

Cumulative Impacts of Conditional Cash Transfer Programs: Experimental Evidence from Indonesia

By Nur Cahyadi, Rema Hanna, Benjamin A. Olken,
Rizal Adi Prima, Elan Satriawan and Ekki Syamsulhakim*

September 2019

Abstract

Conditional cash transfers provide income and promote human capital investments. Yet, evaluating their longitudinal impacts is hard, as most experimental evaluations treat control locations after a few years. We examine such impacts in Indonesia after six years, using an experimental rollout in which the experiment remained largely intact. We find static effects on many targeted indicators: childbirth using trained professionals increased dramatically, and under-15 children not in school fell by half. We observe impacts requiring cumulative investments: stunting fell by 23 percent. While human capital accumulation increased, the transfers did not lead to transformative economic change for recipient households.

* Cahyadi: National Team for Acceleration of Poverty Reduction (TNP2K), Grand Kebon Sirih Building, Jl Kebon Sirih no 35, Jakarta 10110, Indonesia (email: nur.cahyadi@tnp2k.go.id); Hanna: Harvard Kennedy School, 79 John F. Kennedy Street, Cambridge, MA, 02138 (email: rema_hanna@hks.harvard.edu); Olken: Department of Economics, MIT, The Morris and Sophie Chang Building, 50 Memorial Drive, Cambridge, MA, 02141 (email: bolken@mit.edu); Prima: Royal Melbourne Institute of Technology, School of Economics, Finance and Marketing, 445 Swanston St, Melbourne VIC 3000, Australia (email: rizal.prima@rmit.edu.au); Satriawan: National Team for Acceleration of Poverty Reduction (TNP2K), Grand Kebon Sirih Building, Jl Kebon Sirih no 35, Jakarta 10110, Indonesia (email: esatriawan@ugm.ac.id); Syamsulhakim: National Team for Acceleration of Poverty Reduction (TNP2K), Grand Kebon Sirih Building, Jl Kebon Sirih no 35, Jakarta 10110, Indonesia (email: ekki.syamsulhakim@fe.unpad.ac.id). We thank Aaron Berman for research assistance and Harsa Kunthara for help in the early stages of data analysis. We also thank Berk Özler, Alessandra Voena, Chris Blattman and David McKenzie for helpful comments and suggestions. Financial assistance for this project came from PNPM Support Facility (PSF), Poverty Reduction Support Facility (PRSF), and Mahkota, all supported by the Australian Department of Foreign Affairs and Trade.

I. INTRODUCTION

Perhaps the most remarkable innovation in welfare programs in developing countries over the past few decades has been the invention and spread of conditional cash transfer programs (CCTs). These programs provide regular cash transfers to poor households to help reduce poverty, but condition the transfers on households making a series of human capital investments in their young children. These conditions typically begin before birth – pre-natal care and deliveries by trained midwives or doctors are usually conditions – and continue through early childhood health investments (for example, immunizations and growth monitoring) and enrollment in primary and junior secondary school. These programs began in the 1990s in Mexico, Bangladesh and Brazil, and today over 63 countries have at least one CCT program (Bastagli et al. 2016), covering millions of families worldwide (Robles, Rubio, and Stampini 2015; World Bank 2018).

The theory behind these conditions – and the reason they start before birth and continue throughout childhood – is that static investments in human capital at every stage of the life cycle will accumulate as children grow, and the cumulative investments in human capital will eventually lead to improvements in child outcomes that may affect inter-generational poverty. For example, Santiago Levy, who helped create the CCT model with Mexico's PROGRESA program in the 1990s, argued, “clearly, achieving good health is a cumulative process, and temporary investments in nutrition are of little help. The same is true of education: children must be supported year after year.... [PROGRESA's] central effects will gradually occur through the accumulation of human capital” (Levy 2006).

Given the worldwide scope of CCT programs, there has been substantial interest in understanding whether static CCT conditions actually lead to cumulative improvements in child outcomes, but it is empirically challenging to answer this question. Many CCTs, starting with PROGRESA, began with randomized controlled trials on a pilot basis prior to scale-up, and the vast majority of the evidence on their impacts comes from these trials (for example, see Behrman and Todd 1999; Gertler 2004). However, most of these programs extended the CCT to the control group after a relatively short pilot period – 18 months

in the PROGRESA case, for example.¹ While this phase-in experimental design is useful for studying the short-run, static impacts of the CCT on the health and education behaviors they incentivize, the fact that the control group is ultimately treated makes it much harder to credibly estimate *cumulative effects* from sustained exposure to the programs over time²

A second, related question is whether these government welfare programs themselves continue to be effective – even in the static sense of maintaining increased compliance with incentivized behaviors. Some have argued that interventions are often less effective when implemented by the government at scale than in a smaller pilot stage, when researchers are more likely to be paying attention to the implementation (see, for example, Bold et al. 2015; and the related discussions in Banerjee et al. 2017 and Muralidharan and Niehaus 2017). More generally, a CCT program’s effects could weaken over time after people’s initial excitement of being in the program fade, or once people learn that the conditions placed on health and education behaviors are not always perfectly enforced. Since most CCT experiments extend the program to the control group after a relatively short time, understanding whether the programs continue to be effective even in a static sense after a short experimental initial period is also challenging.

This study aims to answer these questions using an unusual, large-scale policy experiment. We study Indonesia’s conditional cash transfer program, known as *Program Keluarga Harapan* (Family Hope Program, or PKH). Starting in 2007, the government introduced PKH in 438 sub-districts across Indonesia (randomly selected from a pool of 736 subdistricts), to a total of about 700,000 households. The unit of

¹ To our knowledge, there are two notable exceptions. First, Barrera-Osorio, Linden, and Saavedra (2017) experimentally evaluate the long-run effect (8-12 years) of an unusual CCT and savings program in Bogotá, Colombia that focused on incentivizing high school enrollment, and studied the effect on tertiary enrollment in universities. Second, in a paper contemporaneous to this one, Molina Millán et al. (2018) study the education effects of the Honduran CCT, which was implemented for five years.

² Some papers exploit the time gap between treatment and control groups receiving the program (even if for a short period), and thus look at medium- or long-run effects for children affected for longer periods by the program (for example, see Behrman, Parker, and Todd 2011; Barham, Macours, and Maluccio 2017; Kugler and Rojas 2018). While this approach is very informative in terms of some outcomes, it could underestimate impacts as the control group is also exposed to the program just a few years later. Others exploit non-experimental variation using discontinuities in who received the transfers (e.g., Filmer and Schady 2014) or cohort analysis across areas with higher and lower program intensity (Parker and Vogl 2017). A third approach is to study long-run effects of temporary programs, such as Baird et al. (2016), who compare treatment and control areas two years following the end of a two-year temporary transfer program.

randomization, the Indonesian subdistrict, is large – a subdistrict has about 50,000 people, and the 736 subdistricts in the experimental sample have a total population of over 36 million people. Targeted households received between 600,000 to 2,200,000 Rupiah (approximately USD 60 to 220) per year, with typical CCT conditions for children (pre- and post-natal care, deliveries with trained birth attendants, regular growth monitoring, immunizations, enrollment and attendance of children in primary and junior secondary school). Households that were enrolled into the program in 2007 continue to receive quarterly benefits today. The World Bank conducted a follow-up survey in 2009, about two years after the rollout, in a randomly chosen subset of 360 sub-districts, intended to be the end of the evaluation (Alatas 2011).

Crucially, while PKH has subsequently been expanded to many more areas in Indonesia – by 2013 it had reached over 3,400 sub-districts, spread over 336 districts in all of Indonesia's provinces, and covered over 2.3 million households – 60 percent of the initial control sub-districts were still not treated nearly six years later. The reason for this expansion to new provinces and districts, rather than to the control group, was that the government prioritized the expansion of the program to new areas, such that the program would be spread throughout the country, rather than focused intensely in a small number of geographic areas.

For research, however, this presents a unique opportunity because the initial randomization of sub-districts to treatment and control status continued to induce random variation in program placement six years later. To study how sustained CCT benefits affect families over this longer time horizon, in 2013 we re-surveyed households that were in the initial experiment. Notably, we successfully re-surveyed at least one household member from 95 percent of the original 14,326 households in the baseline survey. We show that the experimental first stage – the regression of whether a household is receiving PKH on whether the household's sub-district was randomized in 2007 to be in the treatment group – is strong (F-statistic over 400). This unique setup – wherein the experiment continued to run at-scale by the government for over six years without any researcher intervention in the program implementation – allows us to examine whether the static effects of CCTs on targeted indicators persisted even as the program continued over time, as well as whether these human capital investments began to cumulate as children grew up exposed to the program.

We start by examining whether conditional cash transfers continued to have impacts on

incentivized behaviors, even after the program had been running at scale for six years. We find remarkable effects on several of the incentivized indicators, which remain significant even accounting for multiple hypothesis testing. In particular, treated households were more likely to have childbirth assisted by a skilled birth attendant (doctor or midwife; increased by 23 percentage points) and delivery at facility for those who had given birth (increased by 17 percentage points). These are dramatic increases – they imply that the CCT program reduced the share of children not born at a health facility by 62 percent, and virtually eliminated all births not delivered by a trained midwife or doctor. Conditional cash transfers also had large impacts on reducing the share of children not in school: school enrollment rates for the targeted age group – 7- to 15-year-olds – were about 4 percentage points higher for the treatment group than for the control group in the six-year follow-up. Since 92.4 percent of control group children were already enrolled in school, this means that the program eliminated 53 percent of non-enrollment.

We then turn to whether continued exposure to the conditional cash transfers began to lead to results on outcomes that require cumulative investments. We find very large impacts on children's propensity to be stunted or severely stunted.³ In fact, we observe a 23 percent reduction in the probability of being stunted (defined as being 2 standard deviations less than the WHO's height-for-age standard), and a 56 percent reduction in the probability of being severely stunted (3 standard deviations less than the WHO's height-for-age standard). There were no detectable stunting effects in the two-year follow-up. We find no impact on malnourishment (2 standard deviations less than the WHO's weight-for-age standard).

To capture the cumulative effects of educational investments, we look at impacts on older children – aged 15-21 at the time of our surveys, who were in the target age range (aged 9-15) at the time PKH began. We find evidence that children aged 15-17 were about 10 percentage points more likely to still be enrolled in school, reducing the non-enrollment rate by 27 percent. We also find some evidence that high school completion rates increased – by 7 percentage points, representing a 29 percent increase (p -value 0.14 after adjusting for multiple hypothesis testing). We find no evidence that this translated to a higher

³ Weight-based measures respond more quickly to nutrition and health status, whereas stunting is thought to respond to accumulated early childhood conditions (UNICEF 2013; Hoddinott, Alderman, Behrman, et al. 2013).

likelihood of wage employment for those aged 18 to 21 years, nor do we find impacts on early marriage.⁴

The final piece of our analysis is to examine whether the continued cash transfers – which add up to an average of US\$970 per household over the six-year period – had a transformative effect on the recipient households themselves. For example, Gertler, Martinez, and Rubio-Codina (2012) find that PROGRESA households invested a fraction of accumulated transfers in productive assets, which could affect the overall poverty status of the household. However, we find no evidence of this here. While the point estimates of the impact on consumption are positive, we cannot distinguish the measured impacts from zero; we also cannot reject effects equal to the size of the transfer, which was about 7 percent of household consumption. What we can definitively rule out, however, are transformational impacts on household consumption: given our confidence intervals, we can reject that household per-capita consumption increased by more than about 10 percent. We also find no observable effect on productive household assets, such as livestock owned, or on fixed assets, such as land.

In short, conditional cash transfers in Indonesia continued to have impacts on the incentivized health and educational investments of households six years after program introduction: in particular, the program continued to impact primary and secondary school education attainment and deliveries in a facility by trained birth attendants. This occurred despite the fact that the level of benefits fell from 14 to 7 percent of monthly household consumption and also despite the fact that the program was being run with business-as-usual practices by the government (without any researcher involvement). And, perhaps more importantly, after continued investment in children over time, we begin to see some substantial results on “cumulative outcomes”, which was the original rationale for sustained payments over time. In particular, stunting was greatly reduced, suggesting large health gains, and school enrollment for high school-age children increased. On the other hand, we see no transformational effects over six years of repeated cash transfers on the incomes of the beneficiary households themselves. Combined, this suggests that if

⁴ Note that unlike the results on stunting, where children were on PKH most of their lives, the children in the 15-21 age category were older to begin with when PKH started. If we believe that cumulative effects come from PKH’s focus on early childhood, these children are less likely to have been impacted than the younger children.

conditional cash transfers are going to indeed break the cycle of poverty, this effect is going to happen through impacts on the subsequent generation, rather than through impacts on households themselves.

The rest of the paper proceeds as follows. We describe the setting, experimental design, and data in Section II. We provide the findings in Section III, while Section IV concludes.

II. SETTING, EXPERIMENTAL DESIGN, AND DATA

A. Program

We study the cumulative (six-year) effects of the Indonesian government's conditional cash transfer program, Program Keluarga Harapan (PKH or "Family Hope Program"). Launched in 2007, the program provides quarterly cash transfers to very poor households with children or pregnant and/or lactating women, with a fraction of the payment conditional on a number of health- and education-related obligations. By providing a sustained flow of payments to families over many years, the program aims "(a) to reduce current poverty and (b) to improve the quality of human resources among poor households" (Alatas 2011).

The government targeted extremely poor households, approximately in the bottom 10 percent of the per-capita consumption distribution. To determine their eligibility, Statistics Indonesia (*BPS*) conducted a door-to-door survey of potentially eligible households; the survey included 29 asset and demographic questions.⁵ They applied a proxy-means test formula to this data, and households that were below a pre-determined cutoff were deemed to be financially eligible. Statistics Indonesia then kept households that also met demographic requirements: households with a pregnant and/or lactating woman, with children aged 0-15 years, and/or with children aged 16-18 years who had not completed 9 years of basic education.

Eligible households began receiving quarterly cash payments through the nearest postal office. The amount of cash was designed to be about 15 to 20 percent of annual household income, depending on the age of the children; payments ranged from Rp. 600,000 (US \$60) to Rp. 2,200,000 (US \$220) per household per year. The payments were made to women in the household. As with most conditional cash

⁵ BPS visited all households that had been included in a previous 2005 unconditional cash transfer program, and they also worked with local officials to visit any potentially poor households who may have not been included.

transfer programs, households were informed that they had to complete a number of conditions to continue receiving the transfers. For example, households with children aged 0 to 6 years needed to ensure that children completed childhood immunizations and take Vitamin A capsules a minimum of twice per year, and also must take children for growth monitoring check-ups (see Appendix Figure 1 for the full list of conditions). Trained facilitators provided beneficiaries with information and advice, and also verified compliance with conditions: one violation would result in a warning letter, a second violation would lead to a 10 percent cut in benefits, and a third violation would lead to program expulsion. However, in practice, the verification system did not begin until at least 2010, and even afterwards, conditions were not always enforced. In that sense, this program is more akin to a ‘labeled’ CCT program, such as the Moroccan program studied by Benhassine et al (2015).

B. Sample, Experimental Design, and Timing

The Government of Indonesia first introduced the conditional cash transfer program in six provinces (West Java, East Java, North Sulawesi, Gorontalo, East Nusa Tenggara, and the capital city of Jakarta). Within each province, the government excluded the richest 20 percent of districts, and then determined which sub-districts within the remaining districts were “supply-side ready” (based on availability of midwives, doctors, and middle schools) to participate in the program. A total of 736 sub-districts (with a total population of about 36 million individuals in 2005) were included in their sample, and 438 of these sub-districts were randomly selected for the treatment group. About 700,000 households in these selected sub-districts were enrolled in the conditional cash transfer program.

Out of the 736 sample sub-districts, 360 sub-districts were randomly chosen for data collection (180 treatment, 180 control). Appendix Figure 2 shows the distribution of sample sub-districts, by experimental assignment, across Indonesia, including both on and off Java.

As shown in Appendix Figure 3, The World Bank conducted a baseline survey in June to August 2007, and the program was launched in these sub-districts soon afterward. The World Bank conducted a follow-up survey from October to December 2009, about two years after the start of the program; the results

are described in Alatas (2011), and we re-analyze these data below to ensure comparability with our analysis. We conducted a follow-up survey from September to November 2013, about six years after the intervention, using identical survey questionnaires.

The evaluation we conduct is possible because subsequent program expansions kept the control group largely intact. Appendix Figure 4 shows the evolution of the PKH program over the time period that we study based on administrative data on the program's expansion. In 2009 (at the time of the two-year follow-up), the program was operating in 99 percent of the locations randomized to treatment and in 22 percent of locations randomized to control, which implied a sub-district-level 'first stage' of 77 percentage points in 2009. By 2013 (six-year follow-up), the program had expanded somewhat, but the experiment still remained intact: the program was operating in 99 percent of locations randomized to treatment, and in 39 percent of locations randomized to control, for an implied sub-district-level 'first stage' of 60 percentage points. Thus, after six years, the original sub-district-level randomization still had substantial bite. Moreover, because the program reached fewer households in areas in the control sub-districts that received the expansion between 2009 and 2013, the first stage for receiving PKH at the *household level* is virtually identical in both 2009 and 2013. As described below, we use the original sub-district randomization as an instrument for treatment.

E. Data, Data Collection, and Experimental Validity

The World Bank collected both a baseline survey and initial follow-up survey in the 360 sub-districts to assess PKH's short-run program impacts (see Alatas 2011 for more details). These surveys were conducted using the same survey instruments as, and in tandem with, the evaluation of the *Generasi* community block grant program, which was being carried out in 300 separate sub-districts but was targeting similar indicators (see Olken, Onishi, and Wong 2014).

As shown in Appendix Table 1, 14,326 households (73,578 individuals) were surveyed at baseline in the 360 sampled sub-districts between June and August 2007. To create this sampling frame, 8 randomly selected villages were drawn from each sub-district, and then one sub-village was selected within each

village.⁶ From within each village, four households were randomly selected from the government's interview lists, stratified such that two households included a pregnant or lactating mother or a married woman who was pregnant within the last two years and the other two included children aged 6-15 years of age. Note that since the survey was conducted in both treatment and control areas (and we do not know who would have received the conditional cash transfers in the control areas), households were randomly selected to be surveyed from the initial asset listing (not the beneficiary list), so not all households would have ultimately received the CCT. There was very little attrition of households in the first follow-up survey that was conducted from October to December 2009: 13,971 households (97.5 percent of baseline) were found and surveyed, and households that split and moved within the sub-district were also surveyed (so the sample size increases slightly in each round).

Both the baseline and follow-up surveys included modules for consumption, demographics, assets, education, and health outcomes. Additionally, they included anthropometric data (height and weight measurements) for children aged 0 to 36 months in the baseline survey, and for children aged 0 to 60 months in both follow-up surveys.

This paper focuses on the medium-run follow-up, which we conducted from September to November 2013. The survey is a panel, and tracked the original households included in the baseline. Overall household attrition was again low: 13,619 households, or 95.1 percent of baseline households, were found, with the attrition rates nearly identical across the treatment and control groups (see Appendix Table 1).⁷

While household attrition was low, it could be that some household members left the sub-district and thus were not tracked. This is not an issue for the young children born after the baseline, since we measure all children present in any household we track, and we track 95 percent of baseline households.

⁶ If there were fewer than 8 villages sampled in a sub-district (since there were not enough eligible households in enough villages), or if there were fewer than 5 potential households to survey in the sub-village, additional sub-villages from the same village were randomly selected.

⁷We followed entire households or household members that moved within the same sub-district. In addition, we surveyed 362 households that were added to the sampling frame in the two-year follow-up, and 751 households that were added in the six-year follow-up, in order to compensate for household attrition.

Thus, attrition should likely not be a concern for outcomes like completed vaccinations or stunting.⁸

However, attrition could potentially be relevant for the oldest children at baseline, who are now teenagers and young adults, and for outcomes in this age range such as high school completion or teenage marriage. In Appendix Table 3, we examine attrition for those who were initially aged 6-15 years in the baseline. We find and re-survey 90 percent at the two-year follow-up (Column 1) and 72 percent at the six-year follow-up (Column 2), but with no differential attrition between the treatment and control group in either survey, either overall or differentially between boys and girls (Columns 3 and 4).⁹

Appendix Table 6 shows that the final sample is balanced across treatments in terms of baseline characteristics. In Column 2, we provide the control group mean for the variable listed in that row, while Column 3 provides the mean for the treatment group. Column 4 provides the difference between the treatment and control for that variable (clustered by sub-district, which is the level of randomization) and Column 5 provides this difference conditional on strata (districts). Of the 14 variables considered, only one difference is statistically significant at the 10 percent level, consistent with chance.

III. RESULTS

We first outline our empirical strategy and show the first stage results. We then examine results on three key dimensions: ongoing impacts on incentivized health and education behaviors, cumulative effects on proxies for human capital, and cumulative effects on household economic outcomes.

A. Empirical Strategy and First Stage Results

⁸ An alternative concern for younger children is that the CCT differentially affected infant mortality and this biases the results for the young children. However, this does not appear to be the case: In Appendix Table 2, we show that there was no observable difference in fertility, miscarriage, stillbirth, or infant mortality rates in either follow-up survey.

⁹ We can further disaggregate by age at baseline to determine if attrition is worse for older children, who may be more likely to leave the household for work or marriage. In Appendix Table 4, we observe that attrition does indeed increase with baseline age. However, this attrition does not appear differential by treatment group: only 4 out of 40 regression coefficients are significant, which is consistent with chance. We also show in Appendix Table 5b that there are no differences in the reasons for child migration across the treatment and control group. In short, while older children were more likely to have migrated than younger children, the probability of migration, as well as the reasons for doing so, do not appear to be associated with treatment status.

While compliance with the randomization protocol was generally high, it was not perfect, and some control areas were treated. In addition, only a subset of households on the initial Statistics Indonesia interview lists ultimately became beneficiaries, as there was a subsequent screening step to determine categorical eligibility. Therefore, we conduct an instrumental variable analysis in which we instrument program receipt ($ReceivedCCT_{hsd}$) with whether households were located in an initial treatment sub-district:

$$Y_{hsd} = \beta_0 + \beta_1 ReceivedCCT_{hsd} + \mathbf{X}'_{hsd} \boldsymbol{\gamma} + \alpha_d + \varepsilon_{hsd} \quad (1)$$

Y_{hsd} is the outcome of interest for household h in sub-district s in district d . $ReceivedCCT_{hsd}$ is a dummy variable for whether the household has ever received the CCT program, while α_d is a set of district fixed effects. For additional precision, we include the following baseline control variables in \mathbf{X}_{hsd} : house roof type, wall type, floor type, household head's education level, household head works in agriculture, head of household works in services, log monthly per-capita expenditure, log household size, and dummies for whether the household has clean water, has its own latrine, has a square latrine, has its own septic tank, and has electricity from the state electric company.¹⁰ We cluster the standard errors by sub-district, the level of the randomization. We adjust p -values for multiple hypothesis testing within each panel of results using the stepdown method of Romano and Wolf (2005; 2016), and report the resulting adjusted p -values in brackets.¹¹ We also report the p -value for the difference between the two- and six-year effects.

Table 1 provides our first-stage estimates. Column 1 shows the results of our analysis of the World Bank's data from the two-year follow-up for comparison, while Column 2 provides our six-year follow-up results. In the last row, we also provide the F-statistic from a test of the instrument.

The regressions show a strong – and almost identical – first stage in both the short-run and medium-run. By the two-year follow-up (Column 1), about 9 percent of the control group reported receiving the

¹⁰ For missing baseline data, we fill in the control variables with zero and create a dummy variable to indicate missing values for each variable.

¹¹ We do this for each family of outcomes. In Tables 2, 4 and 5, we group outcomes by survey-round (i.e., columns). In Table 3, 6 and 7, we group outcomes by survey round and panel, since some of the panels are essentially summary variable of other variables (e.g., enrollment for age 7-15 is a weighted average of enrollment by categories, with the categories broken down in subsequent panels), or because the outcomes are from a different family of outcomes (e.g., marriage versus education).

CCT, with a 37.5 percentage point increase in the treatment group (p -value less than 0.001). The results are similar in the 6-year data (Column 2), with a 36.8 percentage point increase in the treatment group relative to the 13.1 percent of households having ever received the CCT in the control group (p -value less than 0.001). The instruments are strong, with F-statistics over 450. Moreover, it is largely the same households continuing to receive the program over time: 83.6 percent of those households who report currently receiving the CCT in the two-year survey also report receive it currently in the six-year survey; conversely, 85.2 percent of the panel households in our data who receive the CCT at six years also reported receiving it at two years.

It is important to note that the randomization is at the sub-district level, and hence control households come from other sub-districts. Virtually all health and education services (health clinics, schools) are contained within sub-districts, so spillovers *across* sub-districts are extremely unlikely in this context, and indeed, this was the reason the randomization was done at such a high level.

A second question is whether there are spillovers to non-treated households *within* treated sub-districts. This assumption of no within-sub-district spillovers is important for the exclusion restriction implicit in estimating equation (1) with instrumental variables. While in other contexts this has been a concern – see, e.g., Angelucci and di Giorgi (2009) in the PROGRESA case – there are two reasons why even within-sub-district spillovers seem very unlikely here. Unlike PROGRESA, which treated over 60 percent of households in treatment villages, PKH was targeted at the poorest of the poor households, and as such treated a far smaller fraction of households in a village – in 2009, for example, the typical treated village had only 78 PKH beneficiary households out of a mean of 1200 households, meaning PKH treated only 6.5 percent of households in a village on average.¹² General equilibrium effects (e.g. congestion at schools or health clinics, or positive spillovers through supply-side changes) are therefore likely to be very

¹² One would expect that the average effect on a non-beneficiary household is therefore likely at least an order-of-magnitude smaller than the average effect on beneficiary households. Since our household survey intentionally sampled households who were likely to be beneficiaries, beneficiary households are about 50 percent of the sampled households in the treatment area, even though they are only 6.5 percent of the total population. This means that the average effect in our sampled households will be driven almost entirely by the effect on treated households.

small in our case, given how small the share of treated households is.¹³

B. Impacts on Incentivized Human Capital Behaviors

Health. We begin by examining incentivized health-seeking behaviors. Table 2 reports results of IV regressions, where we estimate separate regressions for the outcomes of interest listed in each row. Our key estimates from the six-year follow-up survey are shown in Column 2; each cell presents the IV effect of PKH treatment analyzed using equation (1).¹⁴ For ease of comparability, in Column 1, we show the results from the two-year survey conducted by the World Bank which we re-analyzed using the same IV specification as (1).

Our analysis shows continued effects of the CCT program on a range of health-seeking behaviors in the medium-run, particularly with regard to maternal health-seeking behaviors. We first examine health-seeking outcomes for women who became pregnant or gave birth within the 24 months prior to each follow-up survey. We find that the transfers continued to have large, positive effects on the probability that childbirths were assisted by trained personnel (doctors or midwives) in the six-year follow-up, and that deliveries were more likely take place in a health facility. Specifically, the estimates imply that the CCT program led to a 17 percentage point increase in delivery at a health facility at the six-year follow-up (24 percent increase), and a 23 percentage point increase in the probability a birth was assisted by a trained midwife or doctor (about double the effect at the 2-year mark; *p*-value of 2-year vs. 6-year difference 0.079).

¹³ Triyana (2016) studies whether there are changes in service provision as a result of PKH in the two-year follow-up. She finds no effect on the number of doctors or traditional attendants, but finds a small increase in the number of midwives. In Appendix Table 7, we examine the effect of PKH on the number of doctors, midwives, traditional birth attendants and schools in the six-year follow-up and do not find an increase on the level of any of these measures of supply-side service availability, suggesting little presence of spillovers through supply responses. In addition, we also conduct an alternative identification strategy that uses baseline assets interacted with sub-district-level treatment status to predict treatment at the individual level, not at the aggregate level. As shown in Appendix Tables 8 and 9, this produces very similar results to the univariate treatment-vs-control sub-district-level instrument (and, in particular, these results are not systematically smaller than the univariate instrument results), further suggesting empirically that spillovers are very small in our context.

¹⁴ Our results are robust to the model specification choices that we made. For example, in Appendix Table 10, we replicate Table 2 using “currently receiving” a CCT, rather than “ever received,” as our variable of interest, because some households received PKH in the two-year follow-up survey but had stopped receiving it by the six-year follow-up; the results look nearly identical (which is not surprising given that the overlap of households in both categories is high). Similarly, in Appendix Table 11, we replicate Table 2, but drop baseline controls. Again, we find similar coefficients, but sometimes we lose some statistical precision when omitting the baseline controls.

These are dramatic changes – they imply that the CCT program reduced the share of children born outside a health facility by 62 percent, and virtually eliminated births not assisted by trained midwives or doctors. These effects remain statistically significant even after adjusting for multiple hypothesis testing.

However, unlike in the short-run, we do not find statistically detectable impacts on pre- and post-natal visits, though the point estimates are positive and we cannot reject that the two- and six-year effects are statistically equal. One potential reason is that the control group increased their overall number of visits in the intervening years (for example, the control mean increased from 6.5 to 7.1 pre-natal visits) and essentially caught up to the treatment group. In terms of the care women and children received, we observe no effect on receiving a full set of iron pills during pregnancy, either in the two- or six-year follow-up.

Health inputs into young children also appear to have improved in the medium-run for children who had ever been covered by PKH since the baseline survey, but any observed effects are relatively weak. While there was no observable impact on immunizations in the two-year follow-up survey for children in our baseline sample, we observe about a 5 percentage point increase in the percent of age-recommended immunizations completed, though this effect does not survive multiple inference adjustment and we cannot reject that the two- and six-year mark effects are statistically equal. We observe no increase in the number of times children between the ages of 6 months and 2 years received Vitamin A. We observe increases in the number of times a child was weighed by a health professional in the last 3 months (for those aged 0 to 60 months), though these increases are no longer significant and smaller in the six-year follow-up compared to the two-year follow-up (the difference between the two- and six-year effects is statistically significant with a *p*-value of .001).

Education. The second component of the CCT incentives focused on increasing enrollment and attendance of primary and junior secondary school age students, i.e., those students aged 7-15 years. Table 3 presents the results for these children. In Panel A, we examine enrollment and attendance for children aged 7-15

(Panel A), and then disaggregate further by ages 7-12 (Panel B) and 13-15 (Panel C).¹⁵

We find substantial increases in enrollment for all children aged 7 to 15: the CCT program increased enrollment rates by 4 percentage points in the six-year survey. Since 92.4 percent of control group children were enrolled in school, the 4 percentage point increase in enrollment represents a 53 percent decrease in the fraction of students who were not enrolled in school; that is, the CCT program eliminated more than half of non-enrollment, making a large dent in the last-mile enrollment problem.¹⁶ This effect is statistically different --slightly smaller-- from what was observed at two years – a 6.4 percentage point increase in enrollment, which represents a 66 percent decline in the non-enrollment rate (the *p*-value of the difference in effects is 0.081) – though the decline in treatment effect from year 2 to year 6 is entirely due to an increase in enrollment in the control group, rather than a decline in the treatment group. It is important to note that these increases are not just nominal enrollment: we observe substantial increases in the percentage of children who reported attending school at least 85 percent of the time in the last two weeks in both the two- and six-year follow-ups.¹⁷

Disaggregating the effects by age group, we see that the six-year effects are concentrated among older students (Panel C). For students aged 13-15, we see increases in school enrollment of 9 percentage points, representing a 52 percent decline in the non-enrollment rate. For students aged 7-12, we do not observe a statistically significant increase in enrollment, but note that the enrollment rate in the control group is 97.2 percent in this age range, so obtaining gains in this age group is likely to be difficult.¹⁸

In sum, the conditional cash transfer remained highly effective at reducing non-enrollment in

¹⁵ Appendix Tables 12 and 13 provide the results using the “currently receiving” CCT variable and with no baseline controls, respectively, and show similar findings to Table 3.

¹⁶ Note that if we redo the FWER multiple inference adjustment among all incentivized indicators, i.e. combining Table 2 and Panel A of Table 4, the statistical significance levels remain unchanged. The multiple-inference adjusted *p*-values for enrollment and attendance become <0.001 and 0.012, respectively, and the results on assisted deliveries and deliveries in facility in Table 2 remain statistically significant.

¹⁷ Note that this attendance measure is defined as 0 for those non-enrolled in school. This measure therefore captures the combination of enrollment and attendance decisions, since both can respond to the CCT program.

¹⁸ Appendix Table 14 examines the effects for boys and girls separately. As shown in Panel B, younger boys in households receiving the CCT program were more likely to be in school, while we find no effect for younger girls—however, girls had very high rates of enrollment to begin with (98 percent were enrolled in some form of school). While somewhat larger in magnitude for older boys than for older girls, we nonetheless observe treatment effects of the CCT program on the enrollment and attendance of both older boys and girls (Panel C).

school for those in the targeted age category, particularly for older students (age 13-15), for whom non-enrollment is a substantial issue. More generally, despite the fact that the program has been running at large scale by the government for six years, and with no researcher monitoring or intervention, it continues to be effective in improving targeted health-seeking and education behaviors.

C. Cumulative Impacts on Proxies for Human Capital

Anthropometric Impacts. The results thus far have shown that health-seeking behaviors continued to be positively affected by the CCT program. This implies that at the time of the six-year survey, young children (those under 5 years old) had spent their entire lives with higher levels of improved health services at various points in their life cycle. A natural question is whether this increased health utilization accumulated and led to changes in health outcomes.

We examine this question in Table 4. We explore anthropometric outcomes for children aged 0 to 60 months.¹⁹ We start by examining measures of stunting. Stunting is considered a measure of cumulative health investments during the first few years of life (Hoddinott, Behrman, Maluccio, et al. 2013; Jayachandran and Pande 2017); it is also thought to be correlated with worse cognitive and economic outcomes later in life (Case and Paxson 2008; Gluwe and Miguel 2008; Hoddinott, Maluccio, Behrman et al. 2011; Guven and Lee 2013). We follow WHO definitions and define stunting as being more than 2 standard deviations below the WHO-standardized height-for-age median; severe stunting is defined as being more than 3 standard deviations below the WHO-standardized height-for-age median.

We observe very large reductions in stunting among children aged 0 to 60 months in the six-year follow-up survey. Stunting declined by roughly 9 percentage points, i.e., a 23 percent reduction in the probability of being stunted. Severe stunting declined by approximately 10 percentage points, i.e., a 56 percent reduction.²⁰ Both boys and girls benefited from the CCT program in terms of decreased stunting

¹⁹ Appendix Table 15 provides the results for those currently receiving the CCT (as opposed to those who ever received the CCT) and Appendix Table 16 does so without controls. The findings remain robust to these specification changes.

²⁰ Appendix Figure 5 estimates the impacts on stunting non-parametrically by child age, and finds similar reductions in stunting across 0- to 60-month-olds; if anything, the figures suggest somewhat larger reductions for older cohorts.

and severe stunting, although the point estimates are slightly larger in magnitude for boys than for girls (see Appendix Table 17). While the point estimates indicate stunting reductions of about 3 percentage points after the program had been in effect for two years, these estimates are not statistically significant. While we cannot reject that the two and six year effects on stunting are the same (p -value of 0.221), we find a statistically significant difference between the two- and six-year effects on severe stunting (p -value of 0.072).

We observe no impacts on malnourishment (i.e., weight-for-age more than 2 standard deviations below WHO standard)—which responds more quickly to health investments—in either the two- or six-year follow-up.²¹

Potential mechanisms for stunting effects. Given that a conditional cash transfer is a bundled intervention (cash + incentives), it is hard to disentangle which specific channels could drive the observed stunting reductions. However, we can explore three potential mechanisms. First, it could be that the increased health-seeking behaviors (shown in Table 2) increased interaction with medical professionals, which in turn reduced stunting through increased maternal knowledge and health behaviors. We do find that treatment mothers are more likely to know and report their child's birthweight, but on net, we do not observe large changes in other indicators of maternal knowledge or behavior (Appendix Table 18). Second, it could be that PKH drives improved nutrition (see Appendix Table 19). We do find changes in child protein intake in response to the CCT program: children aged 18-60 months were roughly 10 to 11 percentage points more likely to have consumed milk, and 10 to 12 percentage points more likely to have consumed eggs, in the

²¹ Most experimental evaluations of CCTs to-date have not shown large effects on stunting, but there are a few exceptions. Fernald et al. (2008) look at stunting in the PROGRESA context in 2003, comparing families who were initially part of the 1997 rollout with those in the control group that received PROGRESA 18 months later, and find that longer exposure to PROGRESA led to reductions in stunting. Kandpal et al. (2016) experimentally measure the effect of the Philippines's Pantawid program on stunting 30-31 months after its introduction, and find reductions in stunting for children aged 6 to 36 months. They argue that part of the reason for the impact could be the program's focus on nutrition (particularly dairy) in family development sessions. Baird et al. (2016) do not report any effect of a UCT or CCT program in Malawi on child height, though the transfers in the program they study were in place for only two years, and ended two years prior to their survey.

week prior to the two-year follow-up survey. Note, however, we do not find similar results in the two-year follow-up survey. Finally, we explore whether reported illness of the children declined, under the hypothesis that sick children would have more stunted growth paths. Appendix Table 20 shows no observable declines in reported acute illness rates for children under 5.

Impacts on Child Labor. Child labor is thought to be an important concern since it likely crowds out human capital accumulation. We therefore examine whether the CCT, which we saw led to substantial gains in schooling for children age 13-15, is also associated with changes in child labor.²² Table 5 shows that the program reduced the fraction of children engaged in wage work by 4.4 percentage points, i.e., a reduction of 48 percent. For those working extensively, which we define as working for a wage at least 20 hours in the past month, the effects are similar: a reduction of 3.0 percentage points, i.e., a decline of 44 percent. Appendix Table 22 shows that reductions in wage work are primarily found for boys, who are more likely to be working for a wage than girls.²³ The fact that we see both substantial increases in school enrollments *and* declines in wage work for the same age groups suggests that the two effects may be related.²⁴

Impacts on High School Education, Labor, Early Marriage, and Early Fertility. Our final set of results for children explores outcomes for children who were aged 9 to 15 when the CCT program was initially rolled out, and hence were between ages 15 and 21 at the time of the six-year follow-up. This allows us to explore the cumulative effects of the CCT on final educational attainment and early adulthood outcomes after the

²² We explore this outcome among children aged 13-15 since child labor among younger children is extremely rare. Appendix Table 21 presents the results for younger children (aged 7 to 12). We do not find any changes for this age group in the 6-year follow-up, but this is not surprising as only 1.6 percent of children in the control group report working for a wage at all, and only 0.4 percent report working more than 20 hours for wage in the past month.

²³ In Appendix Table 23, we explore alternative measures of non-wage work. In the medium-run, we observe some reductions in working more than 20 hours a month for a family business in for those aged 13-15 (but this is not significant in all specifications) and no effect on “helping out at home.”

²⁴ In Appendix Table 24, we show that the CCT program led to a decline in the number of students who were both enrolled in school *and* working for a wage at the same time. This fact – that the CCT program reduces, not increases, the number of students doing both wage work *and* being enrolled in school – suggests that the effects we see on enrollment and work are not coming exclusively from a time-budget constraint, but rather may be related to the income effects of the CCT.

incentives have ended.

We begin by exploring educational outcomes for this cohort, shown in Table 6, Panel A.²⁵ We find large increases in the probability of those aged 15 to 17 attending any kind of school in the six-year follow-up, with some of this effect driven by increases in high school enrollment.²⁶ We also find some evidence of an increase in high school completion rates for those aged 18 to 21 (about 7 percentage points, with a FWER-adjusted *p*-value of 0.139). As shown in Appendix Table 26, most of the increases in educational attainment for these age categories are driven by boys, who show very large impacts on high school enrollment (13 percentage points, representing a 38 percent increase) and completion rates (9.7 percentage points; 42 percent increase). We find no impact on high school enrollment or completion rates for girls.

We then explore work outcomes for this age group, shown in Panel B of Table 6. We find no impacts on the probability of wage work, for either 16- to 17-year-olds or 18- to 21-year-olds. As shown in Appendix Table 26, we find no effect for either boys or girls. For 16- to 17-year-olds, one might expect a decrease in wage work due to the increases in school enrollment documented above; while we do not observe this, as shown in Panel A of Appendix Table 27, we do observe some decreases in helping out with the family business or housework (particularly for girls). For 18- to 21-year-olds, who are more likely to be out of school, we may have expected higher employment rates for wage work. However, this does not appear to be the case.²⁷ It is worth noting, however, that these 18- to 21-year-olds were already teenagers at the time the program started, and thus spent fewer of their formative years in the program than the young children for whom we observe reductions in stunting.

In Panel C, we explore whether the CCT program led to changes in age of marriage. Age of marriage could be delayed from the cash transfer's income effect (Baird et al. 2011), or from a delay in marriage due to the practical side of being enrolled in school longer. However, we find no evidence that the

²⁵ We omit outcomes for ages 18 to 21 from our reported two-year regressions, because virtually no respondents who were aged 9 to 15 at baseline had reached age 18 by the time of the two-year follow-up survey.

²⁶ As shown in Appendix Table 25, we observe increases in 15-year-olds and 17-year-olds attending any type of school. The high school effect appears largely driven by 17-year-olds.

²⁷ We do observe that the boys within this age category were more likely to help out with the family business (Panel B of Appendix Table 27), while girls were somewhat less likely to help out in any family business or housework.

CCT program changed the propensity to marry for those aged 16 to 17, or for those aged 18 to 21, though standard errors are large relative to the mean. Finally, we investigate changes in fertility (Appendix Table 28), and while we qualitatively observe postponement of births—decreases in fertility for girls aged 16 to 17 and increases for those aged 18 to 21—we cannot reject that these coefficients are different from zero.

D. Cumulative impacts on recipient households: Consumption, Work and Assets

The third main question we explore is whether the accumulation of repeated cash transfers had transformative effects on the economic condition of the recipient households themselves. The CCT program provides a quarterly transfer of cash to households for around 6 years. The cash payment is around 7 to 15 percent of total household consumption, cumulatively adding up to between \$360 and \$1320 – an average of \$970 – per household. We, therefore, ask whether this assistance was large enough to have a “transformative effect” on households, shifting them out of poverty. One mechanism for this could be that households save part of the transfers over time, and use this to invest in productive assets. For example, Gertler, Martinez, and Rubio-Codina (2012) find that PROGRESA beneficiary households invested a fraction of their accumulated transfers in productive assets.

In Table 7, we examine the impacts of the transfer on household consumption, adult employment, and household assets.²⁸ While we observe positive impacts of the transfer on overall log per-capita consumption (Panel A), we cannot distinguish these measured impacts from zero; on the other hand, we cannot reject an increase of consumption equal to the amount of the transfer at the six-year follow-up. As shown in Panel B, we do not observe changes in employment rates, although even in the control group, nearly 94 percent of the control group was employed regardless. Finally, the CCT program did not lead to increases in assets as transfers accumulated over time, such as livestock ownership.

In short, we do not observe transformational effects on household economic outcomes. To the extent that the CCT program leads to substantially large changes in material household welfare, these will

²⁸ In Appendix Table 29, we disaggregate expenditure outcomes by province. While we find noisy estimates, the increase in log consumption appears largest in the relatively poorer province of East Nusa Tenggara.

likely come through the effects on the next generation, who experience increased health and education, rather than a reduction in overall poverty of the current generation.

IV. CONCLUSION

The decision to redistribute through targeted transfers is a complex one. Some arguments are at core ethical, arguing that a society should protect the vulnerable and give them some additional help. But other arguments are economic, asserting, for example, that transfers allow households to make business investments that can have transformational impacts on household income and reduce poverty. Still others make arguments based on the intergenerational transmission of poverty, with transfers as a mechanism to help increase investments in child health and education.

We evaluate these claims in the context of a large-scale, government-run conditional cash transfer program, which provides moderately sized, regular, financial assistance to households that adhere to conditions that aim to improve investments in child health and education. We find that even though the program has been running for six years—without any researcher involvement in later years, and with a dynamic economic landscape—the program continues to promote remarkable health and educational investments in children explicitly targeted by the program. For example, six years after the program launched, we observe dramatic increases in usage of trained health professionals and facilities for childbirth, and a reduction of more half of the share of children aged 7 to 15 who are not enrolled in school.

Perhaps even more importantly, for children who have grown up their entire lives in households receiving these transfers, we also begin to observe impacts on outcomes that may require cumulative investments: for example, six years after the program began, we observe large reductions in stunting for young children and increased school enrollment for older teenagers. While this does not yet translate to increases in employment for individuals who have just started to enter the labor force, these are children who were already teenagers at the time the program started, and have thus spent fewer of their formative years in the program. The stunting results suggest that effects may be larger in the very long-run for children who benefited from the program during early childhood.

In contrast, we do not observe any impact on beneficiary households' current consumption, employment, or assets—suggesting that the additional help that the program provides does not have a transformational poverty reduction effect for those currently on the program. Rather, given that our results show that CCTs help poor households make significant investments in their children's health and education, an important part of the economic gains of CCTs likely could come from reductions in the intergenerational transmission of poverty.

WORKS CITED

- Alatas, Vivi. 2011. *Program Keluarga Harapan: Impact Evaluation of Indonesia's Pilot Household Conditional Cash Transfer Program*. Washington, DC: World Bank.
- Angelucci, Manuela, and Giacomo De Giorgi. 2009. "Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?" *American Economic Review* 99 (1): 486-508.
- Baird, Sarah, Craig McIntosh, and Berk Özler. 2011 "Cash or Condition? Evidence from a Cash Transfer Experiment." *Quarterly Journal of Economics* 126 (4):: 1709-1753.
- Baird, Sarah, Craig McIntosh, and Berk Özler. 2016. "When the Money Runs Out: Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?" World Bank working paper.
- Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives* 31 (4): 73-102.
- Barham, Tania, Karen Macours, and John Maluccio. 2017. "Are Conditional Cash Transfers Fulfilling Their Promise? Schooling, Learning, and Earnings after 10 Years." CEPR Discussion Paper 11937.
- Barrera-Osorio, Felipe, Leigh L. Linden, and Juan Saavedra. 2017. "Long Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia." *American Economic Journal: Applied Economics*, forthcoming.
- Bastagli, Francesca, Jessica Hagen-Zanker, Luke Harman, Valentina Barca, Georgina Sturge, and Tanja Schmidt. 2016. "Cash Transfers: What Does the Evidence Say?" Overseas Development Institute Report.
- Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. "Turning a shove into a nudge? A 'labeled cash transfer' for education." *American Economic Journal: Economic Policy* 7 (3): 86-125.
- Behrman, Jere R., and Petra E. Todd. 1999. "Randomness in the Experimental Samples of PROGRESA—Education, Health, and Nutrition Program." IFPRI Discussion Paper.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. 2015. "Interventions and Institutions: Experimental Evidence on Scaling up Education Reforms in Kenya." Unpublished paper.
- Case, Anne, and Christina Paxson. 2008. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy* 116 (3): 499–532.
- Fernald, Lia C.H., Paul Gertler, and Lynnette Neufeld. 2008. "Role of Cash in Conditional Cash Transfer Programmes for Child Health, Growth, and Development: An Analysis of Mexico's

Oportunidades." *Lancet* 371(9615): 828-837.

Filmer, Deon, and Norbert Schady. 2014. "The Medium-Term Effects of Scholarships in a Low-Income Country." *Journal of Human Resources* 49(3): 663-694.

Gertler, Paul. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *American Economic Review* 94(2): 336-341.

_____, Sebastian Martinez, and Marta Rubio-Codina. 2012. "Investing Cash Transfers to Raise Long-Term Living Standards." *American Economic Journal: Applied Economics* 4(1): 164-192.

Guven, Cahit, and Wang Sheng Lee. 2013. "Height and Cognitive Function at Older Ages: Is Height a Useful Summary Measure of Early Childhood Experiences?" *Health Economics* 22(2): 224-233.

Hoddinott, John, Harold Alderman, Jere Behrman, Lawrence Hadadd, and Susan Horton. 2013. "The Economic Rationale for Investing in Stunting Reduction." *Maternal & Child Nutrition* 9(S2): 69-82.

Hoddinott, John, Jere R. Behrman, John A. Maluccio, Paul Melgar, Agnes R. Quisumbing, Manuel Ramirez-Zea, Aryeh D. Stein, Kathryn M. Yount, and Reynaldo Martorell. 2013. "Adult Consequences of Growth Failure in Early Childhood." *American Journal of Clinical Nutrition* 98(5): 1170-78.

Jayachandran, Seema, and Rohini Pande. 2017. "Why are Indian Children So Short? The Role of Birth Order and Son Preference." *American Economic Review* 107(9): 2600-2629.

Kandpal, Eeshani, Harold Alderman, Jed Friedman, Deon Filmer, Junko Onishi, and Jorge Avalos. 2016. "A Conditional Cash Transfer Program in the Philippines Reduces Severe Stunting." *Journal of Nutrition* 146(9): 1793-1800.

Kugler, Adriana D., and Ingrid Rojas. 2018. "Do CCTs Improve Employment and Earnings in the Very Long-Term? Evidence from Mexico." NBER Working Paper 24248.

Levy, Santiago. 2006. *Progress Against Poverty: Sustaining Mexico's PROGRESA-Oportunidades Program*. Washington, DC: Brookings Institution.

Molina Millán, Teresa, Karen Marcours, John A. Maluccio, and Luis Tejerina. 2018. "Experimental Long-term Effects of Early Childhood and School-age Exposure to a Conditional Cash Transfer Program." Working paper.

Muralidharan, Karthik, and Paul Niehaus. 2017. "Experimentation at Scale." *Journal of Economic Perspectives*, 31(4): 103-24.

Olken, Benjamin A., Junko Onishi, and Susan Wong. 2014. "Should Aid Reward Performance?

Evidence from a Field Experiment on Health and Education in Indonesia.” *American Economic Journal: Applied Economics* 6(4): 1-34.

Parker, Susan W., and Tom S. Vogl. 2017. “Do Conditional Cash Transfers Improve Economic Outcomes in the Next Generation? Evidence from Mexico.” Working paper.

Robles, Marcos, Marcela G. Rubio, and Marco Stampini. 2015. “Have Cash Transfers Succeeded in Reaching the Poor in Latin America and the Caribbean?” Inter-American Development Bank Policy Brief No. IDB-PB-246.

Romano, Joseph P., and Michael Wolf. 2005. “Stepwise Multiple Testing as Formalized Data Snooping.” *Econometrica* 73(4): 1237–82.

_____. 2016. “Efficient Computation of Adjusted p -Values for Resampling-Based Stepdown Multiple Testing.” Working Paper.

Triyana, Margaret. 2016. “Do Health Care Providers Respond to Demand-Side Incentives? Evidence from Indonesia.” *American Economic Journal: Economic Policy* 8(4): 255-88.

UNICEF. 2013. *Improving Child Nutrition: The Achievable Imperative for Global Progress*. New York: UNICEF.

World Bank. 2018. *The State of Social Safety Nets 2018*. Washington, DC: World Bank.

Table 1: First-Stage Regressions, Household Level

Outcome: Received CCT	(1) 2-Year	(2) 6-Year
Treatment	0.375 (0.017)	0.368 (0.017)
Observations	14757	15667
R^2	0.258	0.242
Control Mean	0.091	0.131
F-statistic	507.797	456.783

Note: This table reports first-stage regressions of CCT receipt status (“Received CCT”) on baseline sub-district treatment assignment. Column 1 reports results from the 2-year follow-up survey and Columns 2-3 report results from the 6-year follow-up survey. In the final row, we report F-statistics from a Wald test of simple hypotheses involving the strength of our chosen instruments. Baseline controls include the following: household roof type, wall type, floor type, head of household’s education level, head of household works in agriculture, head of household works in services, household has clean water, household has own latrine, household has square latrine, household has own septic tank, household has electricity from PLN, log monthly per-capita expenditure, and log household size. Includes district fixed effects. Standard errors are clustered by sub-district.

Table 2: IV Effect of CCT on Health-Seeking Behaviors

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
Number of pre-natal visits	1.048 (0.473) [0.140] 6.493	0.560 (0.582) [0.706] 7.147	0.485
Delivery assisted by skilled midwife or doctor	0.115 (0.056) [0.166] 0.640	0.233 (0.059) [<0.001] 0.770	0.079
Delivery at health facility	0.112 (0.062) [0.230] 0.457	0.171 (0.066) [0.058] 0.725	0.430
Number of post-natal visits	0.842 (0.272) [0.023] 1.234	0.403 (0.317) [0.663] 1.778	0.275
90+ iron pills during pregnancy	0.025 (0.049) [0.831] 0.179	-0.035 (0.044) [0.706] 0.131	0.356
% of immunizations received for age	0.038 (0.029) [0.444] 0.754	0.048 (0.029) [0.427] 0.786	0.788
Times received Vitamin A (6 months - 2 years)	-0.022 (0.208) [0.903] 1.639	-0.095 (0.205) [0.706] 1.817	0.799
Times weighed in last 3 months (0-60 months)	0.919 (0.130) [<0.001] 1.791	0.250 (0.192) [0.663] 1.954	0.001

Note: Each row in this table represents a separate outcome variable. Each table entry includes: 1) the regression coefficient, 2) the cluster-robust standard error, 3) an adjusted *p*-value controlling the family-wise error rate within each column, as described by Romano and Wolf (2005; 2016), and 4) the control mean. Outcomes from “number of pre-natal visits” to “90+ iron pills during pregnancy” are coded for women in our sample who had been pregnant within the past two years. Outcomes from “% of immunizations received for age” onward are coded for children who were ages 0-36 months at baseline. These child-related regressions also include age-bin controls for each month of age up to 1 year, and for each quarter-year of age for ages 1 and above, in addition to baseline controls and fixed effects listed in Table 1. Standard errors, clustered by sub-district, are shown in parentheses.

Table 3: IV Effect of CCT on Incentivized Education Indicators

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
<i>Panel A: Enrollment for Ages 7-15</i>			
Enrolled in school (any level)	0.064 (0.013) [<0.001] 0.903	0.040 (0.012) [<0.001] 0.924	0.081
>85% attendance last two weeks	0.070 (0.016) [0.001] 0.830	0.057 (0.017) [<0.001] 0.856	0.492
<i>Panel B: Outcomes for Ages 7-12</i>			
Enrolled in school (any level)	0.037 (0.009) [<0.001] 0.960	0.012 (0.008) [0.181] 0.972	0.013
Enrolled in primary school	0.012 (0.014) [0.356] 0.887	0.011 (0.016) [0.505] 0.879	0.928
>85% attendance last two weeks	0.041 (0.016) [0.023] 0.881	0.034 (0.017) [0.102] 0.895	0.745
<i>Panel C: Outcomes for Ages 13-15</i>			
Enrolled in school (any level)	0.121 (0.032) [<0.001] 0.783	0.090 (0.027) [0.002] 0.826	0.383
Enrolled in secondary school	0.075 (0.037) [0.042] 0.585	0.054 (0.034) [0.106] 0.609	0.651
>85% attendance last two weeks	0.132 (0.033) [<0.001] 0.723	0.099 (0.029) [0.002] 0.777	0.364

Note: This table examines school enrollment and attendance outcomes. See Table 2 notes for explanation of table entries. Baseline controls and fixed effects are as listed in Table 1. *p*-values are adjusted within each panel rather than within entire columns. Standard errors, clustered by sub-district, are shown in parentheses.

Table 4: IV Effect of CCT on Child Nutrition and Health Outcomes, 0-60 Months

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
Stunted	-0.028 (0.035) [0.822] 0.513	-0.089 (0.039) [0.061] 0.390	0.229
Severely stunted	-0.023 (0.034) [0.855] 0.306	-0.100 (0.029) [<0.001] 0.180	0.072
Malnourished	-0.008 (0.028) [0.946] 0.332	-0.009 (0.033) [0.943] 0.274	0.981
Severely malnourished	0.004 (0.018) [0.946] 0.097	-0.003 (0.020) [0.943] 0.068	0.786

Note: This table examines child anthropometric outcomes. “Stunted” indicates children with height-for-age z -scores below -2, and “severely stunted” indicates children with height-for-age z -scores below -3. “Malnourished” indicates children with weight-for-age z -scores below -2, and “severely malnourished” indicates children with weight-for-age z -scores below -3. See Table 2 notes for explanation of table entries. Baseline controls and fixed effects are as listed in Table 1. Regressions also include age-bin controls for each month of age up to 1 year, and for each quarter-year of age between 1 and 5 years. Standard errors, clustered by sub-district, are shown in parentheses.

Table 5: IV Effect of CCT on Child Labor, Ages 13-15

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
Worked for wage last month	-0.041 (0.021) [0.050] 0.098	-0.044 (0.020) [0.057] 0.092	0.932
Worked 20+ hours for wage last month	-0.046 (0.016) [0.002] 0.061	-0.030 (0.017) [0.081] 0.055	0.382

Note: This table examines the effect of the conditional cash transfer on child labor outcomes based on survey responses. Outcomes are dummy variables indicating if children ages 13-15 performed any work for wage (or 20+ hours of wage work) in the past month. This definition does not include household labor. See Table 2 notes for explanation of table entries. Baseline controls and fixed effects are as listed in Table 1. Standard errors, clustered by sub-district, are shown in parentheses.

Table 6: IV Effect of CCT on Medium-Run Education, Work, and Marriage Outcomes

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
<i>Panel A: School Enrollment/Completion Outcomes</i>			
Enrolled in school (Ages 15-17)	0.069 (0.047) [0.234] 0.536	0.105 (0.045) [0.052] 0.616	0.481
Enrolled in high school (Ages 15-17)	0.016 (0.039) [0.648] 0.301	0.074 (0.041) [0.139] 0.393	0.207
Completed high school (Ages 18-21)		0.074 (0.041) [0.139] 0.258	
<i>Panel B: Labor Outcomes (Ages 16-21)</i>			
Worked for wage last month (Ages 16-17)	-0.068 (0.053) [0.286] 0.258	0.032 (0.041) [0.665] 0.221	0.104
Worked 20+ hours for wage last month (Ages 16-17)	-0.063 (0.049) [0.286] 0.188	0.004 (0.038) [0.914] 0.172	0.228
Worked for wage last month (Ages 18-21)		-0.059 (0.048) [0.482] 0.478	
Worked 20+ hours for wage last month (Ages 18-21)		-0.043 (0.047) [0.665] 0.423	
<i>Panel C: Marriage Outcomes (Ages 16-21)</i>			
Married (Ages 16-17)	-0.026 (0.020) [0.211] 0.041	-0.012 (0.025) [0.835] 0.056	0.624
Married (Ages 18-21)		-0.017 (0.036) [0.835] 0.186	

Note: This table explores schooling, labor, and marriage outcomes for children who were between the ages of 6-15 (i.e., schooling age) during the baseline survey and initial CCT rollout. Outcomes for ages 18-21 are omitted from Column 1 (2-year follow-up) because virtually none of these children had turned 18 by the time of the follow-up survey. See Table 2 notes for explanation of table entries. Baseline controls and fixed effects are as listed in Table 1. *p*-values are adjusted within each panel rather than within entire columns. Standard errors, clustered by sub-district, are shown in parentheses.

Table 7: IV Effect of CCT on Household Economic Outcomes

Outcome:	(1) 2-Year	(2) 6-Year	(3) p-value (2-Yr. = 6-Yr.)
<i>Panel A: Log Per-Capita Household Expenditure</i>			
Log per-capita expenditure	-0.006 (0.035) [0.951] 12.353	0.037 (0.037) [0.485] 12.898	0.329
<i>Panel B: Household Land + Livestock Investment</i>			
Owes any land	-0.011 (0.017) [0.770] 0.915	0.007 (0.021) [0.916] 0.909	0.406
Head of household employed	0.001 (0.014) [0.939] 0.940	-0.004 (0.011) [0.916] 0.943	0.793
Total number of livestock owned	-0.529 (0.468) [0.574] 3.883	-1.203 (1.575) [0.814] 4.753	0.670

Note: This table reports effects on various household-level consumption and investment outcomes. In Panels A and B, households above the 99th percentile for each category of expenditure are dropped from the regressions for that specific category. See Table 2 notes for explanation of table entries. Baseline controls and fixed effects are as listed in Table 1. *p*-values are adjusted within each panel rather than within entire columns. Standard errors, clustered by sub-district, are shown in parentheses.