling whether effects are driven by cumulative or current exposure to the CCT, or both.

Some methodological lessons for assessing the long-term impacts of a next generation of programs aimed at human capital accumulation follow. Estimating the full benefits of such interventions necessarily requires a decades-long horizon. Where possible, evaluation designs should plan *ex ante* for establishing control groups that can serve as credible counterfactuals for that horizon. This might involve strategies for minimizing attrition, for instance through continuous follow-up (as demonstrated recently by Duflo, Dupas and Kremer 2017). Where experimental control groups cannot be excluded from the program indefinitely, it may also require planning for both experimental and non-experimental control groups identified at baseline (for instance through RDD or matching). This could help validate the non-experimental control group (for example against short-term experimental results), and allow more rigorous long-term assessment once the experimental control is phased into the program.

To conclude, our interpretation of the accumulating evidence is that while there is robust evidence on some important long-term impacts (in particular on schooling), there are a number of important unknowns central to the CCT objectives, in particular in the labor market. 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 mature adulthood, additional opportunities to examine the more "permanent" returns to human capital become available. Uncovering such opportunities, and developing new strategies going forward 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

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

Alam, A., Baez, J. E., & Del Carpio, X. V. (2011). “Does Cash for School Influence Young Women's Behavior in the Longer Term? Evidence from Pakistan.” *World Bank Policy Research Paper* 5669. Washington, DC, United States: The World Bank.

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

Attanasio, O., Gómez, L. C., Heredia, P. and Vera-Hernández, M. (2005). “The Short-Term Impact of a Conditional Cash Subsidy on Child Health and Nutrition in Colombia. Report Summary: Familias 03.” London, UK: Institute for Fiscal Studies.

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., Ferreira, F.H.G., Özler, B., and Woolcock, M. (2014) “Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes”, *Journal of Development Effectiveness* 6(1): 1-43

Baird, S., McIntosh, C., and Özler, B. (2018). “When the Money Runs Out: Do Cash Transfers Have Sustained Effects?” Mimeo, George Washington University.

Barham, T. (2011). “A Healthier Start: The Effect of Conditional Cash Transfers on Neonatal and Infant Mortality in Rural Mexico.” *Journal of Development Economics* 94(1): 74-85.

Barham, T., Macours, K., and Maluccio, J.A. (2013). “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. (2017). “Are Conditional Cash Transfers Fulfilling their Promise? Schooling, Learning and Earnings After 10 Years.” CEPR Discussion Paper 11937.

Barham, T., Macours, K., and Maluccio, J.A. (2018). “Young Women’s Labor Market and Fertility Outcomes 10 years after the start of a CCT.” Mimeo, University of Colorado-Boulder.

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. (2017). “Medium- and Long-Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia.” *American Economic Journal: Applied Economics,* forthcoming.

Barrientos, A. (2018). Social Assistance in Low and Middle Income Countries Database (SALMICs). Global Development Institute: Manchester UK.

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 (eds.) *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.

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

Duque, V., Rosales, M. F., and Sanchez, F., (2018). “How Do Early-Life Shocks Interact with Subsequent Human-Capital Investments? Evidence from Administrative Data.” Mimeo, University of Michigan.

Fernald, L.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., Gertler, P.J. and Neufeld, L.M. (2010). “Author’s Reply”. *The Lancet* 375(9719): 981.

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.

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

Fitzgerald, J.M., Gottschalk, P. and Moffitt, R.A. (1998). “An analysis of sample attrition in panel data.” *Journal of Human Resources*, vol. 33(2): 251–99.

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

Glewwe, P., and K. Muralidharan, K. (2015). “Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications.” In *Handbook of the Economics of Education*, 5: 653-743. Elsevier.

Hahn, J., Smyth, R., Islam, A., Yang, H.S., and Nuzhat, K. (2018). “Education, Marriage, and Fertility: Long-term Evidence from a Female Stipend Program in Bangladesh.” *Economic Development and Cultural Change* 66(2): 383-415

Ham, A. and Michelson, H., (2018). “Does the form of delivering incentives in conditional cash transfers matter over a decade later?” *Journal of Development Economics* 134: 96-108.

Ibarrarán, P., N. Medellín, F. Regalia and M. Stampini, (eds.). (2017). *How Conditional Cash Transfers Work: Good Practices after Twenty Years of Implementation*. Washington, DC, United States: Inter-American Development Bank.

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

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

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

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

Molina Millán, T. and Macours, K. (2017). “Attrition in Randomized Control Trials: Using Tracking Information to Correct Bias” CEPR Discussion Paper 11962.

Molina Millán, T., Macours, K., Maluccio, J., and Tejerina, L. (2018) “Experimental long-term impacts of early childhood and school age exposure to a conditional cash transfer.” IDB Working Paper Series.

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

Parker, S. and Todd, P. (2017). “Conditional Cash Transfers: The case of Progresa/Oportunidades.” *Journal of Economic Literature* 55(3): 866-915.

Parker, S., Rubalcava, L., and 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.” Mimeo, CIDE.

Parker, S. and Vogl, T. (2018). “Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico.” *NBER Working Paper* No. 24303.

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

Sanchez Chico, A., Macours, K., Maluccio, J., Stampini. M. (2018). “Six years of *Comunidades Solidarias Rurales*: Impacts on School Entry of an Ongoing Conditional Cash Transfer Program in El Salvador.” IDB Working Paper Series 908.

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).

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 no. 185.

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

6. Methodology Appendix: Search Methods and Inclusion Criteria for Studies in the Review

Much of the relevant literature on the long-term effects of CCT programs in developing countries is unpublished or appears only in report form. Consequently, the identification of potential CCTs and related studies for inclusion in the review required several steps. We began by identifying the CCT programs around the world, starting from the comprehensive list of programs initiated before 2009 provided in Appendix A of Fiszbein and Schady (2009), complemented by a more recent inventory provided by Barrientos (2018). This exercise yielded a potential set of CCTs and their main design features. To identify the body of research corresponding to these programs, we conducted a systematic search of CCT evaluation studies contained in the International Initiative for Impact Evaluation (3ie) repository of impact evaluations (available at http://www.3ieimpact.org) and conducted an internet search for any additional studies evaluating long-term impacts.

For greater comparability (and to permit more robust conclusions), we focused on CCTs with program designs broadly similar to *PROGRESA* in Mexico, which we treat as the benchmark program. However, as the double requirement of early childhood (nutrition and health) *and* school-age (education) conditions would only yield programs in Latin America, we also included school-age targeted, education-only CCTs. We analyze the two program types separately.

Specifically, an *individual CCT* was eligible for inclusion in the review if it fulfilled the following two criteria:

1. *The program was conditional.* We categorized a program as conditional if it had one or more conditions explicitly related to nutrition, health or schooling (typically preventive health visits, attendance at information sessions on nutrition and health; school enrollment and attendance). But we excluded programs with conditions primarily related to outcomes, such as school test performance, to distinguish between CCTs and academic merit scholarship programs.[xi] Unconditional cash transfer (UCT) programs that do not impose conditionalities were excluded under this criterion.

2. *The program encouraged human capital investment in children.* We only considered CCT programs whose goals included directly improving human capital investment in the nutrition, health or education of children. Therefore, we focused on those interventions covering young children or school-age children.

Having thus identified the eligible set of similar (though by no means identical) CCTs, criteria for inclusion in the review of an *individual study* (evaluating an eligible CCT) focused on the timing of exposure and ages at which outcomes were examined. A study was included if it fulfilled at least one of the following criteria:

1. The study examined whether and how exposure to the CCT in utero or during early childhood (e.g. under age 6) affected outcomes at primary-school ages or older. These generally included impacts on nutrition, health or schooling.

2. The study examined whether and how exposure to the CCT during school ages affected outcomes in young adulthood (approximately over age 18). These generally included impacts on schooling, learning, labor market outcomes and marriage markets.

While not explicit in the above criteria, other important aspects of the methodology warrant emphasis. There was no exclusion related to research design, for example we included both experimental and non-experimental designs, nor on publication status. The duration of CCT exposure or the length of time since the CCT exposure also were not explicit criteria for selecting studies, but instead we focused explicitly on impacts measured during a later phase in children's life cycle compared to when program exposure had begun.

Table 1. Overview of countries with CCT programs included

Country	GDP PC PPP (2016)[1]	HDI (2015)[2]	Mean years of schooling (2015)[2]	CCT Program	Program coverage	Type of evidence	Years since program start	Years after program end	Program Components
Mexico	17,275	0.762	8.6	*PROGRESA*	national	Exp & Nonexp	4/6/8/10/13	ongoing	Health & Education
Colombia	14,154	0.727	7.6	*Familias en Acción*	national	Nonexp	9/10/12	ongoing	Health & Education
Nicaragua	5,540	0.645	6.5	*Red de Protección Social*	regional	Exp & Nonexp	10	5	Health & Education
Honduras	4,737	0.625	6.2	*PRAF-II*	regional	Exp	13	8	Health & Education
Ecuador[3]	11,242	0.739	8.3	*Bono de Desarollo Humano*	national	Exp & Nonexp	10	ongoing	Health & Education
El Salvador	8,617	0.680	6.5	*Comunidades Solidarias Rurales*	regional	Nonexp	6	ongoing	Health & Education
Cambodia	3,737	0.563	4.7	CESSP Scholarship Program	regional	Nonexp	5	2	Education
Pakistan	5,235	0.550	5.1	Punjab Female School Stipend Program	regional	Nonexp	5	ongoing	Education
Malawi	1,169	0.476	4.4	Schooling, Income and Health Risk Study	pilot	Exp	4	2	Education
Colombia	14,154	0.727	7.6	*Subsidios*	pilot	Exp	8/12	11	Education

Notes: 1. 2016 World Development Indicators, World Bank. 2. 2016 Human Development Index, Human Development Report, UNDP. 3. Conditionalities not enforced. Exp = experimental; Nonexp = non-experimental.

Table 2a. Overview of impacts on outcomes in young adulthood from exposure during school ages

Section of this paper	Country	CCT Program	Population	Impact estimate	Higher levels of schooling	Learning	Labor force participation	Income
2.1.1	Mexico	*PROGRESA/Oportunidades*	F/M	Exp Differential	+/+	0/0	0/?	n.a.
			F/M	Nonexp Absolute	+/+	n.a.	+/?	+/0
			F/M	Nonexp Differential	+/+	n.a.	+/+	0/0
2.1.2	Colombia	*Familias en Acción*	F/M	Nonexp Absolute	+/+	?/?	n.a.	n.a.
			F/M	Nonexp Differential	+/+	+/+	+/0	n.a.
2.1.3	Nicaragua	*Red de Protección Social*	F/M	Exp Differential	0/+	0/+	+/+	+/+
			F/M	Nonexp Absolute	+/+	+/+	0/0	0/0
2.1.4	Honduras	*PRAF-II*	F/M	Exp Absolute	+/+	n.a.	?/0	0/0
2.1.5	Ecuador[1]	*Bono de Desarollo Humano*	F/M	Nonexp Absolute	+/+	n.a.	0/0	n.a.
2.2.1	Cambodia	CESSP Scholarship Program	F/M	Nonexp Absolute	+/+	0/0	0/0	0/0
2.2.2	Pakistan	Punjab Female School Stipend Program	F	Nonexp Absolute	+	n.a.	-	n.a.
2.2.3	Malawi	Schooling, Income and Health Risk Study	F	Exp Absolute	+	0	0	0
2.2.4	Colombia	*Subsidios*	F/M	Exp Absolute	+/+	n.a.	n.a.	n.a.

Notes: See corresponding section for descriptions of results, some of which only pertain to specific subpopulations. Results that studies do not report separately by gender were assumed to be the same for females (F) and males (M). 0 indicates no significant impact. ? indicates mixed results. - or + indicate negative or positive results. n.a. indicates no information available. 1. Conditionalities not enforced. Exp = experimental; Nonexp = non-experimental.

Table 2b. Overview of impacts on outcomes at school ages from exposure during early childhood

Section of this paper	Country	CCT Program	Population	Impact estimate	Anthropometrics	Cognitive development	Socio-emotional	Schooling	Learning
3.1.1	Mexico	*PROGRESA/Oportunidades*	F/M	Exp Differential	0/0	0/0	+/+	0/0	n.a.
			F/M	Nonexp Absolute	n.a.	n.a.	n.a.	+/+	n.a.
			F/M	Nonexp Differential	+/+	+/+	+/+	n.a.	n.a.
3.1.2	Colombia	*Familias en Acción*	F/M	Nonexp Absolute	n.a.	+/+	n.a.	n.a.	n.a.
			F/M	Nonexp Differential	+/+	0/0	n.a.	+/+	+/+
3.1.3	Nicaragua	*Red de Protección Social*	M	Exp Differential	0	+	n.a.	n.a.	n.a.
3.1.4	Honduras	*PRAF-II*	F/M	Exp Absolute	n.a.	n.a.	n.a.	+/+	n.a.
3.1.5	Ecuador[1]	*Bono de Desarollo Humano*	F/M	Exp Absolute	n.a.	0/0	0/0	n.a.	0/0
3.1.6	El Salvador	*Comunidades Solidarias Rurales*	F/M	Nonexp Absolute	n.a.	n.a.	n.a.	+/+	n.a.

Notes: See corresponding section for descriptions of results, some of which only pertain to specific subpopulations. Results that studies do not report separately by gender were assumed to be the same for females (F) and males (M). 0 indicates no significant impact. ? indicates mixed results. - or + indicate negative or positive results. n.a. indicates no information available. 1. Conditionalities not enforced. Exp = experimental; Nonexp = non-experimental.

[i] Teresa Molina Millán 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 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 Caridad Araujo, Pedro Cueva, Pablo Ibarrarán, Nadin Medellín, Ferdinando Regalia, Norbert Schady, participants in an IDB seminar, anonymous reviewers and the editor for valuable 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.

[ii] Fiszbein and Schady (2009); Stampini and Tornarolli (2012); Levy and Schady (2013); and Ibarrarán et al. (2017).

[iii] Other recent work examining short-term educational outcomes of CCTs includes reviews by Murnane and Ganimian (2014) and Glewwe and Muralidharan (2015) and meta-analyses by Saavedra and García (2012), Baird et al. (2014) and McEwan (2015). Parker and Todd (2017) provide a comprehensive review of the impacts of the nutrition, health and educational components of the Mexican CCT.

[iv] There is, however, 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 CCT *PROGRESA* made investments that led to consumption gains beyond those associated with the ongoing transfers more than five years after the start of the program. Also for *PROGRESA*, Parker and Vogl (2018) find improvements in household-level asset indices after 13 years.

[v] Although it changed names (first to *OPORTUNIDADES* then to *PROSPERA*), as the long-term evidence described in this review relates back to the program started in 1997, we refer to it throughout by its original name *PROGRESA*.

[vi] It is primarily for this reason that we do not summarize results from two other unpublished studies that focus on subsequent rounds for which attrition is even higher. Rodríguez-Oreggia and Freije (2012) use the subsequent round 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 they consider, so that both the internal and external validity of the study appear to be weak. Kugler and Rojas (2018) also use data from the 2003 and 2007 rounds, and for some outcomes complement this with data from administrative program recertification surveys through 2015. The latter surveys only provide information for individuals still living in original baseline households and only if those households are still in the program. Reported attrition rates are very high for the later years (95 percent and higher) and unbalanced in several years, which jeopardizes both internal and external validity. Their empirical strategy compares outcomes of teenagers and young adults with different lengths of exposure, and the identifying assumption is that younger and older cohorts would have similar education and employment profiles in the absence of the CCT. Given that ages span 10 years (between 15-24 years old when outcomes are observed), this assumption is difficult to justify and it is possible that estimated effects partly reflect age effects rather than length of exposure, even for the earlier 2003 round with lower attrition.

[vii] For example, Barham (2011) uses the rollout of *PROGRESA* at the municipality level between 1997-2000 to examine the effects of the program on infant mortality, and demonstrates that pre-program municipality characteristics differ between municipalities phased in during even this three-year period.

[viii] We do not consider the effects of the unconditional cash transfers since they fall outside our inclusion criteria.

[ix] The paper also analyzes impacts on marriage quality (husband characteristics) and children's nutritional status for the subset of women that is married or had children by the time of the follow-up survey. But, as the paper indicates, given selection into both marriage and fertility, longer-term follow-up may be needed to fully understand the sustained impacts on such outcomes.

[x] 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).

[xi] For example, because of its inclusion of both test score requirements as well as a condition that beneficiaries remained unmarried, we exclude the Female Secondary Education Stipend Programme (FESP) in Bangladesh (Hahn et al. 2018).