When the Money Runs Out:

Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?

February 6, 2018

Sarah Baird  
Department of Global Health  
Milken Institute School of Public Health  
George Washington University  
950 New Hampshire Ave. NW  
Washington, DC 20052  
sbaird@gwu.edu  
202-994-0270

Craig McIntosh  
School of Global Policy and Strategy  
University of California, San Diego  
9500 Gilman Dr.  
La Jolla, CA 92093-0519  
ctmcintosh@ucsd.edu  
858-822-1125

Berk Özler  
Development Economics Research Group  
The World Bank  
1818 H Street NW, Mail Stop MC3-306  
Washington, DC 20433.  
Phone: 202-458-5861  
Email: bozler@worldbank.org
When the Money Runs Out:

Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?*

Sarah Baird, Craig McIntosh, and Berk Özler

February 6, 2018

Abstract

The five-year evaluation of a cash transfer program targeted to young women points to both the promise and limitations of cash transfers for persistent welfare gains. Conditional cash transfers produced sustained improvements in education and fertility for initially out-of-school females, but caused no gains in other outcomes. Significant declines in HIV prevalence, pregnancy and early marriage observed during the program among recipients of unconditional cash transfers (UCTs) evaporated quickly after the cessation of support. However, children born to UCT beneficiaries during the program had significantly higher height-for-age z-scores at follow-up pointing to the potential importance of cash during critical periods.

Keywords: Cash Transfers, Long-term Impacts, Human capital

JEL Codes: C93, I15, I21, I38, J12, J13

* Sarah Baird (corresponding author), email: sbaird@gwu.edu, George Washington University; Craig McIntosh, email: ctmcintosh@ucsd.edu, University of California, San Diego; Berk Özler, email: bozler@worldbank.org, The World Bank. We thank seminar and conference participants at the Center for the Study of African Economies, Columbia University, CU Denver, IFPRI, Mannheim, McGill University, Middlebury College, Monash Development Workshop, Oregon State University, Otago Development Workshop, Otago International Health Research Network, Labor Econometrics Workshop, PopPov Annual Research Conference, University of Maryland, University of Oklahoma, University of Oregon, University of Southern California, University of California, Berkeley, Washington Area Development Economics Symposium, Yale University, and the World Bank. We thank everyone who provided this project with great fieldwork and research assistance and are too numerous to list individually. We gratefully acknowledge funding from the Global Development Network, the Bill and Melinda Gates Foundation, 3ie, NBER Africa Project, and the World Bank. Ethical review committees at the National Health Sciences Research Council (Malawi, Protocol #569) the University of California at San Diego (USA, Protocol #090378), and George Washington University (USA, Protocol #061037) approved the study design. The trial is registered at AEA RCT Registry (#AEARCTR-0000036). The findings, interpretations, and conclusions expressed are entirely those of the authors.
1. INTRODUCTION

The past decade has witnessed an impressive growth in the number, volume, and types of cash transfer programs in developing countries. A rigorous evidence base has shown that cash transfers can have significant effects on household consumption and educational attainment, even if the poor receive these transfers with few strings attached (Fiszbein et al. 2009; Baird et al. 2013; Haushofer and Shapiro 2016; Garcia and Saavedra 2017). However, most of the evidence relies on short-term follow-ups, which leaves open the question of whether such programs can improve the wellbeing of their beneficiaries well after the cessation of support.1 This question is particularly pertinent for Conditional Cash Transfer (CCT) programs, which are built on the premise that they not only fight current poverty, but they also promote human capital accumulation for the next generation. As cash transfer programs continue to grow as major vehicles for social protection, it is increasingly important to understand if these programs break the cycle of intergenerational poverty, or whether the benefits simply evaporate when the money runs out.2

Few papers have empirically really tested this core premise because only a few programs were set up for rigorous long-term evaluation of their overall impacts (Molina-Millan et al. 2016). Even when researchers have examined longer-term effects of cash transfers for the transition from adolescence to adulthood, these studies have generally been limited to educational attainment and labor market participation. We build on the existing literature in two important ways. First, we are able to cleanly estimate the causal impact of a two-year CCT program targeted at adolescent females in Malawi more than two years after the program ended using both a pure experimental control group that never received treatment and a treatment group that was offered equal-sized unconditional cash transfers (UCT).3 Second, we collected data on a rich set of outcomes

---

1 Evaluations of government cash transfer programs that provide small, monthly, and often conditional transfers typically report 12- to 24-month impacts. One reason for the lack of evidence on longer-term impacts is the fact that most of the evaluations have a delayed treatment design, where all eligible households become part of the program within 1-2 years. This has caused researchers interested in longer-term effects to compare the outcomes of early vs. late treatment groups (see, e.g., Behrman, Parker and Todd (2011) and Gertler, Martinez and Rubio-Codina (2012)). In sub-Saharan Africa, cash transfer programs tend to be unconditional, targeting vulnerable households with children, although schooling conditions exist in some (see, e.g., The Transfer Project: https://transfer.cpc.unc.edu/). Evaluations of these programs have similar durations to CCT programs (see, e.g., Handa et al. (2016)).

2 There is a recent wave of transfer programs, generally conducted by NGOs, which aim to lift households out of poverty using larger lump-sum transfers during a limited period of support (Banerjee et al. 2015; Haushofer and Shapiro 2016). Evaluations of these programs are generally concerned with current poverty reduction rather than human capital accumulation among children. As such, while the question of sustained effects is also pertinent for these studies, they are less relevant for our examination of longer-term impacts on adolescent beneficiaries.

3 To our knowledge, the only other CCT evaluation that examines longer-term effects in comparison to an experimental control group that was never treated is Barrera-Osorio, Linden and Saavedra (2017), which examines
(education, childbearing and marriage, health, labor market outcomes, empowerment, and subjective wellbeing) for the target population of young females, as well as information about their children and husbands as these young women gave birth and married. The resulting analysis is a comprehensive assessment of the relative effects of CCTs and UCTs targeted to adolescents for two years during a period of transition into adulthood.\textsuperscript{4}

For any intervention to have a sustained effect, it needs to lead to an increase in the stock of some asset that produces a stream of returns in the future, i.e. some accumulation of capital – whether it takes the form of human, physical, or social capital. However, the causal pathway from program implementation to final outcomes can be circuitous. For example, a program that provides cash grants to groups of unemployed youth for income generating activities may have lasting effects on earnings through the accumulation of physical (productive assets) and human (vocational skills) capital (Blattman, Fiala and Martinez 2014). Alternatively, large unconditional cash grants to poor households may increase future earnings by increasing investments in productive assets, such as livestock (Haushofer and Shapiro 2016); while small monthly conditional cash transfers over a long period of time may lead to increased consumption after beneficiaries exit the program by increasing savings and investments in small-scale agriculture (Gertler, Martinez and Rubio-Codina 2012) or by stimulating entrepreneurial activity (Bianchi and Bobba 2013).

For programs targeting younger people, the causal pathway to improved welfare over the long run is more likely to be human capital accumulation, either in the form of education and skills or health – especially reproductive and sexual health for adolescent females. Even when young women attain higher schooling and delay childbearing and marriage, low quality education, credit education outcomes eight and twelve years after treatment.\textsuperscript{4} Cash transfers during adolescence may be particularly effective as this is a critical period to expand one’s capabilities by investing in human capital. In fact, adolescent girls are viewed as a key demographic target group to successfully break the intergenerational transmission of poverty in developing countries (Levine et al. 2008). Unfortunately, for many boys and girls in developing countries, adolescence entails a fleeting transition from childhood to adulthood, when they are suddenly expected to “behave as adults even though they are not biologically, cognitively, or emotionally ready to assume adult responsibilities” (Naudeau, Hasan and Bakilana 2015). Adolescent females in particular face a multitude of hazards – ranging from school dropout, to child marriage and teen pregnancy, to physical and mental health problems, to gender based violence (Baird and Özler 2016). Young people’s capabilities and functionings (Heckman and Corbin 2016) during this period not only have immediate consequences to their own lives, but also longer-term benefits to their offspring and communities at large (Lloyd and Young 2009; Duflo 2012). Interventions that help adolescent girls reach their full potential by increasing their education, improving their skills, and delaying childbearing have the potential to create a virtuous cycle that improves health, especially child health, and women’s empowerment – ultimately leading to higher economic growth (Canning, Raja and Yazbeck 2015).
constraints, and low demand for skilled labor can stunt income gains. Without economic independence, women cannot attain higher agency, intra-household bargaining power, and empowerment.

Programs targeted to adolescent girls may not only delay marriage and childbearing, but may also benefit the development of their own children. A distinct and mostly U.S.-based literature, largely using quasi-experimental methods, has examined the very long-term effects of being exposed to cash, ‘near cash,’ or other safety net programs during childhood (e.g. Currie and Almond 2011; Aizer et al. 2016; Hoynes, Schanzenbach and Almond 2016; Chetty, Hendren and Katz 2016) and has demonstrated beneficial effects on a host of outcomes as adults.

In this paper, we report the effects of a cash-transfer experiment more than two years after it ended, tracking a broad range of outcomes for females aged 18-27. Our earlier work has demonstrated the short-term effectiveness of cash transfers in improving school participation and test scores, as well as reducing the incidence of pregnancy, marriage, psychological distress, and sexually transmitted infections during adolescence, indicating the possibility of finding longer-term improvements in well-being as young adults (Baird, McIntosh and Özler 2011; Baird et al. 2012; Baird, De Hoop and Özler 2013). Here, following a pre-analysis plan, we first examine human capital accumulation, marriage and fertility, labor market outcomes, and empowerment among the beneficiaries to assess the persistence of the short-term effects. Then, as the majority of the study participants were married and/or had children at the latest follow-up, we examine their marriage market outcomes and their children’s physical development using data we collected on their husbands and anthropometric measurements of their children.

We find that the short-term improvements in the UCT arm observed during the program failed to translate into increased welfare in the longer-term. Substantial reductions in teen marriages, total live births, and HIV infections, as well as improvements in psychological wellbeing and nutritional intake observed at the end of the program were no longer apparent two years after the end of the intervention. In this group, the end of the cash transfer program was immediately followed by a marriage and baby boom among the beneficiaries, who reported lower levels of empowerment and had husbands with lower cognitive ability compared with both the

---

5 At baseline, our target population was never-married females, aged 13-22.
6 This result adds to a growing body of evidence across diverse programs and contexts on the lack of sustained impact of UCTs except under very specific conditions. See for example Brudevold-Newman et al. (2017) and Hicks et al. (2018).
CCT and the control groups. However, consistent with improved physical and mental health during the program, we find evidence of improved height-for-age z-scores (HAZ) among children born to the beneficiaries during the program.

CCTs, on the other hand, caused sustained effects on school attainment, incidence of marriage and pregnancy, age at first birth, total number of births, and desired fertility – but only among the stratum of adolescent females who had already dropped out of school at baseline and were all assigned to CCTs. Conditional transfers were highly effective in allowing a very large share of this group to return to school.\(^7\) In contrast with the marital outcomes in the UCT group, the increased educational attainment in this group was accompanied by assortative matching: their husbands were significantly more likely to have completed secondary school. However, even in this group, we find no gains in other important outcomes, such as individual earnings, per capita household consumption, subjective wellbeing, health, or empowerment. Among the stratum in school at baseline, CCTs did not have any lasting effects, positive or negative, likely because the transfers were mostly inframarginal with respect to school attainment: 88% of the control group in this stratum completed primary school two years after the end of the program.

Our paper speaks to a number of distinct literatures. First, it adds to a growing literature on the medium- to long-term effects of cash transfer programs in developing countries.\(^8\) There are now several longer-term evaluations of cash transfers programs (mostly of CCTs) that indicate that while cash transfer programs might improve school attainment among adolescent beneficiaries, gains in terms of learning, employment, and income are limited or non-existent as they become young women (Baez and Camacho 2011; Behrman, Parker and Todd 2011; Barham, Macours and Maluccio 2013; Filmer and Schady 2014; Araujo, Bosch and Schady 2016).\(^9\)\(^10\)

---

\(^7\) These findings align nicely with Duflo, Dupas and Kremer (2017) who find that by age 25 Ghanaian students who were offered a secondary school scholarship were 26 percentage points more likely to complete secondary school and had 0.217 fewer children—again suggesting the importance of the magnitude of the education effect.

\(^8\) It also builds on Baird, McIntosh and Özler (2011) and adds to the small literature that directly compares CCTs with UCTs either experimentally (Akresh et al. 2013; Benhassine et al. 2015) or quasi-experimentally (Schady and Araujo 2008; de Brauw and Hoddinott 2011; Attanasio et al. 2015).

\(^9\) The evaluation of a school-based intervention in Kenya testing the effects of education subsidies found significant reductions in school dropout, pregnancy, and marriage among girls in the short- and medium-run, and school attainment, marriage, and childbearing by age 16 in the longer-run (Duflo, Dupas and Kremer 2015). Molina-Millan et al. (2016) provides a review of longer-term effects of CCTs in Latin America and finds that the evidence is mixed. Molyneux, Jones and Samuels (2016) strike a similarly cautious tone about the transformative effects of social protection programs.

\(^10\) There is a vast literature on the effects of programs for pregnant women and mothers on child outcomes. Manley, Gitter and Slavchevska (2013) provide a review of the effects of cash transfers on children’s nutritional status in low- and middle-income countries (LMIC).
finding that CCT programs can substantially increase school attainment among vulnerable populations without substantive effects on test scores, cognitive skills, employment, or earnings is consistent with these results.\textsuperscript{11}

Second, we add to the literature on the effects of human capital accumulation and increased age at marriage on marriage market outcomes (Field and Ambrus 2008; Anderson and Bidner 2015; Ashraf et al. 2016). Theory suggests that these two factors affect spousal quality in opposite directions: increased education typically having been found to improve marital outcomes while delaying marriage worsens them, \textit{ceteris paribus}. Our findings provide empirical support for these predictions.

Third, our study contributes to a large literature on the effects of programs that support pregnant women and young children. Policies for child development often target the first 1,000 days from conception to the second birthday (Barham, Macours and Maluccio 2013). What is novel in our study is that we examine the effects of targeting cash transfers to adolescent females of childbearing age and provide evidence on the important policy question of how to time interventions to protect early childhood development.\textsuperscript{12} Our findings suggest that unconditional income support for adolescent girls and young women of childbearing age might cause significant increases in height-for-age z-scores of their children.

The remainder of this paper is structured as follows. Section 2 describes the study setting, study design, and data collection instruments. Section 3 presents our estimation strategy. Sections 4 presents program impacts on the core respondents, followed by an examination of some key characteristics of their husbands and children. Section 5 concludes.

\textbf{2. Study Setting, Design, and Data Sources}

\textbf{2.1 Study Setting}

The “Schooling, Income, and Health Risk” study (SIHR) follows young women who were enrolled as never-married 13-22 year olds in Zomba, Malawi in 2007. We interviewed them for the fourth time in 2012 – approximately five years after baseline and more than two years after the cessation of the cash transfer experiment in December 2009, tracking the adolescents as many of

\textsuperscript{11} There are also mixed results from programs that target adolescent females directly such as programs that provide a safe space to meet on a regular basis and develop life skills (Bandiera et al. 2017; Buehren et al. 2017) or programs that combine training and mentoring with financial incentives to delay marriage (Buchmann et al. 2017).

\textsuperscript{12} Currie and Almond (2011) state “…one of the more effective ways to improve children’s long term outcomes might be to target women of child bearing age in addition to focusing on children after birth.”
them moved on to establish their own families. These longitudinal data paint a very rich picture of the transition from adolescence into adulthood in this context. By 2012, in the control group, the study stratum that had dropped out of school at baseline had effectively completed their schooling with an average of a seventh-grade education; 81% were married, 92% had been pregnant, and only 6% had spent any time in self-employment or paid work during the past week. More than one in eight (13.5%) had been infected with HIV. The stratum of baseline schoolgirls are better-off and younger, and therefore had not proceeded as far in their transition to adulthood: in 2012, their average years of schooling was 10.4 and increasing, with only 40% ever married, 50% ever pregnant, and 5.5% HIV-positive.

In the latest follow-up survey of the study sample, which was more than two years after the cessation of cash transfers, we attempted to trace the pathways through which experimentally induced changes in human capital may translate into longer-term outcomes. Zomba is an almost exclusively agricultural economy characterized by low educational attainment and few opportunities for formal employment. As of 2009, this district was the third poorest in Malawi (in our sample, real monthly per-capita exchange rate comparable consumption in 2008 was USD 20.6). Secondary school completion rates are low – in our sample, among baseline schoolgirls, half of whom had completed primary school at baseline, only 17.0% had completed secondary school as of 2012. Although most adults 15 and over participate in some form of employment, the majority do not receive a formal income. In 2008, only 6% of the adult population in Zomba received a formal income (Zomba City Assembly 2009), a number that is reflected in our data with 6% of baseline dropouts and 3% of baseline schoolgirls participating in any formal work. This context is typical for many parts of rural Africa, and, hence, is an important environment in which to understand the constraints adolescents face as they transition to adulthood.

2.2 Study Design

Our study began by listing all eligible households within 176 Enumeration Areas (EAs) of the 550 EAs in Zomba District. This never-married, 13-22 year-old target population was then divided into two main strata: those who were already out of school at baseline (baseline dropouts) and those who were still in school at baseline (baseline schoolgirls). Baseline dropouts comprised only 15% of target population, so were all recruited into the study. Baseline schoolgirls were sampled into the study at probabilities increasing in age and rural status.
Treatment was assigned first at the enumeration area (EA) level; 88 to treatment and 88 to control. All baseline dropouts in treatment EAs received CCTs, while we experimented with attaching conditions to the cash transfers within the larger cohort of baseline schoolgirls. For them, 46 EAs were assigned to CCTs, 27 were assigned to UCTs and 15 were assigned to receive no transfers in order to study spillovers (from baseline dropouts in those EAs). The amount of money received by the household head was randomized between $4 and $10 at the EA level, and the core respondents were assigned their own individual transfer amounts – ranging from $1 and $5 – in a public lottery. The share of eligible girls offered cash transfers was randomly varied across clusters to estimate spillover effects. Offer letters were distributed in December 2007, payments began in February 2008 and continued through the end of 2009. Four rounds of data took place: Round 1-Baseline (2007), Round 2 (2008), Round 3 (2010), and Round 4 (2012). Figure I presents an illustration of the study design, and a more detailed description of the experiment can be found in (Baird, McIntosh and Özler 2011).

Girls receiving UCTs simply had to show up at a local distribution point each month to pick up their transfers. Monthly school attendance for all girls in the CCT arm was checked and payment for the following month was withheld for any student whose attendance was below 80% of the number of days school was in session for the previous month. However, participants were never removed from the program for failing to meet the monthly 80% attendance rate, meaning that if they subsequently had satisfactory attendance, their payments would resume. Other design aspects of the program were kept identical so as to be able to isolate the marginal effect of imposing a schooling conditionality on outcomes of interest among baseline schoolgirls.

13 The average total transfer to the household of $10/month for 10 months a year is nearly 10% of the average household consumption expenditure of $965 in Malawi in 2009 (World Bank, 2010). This falls in the range of cash transfers as a share of household consumption (or income) in other countries with similar CCT programs. The transfers were offered to all eligible girls in our target demographic and were not targeted by poverty status.

14 In experiments like SIHR, it is important to try to understand what the beneficiaries expected as to the program’s timing and duration (Bazzi, Sumarto and Suryahadi 2015). When the initial offers were made, the beneficiaries were told that the program only had funding for one year, but that efforts were being made to extend it into a two-year program. Towards the end of the first year, upon successfully obtaining additional funding, we circulated new offer letters informing the beneficiaries that the program would be continued for one more year, but not more. This message was repeated regularly at the cash distribution points by the program staff during the second and final year of the intervention.

15 The size of the transfers, the identity of the recipients, or the intensity of treatment within the cluster did not prove to be influential on the primary outcomes of interest. Because these were randomized across the control, CCT, and UCT arms, estimates of average treatment effects remain highly robust to these controls.

16 For households with girls eligible to attend secondary schools at baseline, the total transfer amount was adjusted upwards by an amount equal to the average annual secondary school fees in the conditional treatment arm. This
2.3 Data Sources and Outcomes

The focus of this paper is data collected in Round 4, which took place in 2012, more than two years after the end of the intervention. However, to provide context to these results, we also present impacts on the same outcomes, when available, for data collected during Rounds 2 and 3. Focusing on the core respondent, the data sources include household surveys (all rounds), biomarker data collection on HIV (Round 2-4) and Anemia (Round 4), and competencies (Round 4). In Round 4, data collection also included anthropometric data (children under 60 months of age) and early child development tests (children 36-59 months old) among the children of study participants, as well as a survey and biomarker data collection among their husbands.

The household surveys at each round consisted of a multi-topic questionnaire administered to the households in which the core respondents resided during the data collection period. They consisted of two parts: one that was administered to the head of the household and the other administered to the core respondent. The former collected information on the household roster, dwelling characteristics, household assets and durables, shocks, and consumption. The survey administered to the core respondent collected detailed information about her family background, schooling status, health, dating patterns, sexual behavior, fertility, marriage, labor market outcomes, and empowerment. In addition to the household survey administered to the core respondent (and to her parents/guardian if she still lived with them), the Round 4 survey included a similar module administered to the husbands of married study participants.

The Round 4 household survey also consisted of a test to measure basic labor market skills of the core respondent, which we termed “competencies.” It included reading and following instructions to apply fertilizer; making correct change during a hypothetical market transaction; sending a text message and using a calculator on a mobile phone, and calculating profits for a hypothetical business scenario. As Round 4 was focused more on the transition into adulthood and labor markets, as opposed to the school attainment and learning focus in Round 3, this test was designed to replace the reading comprehension, math, and cognitive skills tests utilized in Round 3, and serve as a measure of a more practical set of skills that might be influenced by increased schooling and needed in the labor market.

---

additional amount ensured that the average transfer amounts offered in the CCT and UCT arms were identical and the only difference between the two groups was the “conditionality” of the transfers on school attendance.
Home-based voluntary counseling and testing for HIV (for core respondents during Rounds 2-4, and their husbands in Round 4) was conducted by Malawian nurses and counselors certified in conducting rapid HIV tests through the Ministry of Health HIV Unit HCT Counselor Certification Program. In addition they tested for hemoglobin and measured the height and weight of all children aged 59 months or younger.

Early childhood development (ECD) tests were administered to all 36-59 month-old children of the study participants. These tests consisted of the Malawi Development Assessment Tool (MDAT) for fine motor skills, language, and hearing, which were administered directly to the child (Gladstone et al. 2008) and the Strengths and Difficulties Test (SDQ), administered to the mother or the guardian responsible for the child (Goodman 2001; Woerner et al. 2004).

Prior to the analysis of data from Round 4, a pre-analysis plan was registered at the AEA RCT Registry (AEARCTR-0000036; https://www.socialscienceregistry.org/trials/36). Our outcomes cover six domains for the core respondent – education and competencies, marriage and fertility, health and sexual behavior, empowerment and aspirations, employment and wages, and consumption – and outcomes in relevant domains for their husbands and children.

3. Estimation Strategy

In this section, we discuss the experimental estimation strategy used to examine program impacts on core respondents. The causal identification of program impacts on husband characteristics and children’s outcomes is more challenging and the estimation strategies used to analyze those outcomes are discussed in Sections 4.6, 4.7, and Appendix A.

To estimate intention-to-treat effects of the program in each treatment arm on our primary outcomes by stratum, we employ a simple reduced-form linear model:

$$Y_{ic} = \alpha + \gamma^c T^c_c + \gamma^u T^u_c + \beta X_{ic} + \varepsilon_{ic}$$

where $Y_{ic}$ is an outcome variable for core-respondent $i$ in cluster $c$, $T^c_c$ and $T^u_c$ are binary indicators for offers in the CCT and the UCT clusters, respectively, and $X_{ic}$ is a vector of baseline

---

17 You can find the pre-analysis plan here: [https://drive.google.com/file/d/1ZAG3WBN7GVNqA21O6DXHNQeYfY8ulaRX/view](https://drive.google.com/file/d/1ZAG3WBN7GVNqA21O6DXHNQeYfY8ulaRX/view)

18 Many of our outcomes are in the form of indexes that are constructed using the following rubric: First, we ensured that all sub-questions are aligned so that higher scores always have a consistent meaning (good or bad). We then calculated the mean and standard deviation of the responses to each sub-question in the control group – separately for baseline schoolgirls and baseline dropouts. We then normalized each sub-question by subtracting the mean and dividing by the standard deviation. Finally we constructed (and then normalized) the raw mean of the normalized variables for all sub-questions within a family of variables to create the final index.

19 A detailed description of all outcomes in this paper can be found here: [https://drive.google.com/file/d/1hvI79ltywocFr-pafqz8_Dtg2ZXXhcHd/view](https://drive.google.com/file/d/1hvI79ltywocFr-pafqz8_Dtg2ZXXhcHd/view)
characteristics. Note that for baseline dropouts we only have the CCT binary indicator. The standard errors $\varepsilon_{ic}$ are clustered at the EA level, which account for both the design effect of our EA-level treatment and the heteroskedasticity inherent in the linear probability model.

In all regressions, we include baseline values of the following pre-specified variables as controls: a household asset index, highest grade attended, a dummy variable for having started sexual activity, and dummy variables for age in years. These variables were chosen because they are strongly predictive of schooling outcomes, hence improving the precision of the impact estimates. We also include indicators for the strata used to perform block randomization – Zomba Town, within 16 kilometers of the town, and beyond 16 kilometers (Bruhn and McKenzie 2009). Age- and stratum-specific sampling weights are used to make the results representative of the target population in the study area.

Appendix Table S1 presents means and standard deviations for nine individual or household characteristics for the study sample at baseline by strata and treatment assignment. As this paper is mainly about program effects more than two years after the end of cash transfers, we conduct all analysis among those who were successfully interviewed in Round 4, which maximizes sample size for the estimation of longer-term impacts. Columns 1 and 2 show descriptive statistics for baseline dropouts, who are older than baseline schoolgirls and come from more disadvantaged backgrounds: for example, 44.5% of the control group had started childbearing at baseline compared to only 2.1% of baseline schoolgirls. In addition to the fact that all baseline dropouts are out of school at baseline and never married, there are no statistically significant differences between the CCT and the control groups for the variables presented in Appendix Table S1. Nor are there any differences between the two treatment groups and the control group among baseline schoolgirls, but the UCT group is, on average, older and has attended higher grades than the CCT group at baseline. Note that this imbalance existed at baseline and is not a result of differential attrition (Baird, McIntosh and Özler 2011). Pre-specified baseline controls used in all impact regressions described above include these two variables. Joint tests of orthogonality presented at the bottom of Appendix Table S1 confirm these findings.

---

20 Conducting the analysis among the Round 4 sample implies that the Round 2 and Round 3 samples are smaller than the Round 4 sample in the analysis. For example, to be included in the Round 3 analysis of impacts, a subject had to be successfully interviewed in both Rounds 3 and 4. In addition to maximizing the sample for Round 4 analysis, which is the focus of this paper, this allows us to demonstrate that the Round 2 and Round 3 impacts, which were reported in earlier publications, hold in this sub-sample and provides some reassurance that differential attrition is not substantially affecting our findings at Round 4.
Appendix Table S2 examines attrition for the same sample of core respondents who were successfully interviewed in Round 4 – first for *baseline dropouts*, then *baseline schoolgirls*. Attrition two years after the end of the cash transfer program is 15.7% in the control group among *baseline dropouts* and this level of attrition is not differential in the CCT arm (column 1). However, interacting attrition with the same pre-specified baseline adjustments used throughout the paper, we find that these interactions are jointly significant (column 2) – primarily due to the fact that CCT beneficiaries in urban areas, which constitutes less than 20% of our sample, were more likely to be lost to follow-up. Attrition in the control group among *baseline schoolgirls* is slightly lower at 12.5%, which is significantly higher than both the CCT and UCT arms (column 3). However, attrition in this stratum is not differential by baseline characteristics between treatment and control, although the F-test for joint significance of UCT interactions is 0.101 (column 4). Furthermore, there is no differential attrition between the CCT and UCT arms – either in levels or by characteristics.

While overall attrition is low, to address any potential bias in impact estimates due to differential attrition by treatment arm – either in levels (CCT and UCT among *baseline schoolgirls*) or in baseline characteristics (CCT among baseline dropouts), we include a thorough analysis of the robustness of our impact estimates in Section 4.5 below. There, we present upper and lower bounds on impact estimates for all primary outcomes (Lee 2009), as well as adjusted estimates using inverse propensity weighting. We also note that impact estimates from earlier follow-up rounds, which did not suffer from differential attrition, replicate in the Round 4 sample used in this paper.

4. Results

We start by presenting the trajectory of program effects on outcomes in four domains, separately for baseline dropouts and baseline schoolgirls: education and competencies, marriage and fertility, health, and, finally, labor market participation and empowerment.21

4.1 Education and Competencies

Table I presents program impacts on highest grade completed and competencies. Among *baseline dropouts*, CCTs led to an increase in highest grade completed of approximately 0.6 years, 21 The reader should note that most of the one- and two-year impacts during and at the end of the program were reported in previous publications, which are clearly cited throughout the paper. What are new here are the findings from two years after the end of the program. Presenting program impacts over time within each domain allows the reader to examine the trajectory of program effects and assess whether earlier impacts were sustained.
which represents a 0.22 standard deviation (SD) increase by Round 4 (Panel A). As a result, the share of beneficiaries with a Primary School Leaving Certificate (PSLC) increased by 5.8 and 8.1 percentage points in Rounds 3 and 4, respectively (Appendix Table S3, Panel A). However, earlier gains in test scores of English reading comprehension, mathematics, and cognitive skills (Table I, columns 4-7) did not translate into increased scores in tests of basic labor market skills, or “competencies,” such as following instructions to apply fertilizer or calculating change in a market transaction (Table I, column 8).

The results for baseline schoolgirls suggest little, if any, effect on school attainment or competencies in either treatment group (Table I, Panel B). Any significant effect in the CCT group at the end of the program was no longer detectable two years later. The reader should note that the mean number of years completed in the control group is 10.4 in Round 4, at which point 88% of the control group had obtained a PSLC (Appendix Table S3, Panel B). Hence, while most of the transfers to baseline schoolgirls were inframarginal with respect to primary school completion, the cash transfer program did not cause any significant gains in secondary school completion either. Similarly, earlier gains in test scores in the CCT group did not translate into improved competencies in the longer-run, with the only significant improvement seen in the UCT group being the ability to send a simple text message using a mobile phone.

The consistent pattern in the CCT arm (for both baselines schoolgirls and dropouts) of short-term improvements in test scores combined with no improvement in long-run competencies has two potential explanations. One of these is that the competencies simply failed to measure variation in skills in a useful way. However, we find this explanation unlikely as the variation in schooling and test scores at the end of the intervention are strongly predictive of competencies two years later: for example, a one year increase in highest grade completed is associated with a 0.21 SD increase in the overall competency score. Mechanically, this would imply an improvement of only 0.13 SD in the overall competency score among baseline dropouts (0.621 x 0.21 = 0.13), which is twice as large as our point estimate of 0.064 SD but within the 95% confidence interval. The more likely explanation is that even though CCTs caused large effects on school attainment and modest ones in test scores by the end of the intervention among baseline dropouts, these learning gains were too small and dissipated within two years.
4.2 Marriage and Fertility

As with the education outcomes, CCTs had large effects on marriage and fertility for baseline dropouts that were sustained at Round 4 (Table II, Panel A). They were 14.0, 15.7, and 10.7 percentage points (pp) less likely to have been ever married at Rounds 2-4, respectively (all significant at 99% confidence). The corresponding reductions were 5.7, 8.1, and 4.0 pp for being ever pregnant (all significant at 90% confidence or higher). Furthermore, there is a negative fertility gradient among CCT beneficiaries, leading to a reduction of 0.147 total live births at Round 4 (p-value < 0.001), which corresponds to a reduction of more than 10% and is consistent with the reduction in stated desired fertility. Age at first marriage and first birth were similarly higher by 0.43 and 0.27 years, respectively.

Among baseline schoolgirls, CCTs had no effects on marriage and fertility at any point during our study period (Table II, Panel B). On the other hand, UCTs were very effective in substantially reducing marriage and pregnancy rates among baseline schoolgirls during and at the end of the program (Baird, McIntosh and Özler 2011). Two years later, there are no longer any differences in ever married, ever pregnant, total number of live births, or even age at first birth between the UCT group and either the control group or the CCT arm. We find that the age at first marriage is delayed by half a year as of Round 4, which is consistent with the fact that girls in the UCT arm who delayed marriage got quickly married following the end of the intervention. Striking spikes in pregnancies and marriages in the UCT group immediately following the end of the transfers are shown in Figure II. The temporary nature of the fertility changes in this group is also reinforced by the fact that desired fertility remains unchanged (Table II, Panel B, column 12). In analysis not shown in the tables, we find that teen pregnancy (defined as starting childbearing at age 18 or younger) was significantly lower in the UCT arm at round 3 (3.8 pp, p-value =0.027) but that this effect had also shrunk by two thirds and was no longer significant by round 4. Beneficiaries of all ages experienced spikes in marriage and pregnancy following the program.

22 The finding of null effects in Round 4 in the UCT arm is not simply a function of lack of power. While the standard errors of binary indicators for marriage and pregnancy are higher in Round 4 than in Round 3 due to the fact that the control means for these variables are increasing towards 0.5 over the course of our study, minimum detectable effects as a percentage of the mean in the control group are actually lower. Furthermore, these minimum detectable effects are comparable to or lower than those presented in similar papers, such as Bandiera et al. (2017). Finally, many of the significant effects among baseline dropouts that we present in Table II are larger than the minimum detectable effects among baseline schoolgirls.
meaning that UCTs reduced the prevalence of neither teen pregnancies nor child marriages by Round 4 – despite large reductions in these quantities at Round 3.

Cash transfers can have effects on marriage and fertility via two channels. The first pathway, apparent in the UCT arm, is through an income effect. In our study, this effect is strong but disappears immediately when the transfers stop – as the transfers have not led to any accumulation of physical or human capital. The other pathway, apparent in the CCT arm among baseline dropouts, is through increased schooling. Increased schooling is strongly associated with delays in marriage and childbearing and reductions in desired and total fertility, but the impacts of transfer programs on schooling have to be substantial to translate into meaningful and statistically significant knock-on effects on marriage and fertility.

4.3 Health

Table III presents program impacts on biomarkers for HIV and anemia – the primary health outcomes specified in our pre-analysis plan. Program effects on HIV prevalence during the program, i.e. at Round 2, were reported in Baird et al. (2012). Despite the improvements in education, delays in marriage and fertility, and the high prevalence of HIV among baseline dropouts (13.5% by Round 4), CCTs did not reduce HIV prevalence in this stratum at any point during the study period (Panel A). Appendix Tables S4 and S5 examine self-reported sexual behavior on the extensive and intensive margin. Both the onset of sexual activity and the likelihood of being sexually active during the past year were lower among program beneficiaries during and immediately after the program, but not two years later. There were no effects on risky sexual behavior, such as having older partners or use of condoms, among those who reported being sexually active. Nor did CCTs have significant effects on psychological wellbeing or nutritional intake (Appendix Table S6).

Among baseline schoolgirls, program impacts on HIV mirror those on marriage and fertility over time: there is no effect of CCTs on HIV at Rounds 3 or 4, but a more than 50% reduction in HIV prevalence in the UCT group at the end of the intervention is no longer there two years later (Table III, Panel B). During the two-year post-intervention period, which saw a spike in pregnancies and marriage in the UCT group, the incidence of HIV was 3.5 percentage points (pp) – compared with 2.0 pp in the control group, but this difference in HIV incidence is not statistically significant. Appendix Table S6 shows that effects of cash transfers were equally transient on mental health and nutritional intake – strongly evident during the program and
disappearing afterwards. There is weak evidence of lower anemia prevalence in the UCT arm in Round 4, but the UCT effect on a continuous measure of hemoglobin levels does not corroborate this finding. Nor does it hold up to multiple hypothesis testing discussed in Section 4.5.

**4.4 Labor Market Participation and Empowerment**

Hardly anyone in our sample spent a significant amount of time in self-employment or paid work during the past week (Table IV, column 3), consistent with labor market conditions in Zomba. Only a third of baseline dropouts and a quarter of baseline schoolgirls report having done any wage work in the past three months (Appendix Table S7). The main activities performed by the *baseline dropouts* in our sample are household chores – such as cooking and cleaning, fetching water and firewood, and looking after children – (69.6%) and subsistence agriculture (19.4%); among *baseline schoolgirls*, 55.2% report household chores as their main activity, 11.1% report subsistence agriculture, while 27.5% are still in school. There are no significant effects on primary outcomes in either stratum, except a negative effect on typical wage among *baseline dropouts*, which may reflect the fact that individuals in the treatment group were in school longer, and thus might have less work experience. Program impacts on secondary labor market outcomes, such as the effective daily wage, labor income in the past five seasons, and any wage work in the past three months, are similarly null (Appendix Table S7).

For *baseline dropouts*, program impacts on empowerment echo those on competencies, health, and labor market participation: despite significant gains in educational attainment, delays in marriage and pregnancy, and reductions in total live births, there are no effects on the overall index of empowerment or subjective welfare (Table IV, Panel A, columns 4 & 5). This finding holds when we examine empowerment by marital status at Round 4 (columns 6 & 7). Appendix Table S8 shows results by the components of the female empowerment index (self-esteem, social participation, preferences for child education, and aspirations).

For *baseline schoolgirls* in the CCT group, we also see no significant impacts on empowerment or subjective wellbeing, although the coefficient estimates are generally positive. However, in the UCT arm, the empowerment index is significantly lower than both the control and the CCT groups (Table IV, Panel B). The -0.159 SD effect (p-value=0.05) on the super-index of overall empowerment among the UCT beneficiaries is reflected in the negative (but insignificant)

---

23 We also examined accumulation of savings, household assets, and productive assets (such as livestock). We find no treatment effects on any of these outcomes in either stratum.
effects in all sub-indices except aspirations (Appendix Table S8, Panel B), and is driven mainly by a large (-0.342 SD; \(p\)-value<0.01) and significant negative effect on empowerment among those who are married (Table IV, Panel B, column 7). The findings indicate a statistically significant divergence in female empowerment between CCT and UCT recipients among baseline schoolgirls two years after the end of the cash transfer program – particularly for those married by Round 4. We further explore these negative impacts on marital empowerment by directly studying husband characteristics below in Section 4.6.

4.5 Robustness of Findings to Attrition and Multiple Hypothesis Testing

Before we move on to analyzing husband and child outcomes, we examine the robustness of program impacts for the young women targeted by our cash transfer program. There are two issues that raise doubt about the findings we presented so far. First, in Section 3, we showed that while the share of our study sample lost to follow-up more than four years after baseline data collection is not high (between 12.5% and 15.7% in the control groups of the two strata), there is evidence of differential attrition in levels (but not characteristics) among baseline schoolgirls, and vice versa among baseline dropouts. As differential attrition has the potential to bias impact estimates and, as such, is a threat to causal inference, we conduct additional analysis to test the robustness of our findings. Second, although we follow a pre-analysis plan, we nonetheless present 14 primary outcomes in Round 4. To allay concerns that some of the statistically significant impacts estimates might have occurred due to chance, we present \(p\)-values for impact estimates that are adjusted for the false discovery rate (FDR).

In Appendix Tables S9-S11, we report the original impact estimates for the 14 primary outcomes presented in Tables I-IV (column 1), along with estimates adjusted for inverse probability weighting (IPW, column 2), as well as lower and upper bound estimates (columns 3 and 4) following Lee (2009). The IPW adjustment is implemented by regressing an indicator variable for being successfully interviewed in Round 4 on treatment indicators, baseline characteristics (the same pre-specified ones used for regression adjustment throughout the paper), and their full interactions. Each individual’s propensity to be part of the Round 4 sample is predicted and impact regressions described in equation (1) are weighted by the inverse of this probability. Lower and upper bound impact estimates are obtained by trimming the sample (from above and below) such that the share of individuals lost to follow-up is equal in study arms.
For *baseline dropouts*, we note that the Lee bounds are tight around the original estimate because the difference in the level of attrition between the control and the CCT groups is very small (Appendix Table S9). Furthermore, IPW-adjusted impact estimates are very close to our original estimates. Nothing in the table suggests that we should significantly revise our interpretation of the key findings of program impacts among *baseline dropouts*. Similarly, for *baseline schoolgirls*, IPW-adjusted estimates are nearly indistinguishable from the original estimates, while the Lee bounds are wider because of the larger difference in attrition levels between the control group and either treatment group (Appendix Table S10). These wider bounds mean that while our original and IPW-adjusted estimates generally indicate a lack of impact of CCTs or UCTs among *baseline schoolgirls* in Round 4, we cannot rule out sizeable impacts for some of the outcomes. Finally, Appendix Table S11 shows that pairwise comparisons of CCT and UCT impacts are completely robust to the adjustments we implement, which confirm that (a) most of the statistically significant differences in schooling, marriage, and fertility that existed between these two treatment arms immediately after the program disappeared two years later, and (b) UCT beneficiaries, on average, have a higher age at marriage and a lower level of overall empowerment than CCT beneficiaries by Round 4.

In Appendix Table S12, we present *q-values* controlling for FDR, as described in Anderson (2008). We use Anderson’s Stata code to calculate FDR-adjusted *q-values*, which uses a simple method proposed by Benjamini and Hochberg (1995) to calculate the smallest *q* at which each hypothesis would be rejected. The *q-values* for the 14 primary outcomes in this study, presented alongside the original *p-values* of the impact estimates for each treatment arm, confirm the robustness of our findings to multiple hypothesis testing adjustments: every statistically significant impact for the CCT arm among *baseline dropouts* has a *q-value* below 0.099, while every *q-value* is greater than 0.289 among *baseline schoolgirls*.

Our analysis so far points to two main findings: first, among the more vulnerable group of baseline dropouts, CCTs improved school attainment and decreased marriage and fertility rates, which were sustained over time. Second, the large effects of UCTs among *baseline schoolgirls*

---

24 The Stata code and the paper that describes the method can be found here: https://are.berkeley.edu/~mlanderson/ARE_Website/Research.html.
during the program have all but disappeared within two years. In this sub-section, we find that these two main findings are robust to attrition and multiple hypothesis testing.²⁵

4.6 Husband Characteristics

The program impacts on empowerment presented above, particularly the negative effects in the UCT group,²⁶ motivate the examination of marriage market outcomes. As described earlier, two years after the end of the transfer program, CCT beneficiaries among baseline dropouts were less likely to be ever married or pregnant, had a smaller number of children, and were older at first marriage and pregnancy. While these gains did not translate into increased empowerment or subjective wellbeing in this group, the program might have nonetheless caused study participants to select spouses with different characteristics.

Table V presents the treatment-control comparison of husband characteristics. For baseline dropouts, the evidence is consistent with assortative matching (Panel A): husbands of CCT beneficiaries have completed 0.56 years more of schooling (p-value=0.11) and are 7.4 pp more likely to have successfully completed secondary school (p-value=0.05). By inducing large numbers of dropouts to return to school, CCTs might have driven them to marry more educated husbands than they would have otherwise. This finding does not appear to be driven by differential selection into marriage.²⁷ These spouses, however, are not different in terms of labor market outcomes, cognitive ability, marital fidelity, mental health, HIV (Table V), or attitudes towards women’s empowerment (Appendix Table S13).

In contrast, the delays in marriage and pregnancy among baseline schoolgirls in the UCT group were transitory, leading to an increase in age at first marriage with no gains in education or reductions in actual or desired fertility. The divergence in empowerment between CCT and UCT recipients among baseline schoolgirls, presented above, is also apparent in the characteristics of their husbands. The coefficient estimates for the overall husband quality index are -0.186 and 0.141 for the UCT and CCT groups, respectively (Table V, Panel B). In particular, the husbands of UCT beneficiaries are 8.8 pp less likely to hold secondary school certificates (MSCE) than the control group (p-value=0.11) and scored approximately 0.36 SD lower in the Raven’s colored progressive

²⁵ Reinforcing the idea that our findings are robust to attrition in Round 4, findings of baseline balance and impact estimates from earlier publications, such as Baird, McIntosh and Özler (2011), replicate in the Round 4 sample.
²⁶ Note that the negative empowerment result among married women remains robust to adjustments for IPW, Lee bounds, and multiple-hypothesis testing.
²⁷ A joint F-test of interactions between treatment (CCT) and baseline attributes predicting selection into the husband sample among baseline dropouts is insignificant.
matrices test (p-value=0.03). The differences between the CCT and UCT groups for the overall husband quality index, as well as MSCE and cognitive ability, are all statistically significant.28

The divergence in these marriage market outcomes between CCT and UCT recipients can be explained by program impacts on education and the timing of childbearing and marriage. Environments in which adolescent marriage is common may feature a preference for young brides (Foster and Khan 2000), meaning that delaying marriage may worsen marriage prospects, resulting in either lower husband quality (or bride price) or higher dowry payments (Field and Ambrus 2008). However, potentially counteracting this effect of increased age at marriage is human capital accumulation: for example, Ashraf et al. (2016) show that higher female education is associated with a higher bride price in Indonesia and Zambia. While bride price is uncommon in Zomba, Malawi (the setting for our study), it is likely that higher education is rewarded in the marriage market in other ways, such as husband quality. These factors lead to a tradeoff between increased age at marriage and higher education, which jointly determine husband quality in the absence of bride prices as a market clearing mechanism (Anderson and Bidner 2015).29

Among baseline dropouts, CCT recipients faced exactly this tradeoff and the evidence suggests that, by and large, they improved their marriage outcomes as a result of staying in school and delaying marriage. However, there was no such tradeoff for UCT beneficiaries: the temporary delays in marriage and pregnancy in this group were due to income effects and not accompanied by gains in educational attainment. An examination of Figure II, which shows the relative timings of births and marriages in Panels A and B, respectively, suggests that a large share of these unions may have been shotgun marriages – forced by pregnancies: the large “baby boom” apparent in the UCT group 10-12 months after the end of the cash transfer program, indicating a spike in pregnancies immediately after the cessation of financial support, is preceded by a similarly-sized “marriage boom” only a few months earlier. Thus, consistent with the broader literature, it appears that the UCT beneficiaries ended up with worse marriage market outcomes and lower levels of empowerment as a result of delaying childbearing and marriage without accumulating additional schooling.

28 In contrast to the baseline dropouts, selection regressions indicate that UCTs induced positive selection into marriage (e.g. women who were more educated at baseline and more urban, i.e. those with a higher expected quality of husbands). Correcting for this selection through IPW (not shown here) makes the negative relationship between UCT and husband quality stronger, suggesting that the negative effects estimates presented here are conservative.
29 Field and Ambrus (2008) report that parents in Bangladesh increase dowry payments for daughters who are late bloomers so that they do not end up worse off in terms of spousal quality.
4.7 Child Outcomes

We conclude this section with a discussion of program impacts on children born to study participants. Policies for child development often target the first 1,000 days – from conception to the second birthday (Barham, Macours and Maluccio 2013), a period during which improvements in family income may be particularly important for children’s development. In our experiment, more than 2,000 babies were born to study participants by Round 4 – with endogenous variation in their duration of exposure to the cash transfer program. We have already demonstrated that well-known channels for growth, such as maternal nutrition and stress (Black, Devereux and Salvanes 2017), improved during the two-year program.

In terms of the timing and structure of the cash transfers, we would expect substantial heterogeneity of program impacts on child outcomes both by when the birth took place and whether the transfers to the mother were conditional on school attendance. As in other countries in the region, fertility and schooling are mutually exclusive in Malawi (Baird, McIntosh and Özler 2011; Ozier 2015), meaning that the condition to regularly attend school effectively screens out most expecting and new mothers in the CCT arm: only in the UCT arm would mothers with newborn children continue receiving transfers. Secondly, even in the UCT arm, a child conceived after the end of the program would have had no direct exposure to the program and, as we have shown earlier, the average mother would have acquired no additional education that could provide subsequent human capital-driven benefits. On the other hand, increased mother’s education can, for example, increase child height (Thomas, Strauss and Henriques 1991), so we might expect to see benefits among children born after the program in the CCT groups – particularly among baseline dropouts, who experienced large gains in school attainment themselves. These causal chains suggest that UCT benefits should be concentrated among children born or in utero during

---

30Agüero, Carter and Woolard (2006) study the effect of Child Support Grants in South Africa for children who were exposed to the program up to three years after birth and find sizeable effects of increased exposure to these unconditional cash transfers on child height. Milligan and Stabile (2009), studying child benefits in Canada, find effects on cognitive and socio-emotional skills of children aged 4-6. Dahl and Lochner (2012) using the variation in Earned Income Tax Credit in the U.S., find that increased income improves children’s test scores. Currie and Almond (2011) review the effects of “near cash” programs, such as food stamps, in the U.S. and find credible evidence of effects on birth weight. Finally, Aizer et al. (2016) and Hoynes, Schanzenbach and Almond (2016) find that children whose parents received cash transfers and food stamps in the U.S. had improved education, health, and income as adults.
the program, while CCTs might be most beneficial to children born after the mother’s additional human capital accumulation took place.\footnote{Increased age at first birth can also have positive effects on child height through improved gynecological maturity and decreased competition for nutrition between the mother and the child in utero, which could operate in both treatment groups that delayed pregnancies.}

As with the husband characteristics, we begin by presenting simple treatment-control comparisons for primary child outcomes. These comparisons, presented in Table VI, appear to show few significant differences; none among CCT children among \textit{baseline dropouts} and only one (out of eight outcomes) among \textit{baseline schoolgirls}. In the UCT group, we observe a significantly higher prevalence of exclusive breastfeeding and better parenting practices, with no significant differences between the UCT and CCT treatment arms.

However, we need to be cautious in interpreting these differences between the treatment and control groups, because we know that the program caused significant changes in fertility patterns (Table II): in other words, the raw treatment-control differences are not interpretable as causal impacts of the program on a specific child, because childbearing is endogenous to treatment. We now pursue two approaches: estimation of heterogeneity by child age, and regression control for selection-driven covariates, to address this endogeneity problem. The technical details of the assumptions required and the sequence of adjustments are outlined in Appendix A.

First, we can examine how treatment effects vary across three ‘epochs’ defined by child age. The first epoch captures those \textit{directly exposed} to the program, meaning those born during the program.\footnote{The percentage of baseline schoolgirls who reported having been ever pregnant was less than 2\% at baseline. Hence, children directly exposed to the program in this stratum are almost exclusively born during the intervention. However, approximately 45\% of baseline dropouts had already started childbearing at baseline. Therefore, our analysis includes children under two at the start of the program, who were at least partially exposed to cash transfers.} The second epoch covers those born within nine months of the end of the program, who were \textit{exposed in utero} for a maximum of nine months. Finally, the third epoch covers those born more than nine months after the end of the program, who were \textit{not exposed} to cash transfers either as children or in utero and could only benefit from the program due to improved outcomes of their mothers. We concentrate our analysis on height-for-age z-scores (HAZ), which is an objectively measured indicator of stunting that affects almost 50\% of children under the age of five in Malawi, and is a strong predictor of productivity as an adult in low income settings (LaFave and Thomas 2016).\footnote{Of the two anthropometric measures that we collected for children aged 0-59 months – height and weight – stunting (height-for-age z-score\(<-2\)) is the key indicator of malnutrition in Malawi: almost half of the children under the age of 5 were categorized as stunted in 2010, while wasting (weight-for-height z-score\(<-2\)) rates are low at 4\% (IFPRI 2010).}
Figure III plots the “raw” differences in HAZ for children under 60 months between the treatment and the control groups. The figures are consistent with the hypothesis that differences in children’s heights are moderated by exposure to the program. Most strikingly, we see a very large difference in HAZ between the UCT and the control group during the program, which steadily declines, disappears by the end of the program, and even turns negative during the final epoch (Panel C). This pattern is consistent with the substantive but transient improvements in the nutritional status and mental health of UCT beneficiaries. In contrast, no significant differences in child height are apparent between the CCT and the control groups during the program – also consistent with the fact that most mothers of children born in this period would have dropped out of school as a result of their pregnancies, thus forgoing any cash transfers (Panels A and B). Column 1 in Tables VII and VIII reports the raw differences in HAZ by epoch, for baseline dropouts and baseline schoolgirls respectively, and confirms these patterns.

Next, we attempt to control for endogenous selection via propensity weighting and regression control. The treatment/control comparisons may combine extensive margin selection effects (such as the types of women who became pregnant, the types of partners they chose, and the age at birth) with a ‘direct’ casual effect of the program on the children actually observed. Unlike many such applications in the natural experimental literature, it is entirely plausible that all of the observed impacts on HAZ arise from the selection effect of unwanted children being delayed by the receipt of the UCT. Following the methodology laid out in Appendix A, we can then sequentially implement a set of selection controls: in Column 2 we use a set of baseline maternal characteristics to predict fertility in each epoch, and include inverse propensity weights based on fertility probabilities in the analysis (as well as including these covariates in the regression) to

2014). Child assessments (MDAT and SDQ) are also objectively measured outcomes of cognitive and socio-emotional development, but the target age group for these assessments (36-59 months) makes them unsuitable for analysis by epoch of exposure to the program because only children born during the first year of the program (less than 200 in the baseline schoolgirl stratum with less than 30 in the UCT arm) were eligible for assessment.

We construct these figures by running a locally weighted treatment effects regression across the distribution of child age (Fan 1992) and plotting the resulting time-specific treatment effects and 95% confidence intervals.

In the study of a negative shock, the most likely extensive margin impact is an increase in mortality among the weakest fetuses and children, thereby pushing upwards the average outcome among surviving cohorts exposed to the shock. The large set of papers studying negative shocks such as pollution (Chay and Greenstone 2003; Adhvaryu et al. 2016; Black et al. 2017), disease (Almond 2006), and hunger (Almond and Mazumder 2011) can therefore typically argue that any negative effects found on surviving children are actually conservative. Because we study a positive shock that may have delayed economically motivated pregnancies that were expected to have worse outcomes, the selection and direct treatment effects in our case both point to superior child outcomes in the treatment. Decomposing these effects is therefore critical.
provide estimates of impact that are doubly robust to maternal type selection. Column 3 includes covariates controlling for paternal type, Column 4 adds flexible controls for child age, while Column 5 adds indicator variables for the mother’s age at birth and interactions of maternal age with all other baseline covariates. Subject to the assumptions laid out in the Appendix A, these estimates allow us to move from the reduced-form ‘raw’ treatment effects to estimates of a ‘direct’ effect – i.e. suggestive *ceteris paribus* impacts of CCTs and UCTs on the children actually born by epoch.

Column 2 in Table VIII, Panel A shows that the maternal selection controls alone reduce the effect of UCTs during the program by almost a half (from .953 to .525 SD), confirming significant positive selection into childbearing during the program in the UCT arm. The other pathways have a limited effect, resulting in a fully adjusted direct effect of .523 SD (column 5). The size of this remaining direct effect is consistent with Barham, Macours and Maluccio (2013), who report that children in Nicaragua who received three years of cash transfers were 0.2-0.4 SD taller; and with Agüero, Carter and Woolard (2006), who find that children in South Africa receiving child support grants for most of the period between 0-3 years of age gained as much as 0.45 SD in HAZ. The bold curves in Figure III plot these ‘direct’, fully adjusted Fan regressions across the month of birth, including the battery of controls included in Column 5 of Tables VII & VIII. The distribution of direct treatment effects in the UCT arm shown in Panel C is remarkably consistent with what we would expect: a significant and positive effect on HAZ among children born during the program, which disappears immediately following the cessation of transfers.

The effects on HAZ in the CCT groups are also as expected: as females who dropped out of school due to pregnancies did not continue to receive transfers, we’d expect little effect on their children born during the program. Conversely, if increased education or delaying childbearing has an effect on child height, we might see effects among children of CCT recipients after the program. Among *baseline dropouts* or *baseline schoolgirls*, we see no significant effects on HAZ for babies born during the program. However, the corrected plots show modest (0.10-0.25 SD) improvements in HAZ for children born after the program to *baseline schoolgirls* who received CCTs (Figure III, Panel B).

The findings here are consistent with the theory that underlies the tradeoff between CCTs for schooling and UCTs: UCTs primarily confer an income effect on children born during the program and no effects on children born later because they do not lead to an accumulation of
capital (human, physical, or social) for the mother. On the other hand, CCTs deny such benefits to the children of non-compliers during the program, but may have modest effects on future children through increased human capital accumulation.

5. Conclusion

The most striking feature of the findings presented in this paper is the transience of the impacts of cash transfers, especially those given unconditionally. Particularly glaring are the fleeting decreases in child marriage and teen pregnancy in the UCT arm, along with psychological distress and HIV – the prevalence of all of which reverted to control group levels within just two years, implying significant but temporary income effects. Within months of the end of the program, a large number of UCT beneficiaries became pregnant, and were married soon thereafter. These delayed marriages, without any concomitant improvements in education, were, on average, to lower quality husbands and may have resulted in decreased empowerment in this group. This negative impact of waiting to marry in the absence of compensating gains is consistent with evidence from South Asia (Field and Ambrus 2008).

On the other hand, there were sustained program effects on school attainment (accompanied by assortative matching), early marriage, and pregnancy for baseline dropouts receiving CCTs. However, these effects did not translate into reductions in HIV or gains in labor market outcomes or empowerment. Several reasons might explain the disconnect between increased school attainment and no improvements in labor market outcomes, empowerment, or health. First, it is possible that increased schooling does not provide one with the skills needed to increase future welfare in this context. There are very few formal sector jobs for women in Malawi and most households depend on subsistence farming and a variety of informal sector activities. We administered tests of skills needed in farming and running small household enterprises and detected no effects in these domains. If safe and well-paying jobs existed for women in Malawi, households might invest in the necessary human capital of adolescent females on their own – perhaps even without the help of any outside interventions (Munshi and Rosenzweig 2006; Jensen

36 We do not see any positive effects of UCTs for babies born within nine month of the end of the program, i.e. those exposed in utero. While this may be considered surprising given the extant evidence on the importance of this period for physical development, it should be remembered that the young mothers are also dealing with the cessation of support during this same period. Changes in lifestyle and increased stress from the loss of regular income during this transitional period may have dampened any beneficial effects of cash transfers on the child in utero.

37 These findings are consistent with Duflo, Dupas and Kremer (2015), who find that education subsidies in Kenya reduce dropout, pregnancy, and marriage, but not sexually transmitted infections. They suggest a model in which choices between committed and casual relationships, rather than unprotected sex alone, affect pregnancy and HIV.
Second, task performance is dependent on not only improvements in cognitive skills, but also on character skills and effort (Heckman and Kautz 2013). Hence, it is possible that CCTs, by providing incentives for formal schooling, improved only cognitive skills, which may not have been sufficient to increase productivity.\(^{38}\)

Our study provides some important guideposts for the design of effective adolescent-focused cash transfer programs. First, the palliative benefits of small and frequent unconditional cash transfers are uncontested and reinforced by our study, but the idea that they can contribute to a sustained improvement in welfare over the longer-run is unproven and not supported here.\(^{39}\) Second, we shed further light on the tradeoffs between the benefits of conditional and unconditional transfers. The lack of knock-on effects from schooling gains in this context implies that the imperative to use conditions to generate increased investments in human capital may be weak when few income-generating opportunities exist. Moreover, by denying adolescent girls and young women cash transfers at precisely the moment when they are most likely to start childbearing, a myriad of potential benefits are missed under CCT programs. A potentially promising way of resolving this tradeoff is to view CCT and UCT programs as complements to each other rather than alternatives: policymakers could provide a basic unconditional cash transfer to adolescent girls topped up by conditional cash transfers for human capital accumulation and desired health behaviors – providing both an incentive to invest in education and health while still guaranteeing a basic level of protection to those who are unable or unwilling to comply with the conditions. Third, and finally, the promising (if only suggestive) evidence of the positive effect of UCTs on children’s height provides an additional reason to consider providing basic UCTs to adolescent females. Indeed, Currie and Almond (2011) have suggested that targeting transfers towards women of childbearing age may be beneficial in the U.S. context, so as to maximize benefits to children in utero. This form of targeting would suffer from remarkably little ‘leakage’

---

\(^{38}\) Heckman and Mosso (2014) state “The most effective adolescent interventions target formation of personality, socioemotional, and character skills through mentoring and guidance, including providing information.” Bandiera et al. (2017) provide suggestive evidence that a mentoring program in Uganda (ELA) that provided young females with “hard” vocational and “soft” life skills may have led to longer-term improvements in welfare.

\(^{39}\) We do not mean to downplay or underestimate the effects of redistributive policies on current poverty and inequality reduction, even if they do not lead to substantive increases in human capital accumulation. Welfare gains from such effects can be as large as, if not larger than, those from human capital investments (Alderman, Behrman and Tasneem 2015).
in the Malawian context; two thirds of women aged 20-24 gave birth by age 20 and virtually all females have started childbearing by age 25 (National Statistical Office and ICF Macro 2005).

Given the medium-term nature of these results, it is natural to ask how much we can infer about longer-run impacts. As our study captures outcomes a little more than two years after the cash transfers stopped, we cannot speak to long-term effects, such as those analyzed in the U.S. context in recent studies (Aizer et al. 2016; Hoynes, Schanzenbach and Almond 2016). To guide our thinking, we return again to the role of productive assets in generating long-term rewards: to make an impact later in life, a program must have meaningfully shifted the stock of some form of capital that can generate returns over the long haul. For baseline dropouts, who were offered CCTs to return to school, the improvement in schooling human capital is sizeable, and they have formed households with more educated partners. For this group, it may be premature to conclude that improvements in education have led to no long-term gains. If the education/wage relationship becomes steeper with age, or if household-level human capital alters the economic trajectory of these households, future follow-up studies may well reveal longer-term benefits. For baseline schoolgirls in the UCT arm, our findings suggest that two years of financial support during adolescence might have been too short—rather than a two-year follow-up window being too short to trace out subsequent impacts.40 Only two years after the end of the program, UCT beneficiaries are, in most respects, in a position indistinguishable from where they would have been in the absence of cash transfers. The unwinding of the program impacts on marriage and pregnancy is immediate and substantial, so given the lack of school attainment or learning effects in this group it is only their children in whom we note some vehicle for durable improvements in human capital.

---

40 However, it should be noted that the Mothers’ Pension program of the early 20th century U.S. had a median duration of three years and was of similar generosity to many cash transfer programs today, including ours (Aizer et al. 2016), and showed long-term effects in health, education, and income among children of program beneficiaries.
REFERENCES


Alderman, Harold, Jere Behrman, and Afia Tasneem. 2015. The contribution of increased equity to the estimated social benefits from a transfer program: An illustration from PROGRESA. IFPRI Discussion Paper 1475.


Baird, Sarah, Francisco HG Ferreira, Berk Özler, and Michael Woolcock. 2013. Relative effectiveness of conditional and unconditional cash transfers for schooling outcomes in
developing countries: a systematic review. *Campbell Systematic Reviews* 9 (8).


Table I: Program impacts on education and learning (beneficiaries)

<table>
<thead>
<tr>
<th></th>
<th>Highest Grade Completed</th>
<th>English Test Score (Standardized)</th>
<th>TIMMS Math Score (Standardized)</th>
<th>Non-TIMMS Math Score (Standardized)</th>
<th>Cognitive Test Score (Standardized)</th>
<th>Competencies Score (Standardized)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>End of Program</td>
<td>Two Years After Program</td>
<td>End of Program</td>
<td>Two Years After Program</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>During Program</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>0.579***</td>
<td>0.558***</td>
<td>0.621***</td>
<td>0.079</td>
<td>0.147***</td>
<td>0.116</td>
<td>0.163**</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.102)</td>
<td>(0.125)</td>
<td>(0.071)</td>
<td>(0.056)</td>
<td>(0.072)</td>
<td>(0.070)</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>6.345</td>
<td>6.967</td>
<td>6.997</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td>697</td>
<td>718</td>
<td>744</td>
<td>704</td>
<td>704</td>
<td>704</td>
<td>742</td>
</tr>
</tbody>
</table>

Panel A: Baseline Dropouts

Panel B: Baseline Schoolgirls

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The cognitive test score is based on Raven’s Colored Progressive Matrices. Math and English reading comprehension tests were developed based on the Malawian school curricula. Five questions (four from the Fourth Grade test and one from the Eighth Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every four years, were added to the math test. Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The skills tested included reading and following instructions to apply fertilizer; making correct change during hypothetical market transactions; sending text messages and using the calculator on a mobile phone, and calculating profits under hypothetical business scenarios. All test scores and the competency index were standardized to have a mean of zero and a standard deviation of one in the control group. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex, and whether the respondent participated in the pilot phase of the development of the testing instruments. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Note that in Rounds 2 and 3, highest grade completed is actually highest grade attended. Parameter estimates statistically different than zero at 99% (***) , 95% (**) , and 90% (*) confidence.
### Table II: Program impacts on marriage and fertility (beneficiaries)

#### Panel A: Baseline Dropouts

<table>
<thead>
<tr>
<th>Age First Marriage</th>
<th>=1 if Ever Married</th>
<th>Age at First Birth</th>
<th>=1 if Ever Pregnant</th>
<th>Desired Fertility</th>
<th>Number of Live Births</th>
</tr>
</thead>
<tbody>
<tr>
<td>During Program</td>
<td>End of Program</td>
<td>Two Years After Program</td>
<td>During Program</td>
<td>End of Program</td>
<td>Two Years After Program</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.140***</td>
<td>-0.157***</td>
<td>-0.107***</td>
<td>0.431***</td>
<td>-0.057*</td>
<td>-0.081***</td>
</tr>
<tr>
<td>(0.029)</td>
<td>(0.037)</td>
<td>(0.032)</td>
<td>(0.155)</td>
<td>(0.030)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>

Mean in Control Group

| 0.291 | 0.575 | 0.809 | 19.644 | 0.610 | 0.784 | 0.924 |

Sample Size

| 698  | 718  | 744  |

#### Panel B: Baseline Schoolgirls

<table>
<thead>
<tr>
<th>Age First Marriage</th>
<th>=1 if Ever Married</th>
<th>Age at First Birth</th>
<th>=1 if Ever Pregnant</th>
<th>Desired Fertility</th>
<th>Number of Live Births</th>
</tr>
</thead>
<tbody>
<tr>
<td>During Program</td>
<td>End of Program</td>
<td>Two Years After Program</td>
<td>During Program</td>
<td>End of Program</td>
<td>Two Years After Program</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.033***</td>
<td>-0.083***</td>
<td>-0.010</td>
<td>0.486**</td>
<td>-0.013</td>
<td>-0.063**</td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.024)</td>
<td>(0.027)</td>
<td>(0.148)</td>
<td>(0.015)</td>
<td>(0.027)</td>
</tr>
</tbody>
</table>

Mean in Control Group

| 0.047 | 0.180 | 0.402 | 18.651 | 0.092 | 0.247 | 0.501 |

Sample Size

| 1,967 | 2,018 | 2,049 | 821    | 1,966 | 2,019 | 2,049 |

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. We correct for inconsistencies in ‘ever married’ and ‘ever pregnant’ across rounds. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***)*, 95% (**), and 90% (*) confidence.
### Table III: Program impacts on HIV and Anemia (beneficiaries)

#### Panel A: Baseline Dropouts

<table>
<thead>
<tr>
<th>=1 if HIV Positive</th>
<th>=1 if Anemic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>During Program</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>0.06</td>
</tr>
<tr>
<td>Sample Size</td>
<td>373</td>
</tr>
</tbody>
</table>

#### Panel B: Baseline Schoolgirls

<table>
<thead>
<tr>
<th>=1 if Conditional Schoolgirl</th>
<th>=1 if Unconditional Schoolgirl</th>
</tr>
</thead>
<tbody>
<tr>
<td>=1 if HIV Positive</td>
<td>=1 if Anemic</td>
</tr>
<tr>
<td></td>
<td>During Program</td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>-0.020**</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
</tr>
<tr>
<td>=1 if Unconditional Schoolgirl</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
</tr>
<tr>
<td>p-value UCT vs. CCT</td>
<td>0.616</td>
</tr>
<tr>
<td>p-value Treatment</td>
<td>0.112</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>0.026</td>
</tr>
<tr>
<td>Sample Size</td>
<td>1,192</td>
</tr>
</tbody>
</table>

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. An individual is considered anemic if her hemoglobin count is less than or equal to 11g/dL if pregnant and less than or equal to 12g/dL if non-pregnant based on WHO guidelines to define mild anemia. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***) , 95% (**) , and 90% (*) confidence.
Table IV: Program impacts on labor market outcomes and empowerment (beneficiaries)

### Panel A: Baseline Dropouts

<table>
<thead>
<tr>
<th>Opportunity Cost of Time</th>
<th>Typical Wage in Past Three Months</th>
<th>Proportion of Hours Spent in Self-Employment or Paid Work in Past Week</th>
<th>Super-Index of Overall Empowerment (Standardized)</th>
<th>Change in Subjective Wellbeing from Five Years Ago to Today</th>
<th>Super-Index of Unmarried Empowerment (Standardized)</th>
<th>Super-Index of Married Empowerment (Standardized)</th>
<th>Married Index of Economic Control (Standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 if Conditional Schoolgirl</td>
<td></td>
<td>-0.037</td>
<td>-0.140**</td>
<td>-0.011</td>
<td>-0.083</td>
<td>-0.032</td>
<td>0.018</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td></td>
<td>0.707</td>
<td>0.375</td>
<td>0.061</td>
<td>0.000</td>
<td>1.120</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td></td>
<td>718</td>
<td>743</td>
<td>744</td>
<td>744</td>
<td>744</td>
<td>289</td>
</tr>
</tbody>
</table>

### Panel B: Baseline Schoolgirls

<table>
<thead>
<tr>
<th>Opportunity Cost of Time</th>
<th>Typical Wage in Past Three Months</th>
<th>Proportion of Hours Spent in Self-Employment or Paid Work in Past Week</th>
<th>Super-Index of Overall Empowerment (Standardized)</th>
<th>Change in Subjective Wellbeing from Five Years Ago to Today</th>
<th>Super-Index of Unmarried Empowerment (Standardized)</th>
<th>Super-Index of Married Empowerment (Standardized)</th>
<th>Married Index of Economic Control (Standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 if Conditional Schoolgirl</td>
<td></td>
<td>-0.051</td>
<td>-0.011</td>
<td>0.003</td>
<td>0.049</td>
<td>0.276</td>
<td>0.111</td>
</tr>
<tr>
<td>1 if Unconditional Schoolgirl</td>
<td></td>
<td>-0.115</td>
<td>0.036</td>
<td>0.002</td>
<td>-0.159*</td>
<td>0.176</td>
<td>-0.094</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td></td>
<td>0.897</td>
<td>0.212</td>
<td>0.029</td>
<td>0.000</td>
<td>0.906</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td></td>
<td>2,002</td>
<td>2,048</td>
<td>2,045</td>
<td>2,049</td>
<td>2,049</td>
<td>1,271</td>
</tr>
</tbody>
</table>

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Opportunity cost of time is calculated by taking the minimum daily wage the respondent would take for one year of work in her village. Detail on the construction of the super-indices can be found in Appendix A and Appendix B. The change in subjective wellbeing asks the respondent where she sees herself on a 10-step ladder comparing five years ago to today, where zero represents the worst possible life she could have and 10 represents the best possible life she could have. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***) and 99% (**) confidence.
Table V: Program impacts marriage market outcomes (husband characteristics)

### Panel A: Baseline Dropouts

<table>
<thead>
<tr>
<th>Husband Quality Super Index (Standardized)</th>
<th>Highest Grade Completed =1 if Passed Primary School (PSLC)</th>
<th>=1 if Passed Junior Secondary School (JCE)</th>
<th>Cognitive Test Score (Standardized)</th>
<th>Typical Wage in Past Three Months =1 if Currently Employed</th>
<th>Sexual Activity and Marital Fidelity (Standardized) =1 if Does Not Suffers from Psychological Distress</th>
<th>=1 if HIV Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean in Control Group</td>
<td>0.000</td>
<td>7.806</td>
<td>0.526</td>
<td>0.314</td>
<td>0.097</td>
<td>1.194</td>
</tr>
<tr>
<td>Sample Size</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>326</td>
<td>325</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>0.141</td>
<td>0.046</td>
<td>0.024</td>
<td>0.012</td>
<td>0.059</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.271)</td>
<td>(0.043)</td>
<td>(0.049)</td>
<td>(0.053)</td>
<td>(0.109)</td>
</tr>
<tr>
<td>=1 if Unconditional Schoolgirl</td>
<td>-0.186</td>
<td>-0.454</td>
<td>0.005</td>
<td>0.017</td>
<td>-0.088</td>
<td>-0.357**</td>
</tr>
<tr>
<td></td>
<td>(0.180)</td>
<td>(0.425)</td>
<td>(0.068)</td>
<td>(0.086)</td>
<td>(0.054)</td>
<td>(0.163)</td>
</tr>
<tr>
<td>p-value UCT vs. CCT</td>
<td>0.084</td>
<td>0.240</td>
<td>0.776</td>
<td>0.954</td>
<td>0.042</td>
<td>0.044</td>
</tr>
<tr>
<td>p-value Treatment</td>
<td>0.145</td>
<td>0.490</td>
<td>0.845</td>
<td>0.964</td>
<td>0.118</td>
<td>0.087</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>0.000</td>
<td>9.743</td>
<td>0.699</td>
<td>0.541</td>
<td>0.258</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td>543</td>
<td>543</td>
<td>543</td>
<td>543</td>
<td>543</td>
<td>539</td>
</tr>
</tbody>
</table>

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The husband quality super index is a standardized index of all other outcomes in this table (except HIV as it is defined on a smaller sample). All variables are constructed so that higher values are better, except for HIV. The cognitive test score is based on Raven’s Colored Progressive Matrices. The husband’s sexual activity and marital fidelity index is constructed from three variables: number of sexual partners ever, number of sexual partners in the past 12 months and an indicator for concurrent multiple partners. Psychological distress is equal to one if the summed General Health Questionnaire-12 score is equal to three or higher, and is zero otherwise. Additional details on the variables can be found in Appendix A and Appendix B. Baseline values of the following variables for the beneficiaries are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to husbands of respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). The husband quality super index regression also includes an indicator for whether any of the sub-components of the indicator are missing. Parameter estimates statistically different than zero at 99% (***)), 95% (**), and 90% (*) confidence.
### Table VI: Program impacts on child outcomes (children of beneficiaries)

#### Panel A: Baseline Dropouts

<table>
<thead>
<tr>
<th></th>
<th>Height-for-Age z-score</th>
<th>Neonatal Mortality</th>
<th>Postneonatal Mortality</th>
<th>Parenting Practices Percentage Score</th>
<th>Exclusively Breastfed for First 6 Months</th>
<th>Malawi Developmental Assessment Tool (3-4 year-olds) (Standardized)</th>
<th>Reported Child Difficulties (3-4 year-olds) (Standardized)</th>
<th>Reported Pro-Social Behaviors (3-4 year-olds) (Standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>=1 if Treatment Dropout</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
</tr>
<tr>
<td></td>
<td>-0.013</td>
<td>0.013</td>
<td>-0.009</td>
<td>-0.003</td>
<td>0.030</td>
<td>-0.086</td>
<td>0.104</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.026)</td>
<td>(0.112)</td>
<td>(0.190)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>-1.351</td>
<td>0.015</td>
<td>0.026</td>
<td>0.496</td>
<td>0.804</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td>742</td>
<td>958</td>
<td>707</td>
<td>861</td>
<td>971</td>
<td>213</td>
<td>223</td>
<td>223</td>
</tr>
</tbody>
</table>

#### Panel A: Baseline Schoolgirls

<table>
<thead>
<tr>
<th></th>
<th>Height-for-Age z-score</th>
<th>Neonatal Mortality</th>
<th>Postneonatal Mortality</th>
<th>Parenting Practices Percentage Score</th>
<th>Exclusively Breastfed for First 6 Months</th>
<th>Malawi Developmental Assessment Tool (3-4 year-olds) (Standardized)</th>
<th>Reported Child Difficulties (3-4 year-olds) (Standardized)</th>
<th>Reported Pro-Social Behaviors (3-4 year-olds) (Standardized)</th>
</tr>
</thead>
<tbody>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.096</td>
<td>-0.014</td>
<td>0.005</td>
<td>0.012</td>
<td>0.029</td>
<td>-0.294***</td>
<td>-0.011</td>
<td>-0.357</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
<td>(0.009)</td>
<td>(0.012)</td>
<td>(0.018)</td>
<td>(0.033)</td>
<td>(0.167)</td>
<td>(0.180)</td>
<td>(0.282)</td>
</tr>
<tr>
<td>=1 if Unconditional Schoolgirl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.065</td>
<td>-0.012</td>
<td>0.001</td>
<td>0.050**</td>
<td>0.126****</td>
<td>0.213</td>
<td>0.035</td>
<td>-0.132</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.029)</td>
<td>(0.039)</td>
<td>(0.376)</td>
<td>(0.173)</td>
<td>(0.309)</td>
</tr>
<tr>
<td>p-value UCT vs. CCT</td>
<td>0.872</td>
<td>0.901</td>
<td>0.734</td>
<td>0.229</td>
<td>0.014</td>
<td>0.172</td>
<td>0.835</td>
<td>0.568</td>
</tr>
<tr>
<td>p-value Treatment</td>
<td>0.666</td>
<td>0.302</td>
<td>0.912</td>
<td>0.215</td>
<td>0.006</td>
<td>0.145</td>
<td>0.974</td>
<td>0.434</td>
</tr>
<tr>
<td>Mean in Control Group</td>
<td>-1.410</td>
<td>0.028</td>
<td>0.013</td>
<td>0.484</td>
<td>0.771</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Sample Size</td>
<td>1,032</td>
<td>1,167</td>
<td>756</td>
<td>1,090</td>
<td>1,169</td>
<td>185</td>
<td>196</td>
<td>196</td>
</tr>
</tbody>
</table>

**Notes:** Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. The parenting practices score is the percentage score on a set of parenting practices. The Malawi Developmental Assessment Tool is a test of fine motor skills, language, and hearing administered directly to the child. The reported child difficulties and reported pro-social behaviors are created using the Strengths and Difficulties Questionnaire (http://www.sdqinfo.com/c3.html). Additional details on the outcome variables can be found in Appendix A and Appendix B. Baseline values of the following variables are included as controls in the regression analyses: gender of the child, age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***) , 95% (**), and 90% (*) confidence.
### Table VII: Program impacts on height-for-age z-scores (children of beneficiaries: baseline dropouts)

<table>
<thead>
<tr>
<th>Panel A: Born During Program</th>
<th>Raw Effect</th>
<th>+ Maternal Selection weights</th>
<th>+ Paternal Selection Controls</th>
<th>+ Child Age</th>
<th>+ Mother Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>-0.015</td>
<td>-0.174</td>
<td>-0.139</td>
<td>-0.154</td>
<td>-0.051</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(0.149)</td>
<td>(0.143)</td>
<td>(0.140)</td>
<td>(0.136)</td>
</tr>
</tbody>
</table>

| Sample Size | 367 | 367 | 367 | 367 | 367 |

| Panel B: Born Within 9 Months of Program Ended |
| =1 if Conditional Schoolgirl |
| 0.353 | 0.518* | 0.394 | 0.411* | 0.577** |
| (0.296) | (0.303) | (0.249) | (0.234) | (0.260) |

| Sample Size | 88 | 88 | 88 | 88 | 88 |

| Panel C: Born More than 9 Months After Program Ended |
| =1 if Conditional Schoolgirl |
| -0.269 | -0.175 | -0.127 | -0.137 | -0.183 |
| (0.168) | (0.192) | (0.161) | (0.154) | (0.152) |

| Sample Size | 287 | 287 | 287 | 287 | 287 |

**Control Structure:**
- Maternal selection controls + propensity weight: X X X X X
- Father selection controls: X X X X
- Cubic in child age in months: X X
- Maternal age in years, age interactions: X

**Notes:** Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights and controls directly for maternal baseline characteristics (stratum indicators, household asset index, highest grade attended, and an indicator for never had sex). Specification (3) adds controls for paternal attributes (highest education level, religion, ethnicity, main activity, and likely HIV status). Specification (4) adds a linear, quadratic, and cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***)**, 95% (**), and 90% (*) confidence.
Table VIII: Program impacts on height-for-age z-scores (children of beneficiaries: baseline schoolgirls)

<table>
<thead>
<tr>
<th>Panel A: Born During Program</th>
<th>Raw Effect</th>
<th>Direct Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender</td>
<td>+ Maternal Selection weights</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>=1 if Conditional Schoolgirl</td>
<td>0.155</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td>(0.192)</td>
</tr>
<tr>
<td>=1 if Unconditional Schoolgirl</td>
<td>0.953**</td>
<td>0.525**</td>
</tr>
<tr>
<td></td>
<td>(0.476)</td>
<td>(0.221)</td>
</tr>
<tr>
<td>p-value UCT vs. CCT</td>
<td>0.091</td>
<td>0.022</td>
</tr>
<tr>
<td>p-value Treatment</td>
<td>0.123</td>
<td>0.040</td>
</tr>
<tr>
<td>Sample Size</td>
<td>315</td>
<td>315</td>
</tr>
</tbody>
</table>

Panel B: Born Within 9 Months of Program Ended

| =1 if Conditional Schoolgirl | 0.251      | 0.156         | 0.235                    | 0.125      | 0.086        |
|                              | (0.279)    | (0.263)       | (0.240)                  | (0.175)    | (0.194)      |
| =1 if Unconditional Schoolgirl | 0.177      | 0.163         | 0.109                    | -0.431**   | -0.434**     |
|                              | (0.514)    | (0.315)       | (0.336)                  | (0.183)    | (0.193)      |
| p-value UCT vs. CCT          | 0.887      | 0.984         | 0.725                    | 0.013      | 0.028        |
| p-value Treatment            | 0.663      | 0.787         | 0.619                    | 0.028      | 0.047        |
| Sample Size                  | 214        | 211           | 211                      | 211        | 211          |

Panel C: Born More than 9 Months After Program Ended

| =1 if Conditional Schoolgirl | -0.011     | 0.497         | 0.149                    | 0.264      | 0.257        |
|                              | (0.187)    | (0.445)       | (0.199)                  | (0.196)    | (0.179)      |
| =1 if Unconditional Schoolgirl | -0.351**   | -0.651***     | -0.336                   | -0.102     | -0.123       |
|                              | (0.174)    | (0.242)       | (0.212)                  | (0.168)    | (0.183)      |
| p-value UCT vs. CCT          | 0.115      | 0.006         | 0.025                    | 0.068      | 0.078        |
| p-value Treatment            | 0.114      | 0.002         | 0.075                    | 0.184      | 0.186        |
| Sample Size                  | 507        | 506           | 506                      | 506        | 506          |

Control Structure:
- Maternal selection controls + propensity weight
- Father selection controls
- Cubic in child age in months
- Maternal age in years, age interactions

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights and controls directly for maternal baseline characteristics (stratum indicators, household asset index, highest grade attended, and an indicator for never had sex). Specification (3) adds controls for paternal attributes (highest education level, religion, ethnicity, main activity, and likely HIV status). Specification (4) adds a linear, quadratic, and cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (**), 95% (**), and 90% (*) confidence.
**Figure I: Research Design**

<table>
<thead>
<tr>
<th>Treatment EAs</th>
<th>Control EAs</th>
</tr>
</thead>
<tbody>
<tr>
<td>(88 Clusters)</td>
<td>(N=88)</td>
</tr>
<tr>
<td><strong>Conditional</strong></td>
<td><strong>Unconditional</strong></td>
</tr>
<tr>
<td>(N=61*)</td>
<td>(N=27)</td>
</tr>
<tr>
<td>Baseline Dropouts</td>
<td></td>
</tr>
<tr>
<td>(N=889)</td>
<td></td>
</tr>
<tr>
<td>Baseline Schoolgirls</td>
<td></td>
</tr>
<tr>
<td>(N=2,907)</td>
<td></td>
</tr>
</tbody>
</table>

- CCT
- Within-village control
- UCT
- Pure control

*In 15 of the 61 conditional treatment clusters only baseline dropouts were treated.*
**Figure II:** Monthly marriage and fertility rates for baseline schoolgirls

**Panel A: Monthly Fertility Rates**

![Graph showing monthly fertility rates with three lines representing different groups: Control, CCT, UCT. The x-axis represents years from 2008 to 2012, and the y-axis shows the smoothed fraction of core respondents who give birth. The graph includes time periods labeled as During Program, Utero, and After Program.]

**Panel B: Monthly Marriage Rates**

![Graph showing monthly marriage rates with three lines representing different groups: Control, CCT, UCT. The x-axis represents years from 2008 to 2012, and the y-axis shows the smoothed fraction of core respondents who get married. The graph includes time periods labeled as During Program, W/in 9 mos, and After Program.]

**Notes:** Figures illustrate the smoothed fraction of core respondents who give birth (Panel A) or get married (Panel B) in each month using retrospective information on the month of birth and marriage, respectively.
**Figure III:** Fan regressions of height-for-age z-scores by month of birth, raw and fully adjusted treatment effects with 95% confidence intervals

Panel A: Baseline Dropouts, CCT

Raw and Direct CCT Dropout effect on Child HAZ

With 95% Confidence Intervals

Panel B: Baseline Schoolgirls, CCT

Raw and Direct CCT Schoolgirl effect on Child HAZ

With 95% Confidence Intervals

Panel C: Baseline Schoolgirls, UCT

Raw and Direct UCT Schoolgirl effect on Child HAZ

With 95% Confidence Intervals

Adjustments include propensity weighting, child and maternal age, maternal and paternal attributes.