When the Money Runs out:
Do Cash Transfers Have Sustained Effects on Human Capital Accumulation?

Sarah Baird, Craig McIntosh, and Berk Özler

Working Paper Series
Center for Effective Global Action
University of California

eScholarship
University of California

When the Money Runs out:

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

Sarah Baird, Craig McIntosh, and Berk Özler

November 30, 2016

Abstract

We examine the medium-term effects of a two-year cash transfer program targeted to adolescent girls and young women. Significant declines in HIV prevalence, teen pregnancy, and early marriage among recipients of unconditional cash transfers (UCTs) during the program evaporated quickly two years after the cessation of transfers. However, children born to UCT beneficiaries during the program had significantly higher height-for-age z-scores at follow-up. On the other hand, conditional cash transfers (CCTs) offered to out-of-school females at baseline produced a large increase in educational attainment and a sustained reduction in the total number of births, but caused no gains in health, labor market outcomes, or empowerment. Our findings point to both the promise and the limitations of cash transfer programs for sustained gains in welfare among young women.

Keywords: Cash Transfers, Long-term Impacts, Human capital
JEL Codes: C93, I15, I21, I38, J12, J13

---

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 (Baird et al. 2013; Fiszbein, Schady and Ferreira 2009 ; Haushofer and Shapiro 2016; Saavedra and Garcia 2013). However, with some recent exceptions discussed below, 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

---

[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., Gertler, Martinez and Rubio-Codina 2012 and Behrman, Parker and Todd 2011). 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're less relevant for our examination of longer-term impacts on adolescent beneficiaries.

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 another treatment group that was offered equal-sized *unconditional* cash transfers (UCT).[3] Second, we not only collected data on a rich set of outcomes (education, childbearing and marriage, health, labor market outcomes, empowerment, and subjective wellbeing) for the target population of young females, but also on their own children and husbands as they started bearing children and getting married. The resulting analysis is a comprehensive assessment of the relative effects of CCT and UCT programs targeted to adolescents for two years during an important period of transition into adulthood.

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 during this period not only have immediate consequences to their own lives, but also longer-term

---

[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 by Barrera-Osorio, Linden and Saavedra (2015), which examines education outcomes eight years after treatment.

benefits to their offspring and communities at large (Lloyd and Young 2009; Duflo 2012).[4] 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).

For any intervention during adolescence 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. Even when young women attain higher schooling and delay childbearing and marriage, low quality education, credit 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. For today's adolescent girls to turn into productive and happier young women with healthier families, programs that improve endowments or aspirations must work in a social and institutional context whereby these newfound forms of capital can generate sustained returns (*Gender Equality and Development* 2011).

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 (Araujo, Bosch and Schady 2016; Baez and Camacho 2011; Barham, Macours and Maluccio 2013; Behrman, Parker and Todd 2011; Filmer and

---

[4] We use capabilities and functionings as described in Heckman and Corbin (2016), page 10: "At a point in time, agents have endowments, including cognitive skills, personality and character skills, and health, as well as access to information, financial resources, and peers. They combine to produce the space of potential actions ("capabilities"). Which actions (functionings) are selected depend on preferences (personal and social), norms, and the efforts of individuals which are shaped in part by both preferences and sociocultural norms."

Schady 2014).[5] There are also programs that target adolescent females directly by providing them a safe space to meet on a regular basis and develop life skills. For example, the Empowerment and Livelihood for Adolescents (ELA) program in Uganda, which provided vocational and life skills training to 14-20 year old girls in development clubs held outside of school, showed significant declines in childbearing, marriage, and having had sex unwillingly within two years, as well as increases in self-employment activities and expenditures on private consumption goods (Bandiera et al. 2015). However, a similar intervention in Tanzania found no effect (Buehren et al. 2015). Early findings from a program that combined training and mentoring with financial incentives to delay marriage until the age of 18 in Bangladesh (Kishoree Konta) indicate that only in the incentive arms did girls have lower marriage rates and improved educational outcomes in the short-run (Glennerster 2013).

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. Aizer et al. 2016; Currie and Almond 2011; Hoynes, Schanzenbach and Almond 2016) and has demonstrated beneficial effects on a host of outcomes as adults. Chetty, Hendren and Katz (2016) show that exposure to lower poverty neighborhoods has long-term benefits but only for children who were young at the time of random assignment and had higher exposure as a result.

In summary, the extant evidence is mixed as to what we can expect from programs that

---

[5] 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.

target adolescent females in developing countries with respect to longer-term benefits: the evidence from cash transfer programs do not look particularly promising for income and employment gains, while most other programs do not yet have long-term evaluations. Furthermore, we know little about impacts of such programs on other important outcomes, such as health, economic empowerment, marriage market outcomes, or early childhood development.[6]

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.[7] 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). Rigorously following a pre-analysis plan, we first look at 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 and at the end of the program failed to translate into increased welfare in the longer-run. 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

---

[6] 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).

[7] At baseline, our target population was never-married females, aged 13-22.

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 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 pre-specified stratum of adolescent females who had already dropped out of school at baseline: CCTs were highly effective in allowing a very large share of this group to return to school. 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 those who were in school at baseline, CCTs did not have any lasting effects, positive or negative, mainly 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: our 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 recent evaluations of CCT or bursary programs discussed above. Second, we

7

add to the literature on the effects of human capital accumulation and increased age at marriage on marriage market outcomes (Anderson and Bidner 2015; Ashraf et al. 2015; Field and Ambrus 2008). Theory suggests that these two factors affect spousal quality in opposite directions with, *ceteris paribus*, increased education improving marital outcomes while delaying marriage worsening them. 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.[8] 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.

2. STUDY SETTING, DESIGN, AND DATA SOURCES

2.1 Study Setting

The "Schooling, Income, and Health Risks" study (SIHR) tracks the lives of a sample of 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

---

[8] 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".

more than two years after the cessation of the cash transfer experiment in December 2009, tracking the adolescents as many of 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, 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 about a third had done any wage work in the past three months. 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 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 conditional cash transfers (CCTs), while a further experiment was performed within the larger cohort of *baseline schoolgirls*. For them, 46 EAs were assigned to CCTs, 27 were assigned to unconditional cash transfers (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.[9] 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.[10] Four rounds of data took place: Round 1-Baseline (2007), Round 2 (2008), Round 3

---

[9] 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 (WDI, 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.

[10] 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

(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).[11]

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*.[12]

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

---

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.

[11] 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.

[12] 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 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.

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.

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; please see Appendix A). 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.[13] A detailed description of all outcomes in this paper is provided in Appendix B.

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 C.

---

[13] 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.

The evaluation of the impact of the Zomba Cash Transfer Program utilizes the experimental design of the intervention for causal identification. 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:

$$Y_{ic} = \alpha + \gamma^c T_c^c + \gamma^u T_c^u + \beta X_{ic} + \varepsilon_{ic} \qquad (1)$$

where Y_{ic} is an outcome variable for core-respondent i in cluster c, T_c^C and T_c^U are binary indicators for offers in the CCT and the UCT clusters, respectively, and X_{ic} is a vector of baseline characteristics. Note that for baseline dropouts we only have the CCT binary indicator. The standard errors ε_{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.[14] 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.

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

---

[14] 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.

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.

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.[15]

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, 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

---

[15] 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.

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 (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 during, 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 the 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 increased by half a year by 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).[16] 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, 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

---

[16] 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. (2015). 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*.

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 young females 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 dropouts*; 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).[17]

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

---

[17] 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.

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) 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 have shown 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 S9-11, we report the original impact estimates for 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.[18] 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 point 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* 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.[19]

4.6 Husband Characteristics

The program impacts on empowerment presented above, particularly the negative effects apparent in the UCT group,[20] motivate the examination of marriage market outcomes. As described earlier, two years after the end of the transfer program, CCT beneficiaries among

---

[18] The Stata code and the paper that describes the method can be found here:
https://are.berkeley.edu/~mlanderson/ARE_Website/Research.html.
[19] 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.
[20] Note that the negative empowerment result among married women remains robust to adjustments for IPW, Lee bounds, and multiple-hypothesis testing.

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.[21] 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 matrices test (p-value=0.03). The differences between the CCT and UCT groups for

---

[21] A joint F-test of interactions between treatment (CCT) and baseline attributes predicting selection into the husband sample among baseline dropouts is insignificant.

the overall husband quality index, as well as MSCE and cognitive ability, are all statistically significant.[22]

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. (2015) 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).[23]

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

---

[22] 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.

[23] 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.

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.

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.[24] 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 2016), 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

---

[24]Agüero, 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.

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 the program, while CCTs might be most beneficial to children born after the mother's additional human capital accumulation took place.[25]

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 *baseline dropouts* and only one (out of eight outcomes) among *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

---

[25] 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.

causal impacts of the program on a specific child, because childbearing is endogenous to treatment. However, as the causal effects of cash transfers targeted to females of childbearing age (rather than to pregnant women or mothers) is an important policy question, we attempt to disentangle the selection-driven components of fertility and parentage from the direct treatment effects on the actual sample of children born. The technical details of the assumptions required and the sequence of adjustments – to move from the overall reduced-form difference between children born to mothers in the treatment and control groups towards a more standard causal effect on the children actually born – are outlined in Appendix C.

To investigate how differential exposure to CCTs and UCTs drives treatment effects, we consider the sample of children born during three epochs. The first epoch captures those *directly exposed* to the program, meaning those born during the program.[26] This cohort is exposed for a maximum of two years, with some combination of in utero and child exposure depending on the exact birth date of the child. The second epoch covers those born within nine months of the end of the program, who were *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 *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).[27]

---

[26] 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.

[27] 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%

Figure III plots the "raw" differences in HAZ for children under 60 months between the treatment and the control groups.[28] 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.

These impacts 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.[29] Following the methodology laid out in Appendix C, we can then

---

(Haddad et al. 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.

[28] 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.

[29] 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, Black et al. 2014, Adhvaryu et al. 2016), 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

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 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 technical appendix, 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

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.

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.[30] 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, particularly those given unconditionally, on adolescent females. 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

---

[30] 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*.

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.[31] 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 (Heath and Mobarak 2015; Jensen 2012; Munshi and Rosenzweig 2006; Oster and Steinberg 2013). 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

---

[31] 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.

incentives for formal schooling, improved only cognitive skills, which may not have been sufficient to increase productivity.[32]

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.[33] 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

---

[32] 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. (2015) 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.

[33] 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).

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' 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.[34] 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

---

[34] 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.

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.

REFERENCES

Adhvaryu, Achyuta, Prashant Bharadwaj, James Fenske, Anant Nyshadham, and Richard Stanley. 2016. Dust and Death: Evidence from the West African Harmattan. *CSAE Working Paper WPS/2016-03.* Centre for the Study of African Economies, University of Oxford.

Agüero, Jorge, Michael Carter, and Ingrid Woolard. 2006. The Impact of Unconditional Cash Transfers on Nutrition: The South African Child Support Grant. SALDRU Working Paper Series No. 06/08.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. The Long-Run Impact of Cash Transfers to Poor Families. *Am Econ Rev* 106 (4): 935-971.

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.

Almond, D. 2006. Is the 1918 influenza pandemic over? Long-term effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114 (4): 672-712.

Almond, Douglas and Bhashkar A Mazumder. 2011. Health capital and the prenatal environment: the effect of Ramadan observance during pregnancy. *American Economic Journal: Applied Economics* 3 (4): 56-85.

Anderson, Michael L. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association* 103 (484): 1481-1495.

Anderson, Siwan and Chris Bidner. 2015. Property Rights over Marital Transfers. *Q J Econ* 130 (3): 1421-1484.

Araujo, M Caridad, Mariano Bosch, and Norbert Schady. 2016. Can Cash Transfers Help Households Escape an Inter-Generational Poverty Trap?. *NBER Working Paper 22670.* National Bureau of Economic Research.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena. 2015. Bride Price and Female Education. *Unpublished manuscript*:

http://scholar.harvard.edu/files/nbau/files/paper_draft_indonesia_zambia_20140821.pdf.

Baez, Javier Eduardo and Adriana Camacho. 2011. Assessing the long-term effects of conditional cash transfers on human capital: evidence from Colombia. *IZA Discussion Paper No. 5751*.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. Cash or Condition? Evidence from a Cash Transfer Experiment. *Q J Econ* 126 (4): 1709-1753.

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).

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. Income shocks and adolescent mental

health. *Journal of Human Resources* 48 (2): 370-403.

Baird, Sarah and Berk Özler. 2016. Sustained Effects on Economic Empowerment of Interventions for Adolescent Girls: Existing Evidence and Knowledge Gaps. CGD Background Paper: http://www.cgdev.org/sites/default/files/sustained-effects-economic-empowerment.pdf.

Baird, Sarah J, Richard S Garfein, Craig T McIntosh, and Berk Özler. 2012. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet* 379 (9823): 1320-1329.

Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2015. Women's empowerment in action: evidence from a randomized control trial in Africa. *Unpublished manuscript*: http://sticerd.lse.ac.uk/dps/eopp/eopp50.pdf.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. A multifaceted program causes lasting progress for the very poor: Evidence from six countries. *Science* 348 (6236): 1260799.

Barham, Tania, Karen Macours, and John A Maluccio. 2013. Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *Am Econ Rev* 103 (3): 467-471.

Barrera-Osorio, Felipe, Leigh L Linden, and Juan E Saavedra. 2015. Medium Term Educational Consequences of Alternative Conditional Cash Transfer Designs: Experimental Evidence from Colombia. *CESR-Schaeffer Working Paper*, no. 2015-026.

Bazzi, Samuel, Sudarno Sumarto, and Asep Suryahadi. 2015. It's all in the timing: Cash transfers and consumption smoothing in a developing country. *Journal of Economic Behavior & Organization* 119:267-288.

Behrman, Jere R, Susan W Parker, and Petra E Todd. 2011. Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? *Journal of Human Resources* 46 (1): 93-122.

Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 289-300.

Black, Sandra E, Aline Bütikofer, P Devereux, and K Salvanes. 2014. This Is Only a Test? Long-Run and Intergenerational Impacts of Prenatal Exposure to Radioactive Fallout. *Unpublished manuscript.*

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2016. Does Grief Transfer across Generations? Bereavements during Pregnancy and Child Outcomes. *American Economic Journal: Applied Economics* 8 (1): 193-223.

Bruhn, Miriam and David McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1 (4): 200-232.

Buehren, Niklas, Markus Goldstein, Selim Gulesci, Munshi Sulaiman, and Venus Yam. 2015. Evaluation of Layering Microfinance on an Adolescent Development Program for Girls in

Tanzania. *Unpublished manuscript:* https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2016&paper_id=1018.

Canning, David, Sangeeta Raja, and Abdo S Yazbeck. 2015. Africa's Demographic Transition: Dividend Or Disaster?

Chay, Kenneth Y and Michael Greenstone. 2003. The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *Quarterly Journal of Economics* 118 (3): 1121-1167.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *Am Econ Rev* 106 (4): 855-902.

Currie, Janet and Douglas Almond. 2011. Human capital development before age five. *Handbook of Labor Economics* 4:1315-1486.

Dahl, Gordon B and Lance Lochner. 2012. The impact of family income on child achievement: Evidence from the earned income tax credit. *Am Econ Rev* 102 (5): 1927-1956.

Duflo, Esther. 2012. Women Empowerment and Economic Development. *Journal of Economic Literature* 50 (4): 1051-1079.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. Education, HIV, and Early Fertility: Experimental Evidence from Kenya. *Am Econ Rev* 105 (9): 2757-97.

Fan, Jianqing. 1992. Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87 (420): 998-1004.

Field, Erica and Attila Ambrus. 2008. Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy* 116 (5): 881-930.

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

Fiszbein, Ariel, Norbert Rüdiger Schady, and Francisco H G Ferreira. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*.

   http://public.eblib.com/choice/publicfullrecord.aspx?p=459451.

Foster, Andrew and Nizam Khan. 2000. Equilibrating the marriage market in a rapidly growing population: Evidence from rural Bangladesh. *Unpublished manuscript*: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.536.2227&rep=rep1&type=pdf.

Gender Equality and Development. 2011. Gender Equality and Development

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

Gladstone, M J, G A Lancaster, A P Jones, K Maleta, E Mtitimila, P Ashorn, and R L Smyth. 2008. Can Western developmental screening tools be modified for use in a rural Malawian setting? *Archives of Disease in Childhood* 93 (1): 23-29.

Glennerster, Rachel. 2013. Empowering Girls. Available at:

   http://www.3ieimpact.org/media/filer_public/2013/07/19/girls_empowerment_in_banglade

sh_delhi.pdf.

Goodman, Robert. 2001. Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry* 40 (11): 1337-1345.

Haddad, Lawrence James, Endang Achadi, Mohamed Ag Bendech, Arti Ahuja, Komal Bhatia, Zulfiqar Bhutta, Monika Blössner, Elaine Borghi, Esi Colecraft, and Mercedes de Onis. 2014. Global Nutrition Report 2014: Actions and accountability to accelerate the world s progress on nutrition

Handa, Sudhanshu, Luisa Natali, David Seidenfeld, and Gelson Tembo. 2016. The impact of Zambia's unconditional child grant on schooling and work: results from a large-scale social experiment. *Journal of Development Effectiveness* 8 (3): 346-367.

Haushofer, Johannes and Jeremy Shapiro. 2016. The Short-Term Impact Of Unconditional Cash Transfers To The Poor: Experimental Evidence From Kenya. *Quarterly Journal of Economics* 131 (4): 1973-2042.

Heath, Rachel and A. Mushfiq Mobarak. 2015. Manufacturing growth and the lives of Bangladeshi women. *Journal of Development Economics* 115:1-15.

Heckman, James J and Chase O Corbin. 2016. Capabilities and Skills. *Journal of Human Development and Capabilities* 17 (3): 342-359.

Heckman, James J and Stefano Mosso. 2014. The economics of human development and social mobility. *IZA DP No. 800.*

Heckman, James J and Tim Kautz. 2013. Fostering and measuring skills: Interventions that improve character and cognition. *NBER Working Paper 19656.* National Bureau of Economic Research.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. Long-Run Impacts of Childhood Access to the Safety Net. *Am Econ Rev* 106 (4): 903-934.

Jensen, Robert. 2012. Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *Q J Econ* 127 (2): 753-792.

LaFave, Daniel and Duncan Thomas. 2016. Height and Cognition at Work: Labor Market Productivity in a Low Income Setting. *NBER Working Paper 22290.* National Bureau of Economic Research.

Lee, David S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* 76 (3): 1071-1102.

Levine, Ruth, Cynthia Lloyd, Margaret Greene, and Caren Grown. 2008. Girls Count: A global investment and action agenda. Reprint, 2009. Washington, D.C.: Center for Global Development.

Lloyd, Cynthia B and Juliet Young. 2009. New lessons: the power of educating adolescent girls: a Girls Count report on adolescent girls. The Population Council, Inc. http://www.popcouncil.org/uploads/pdfs/2009PGY_NewLessons.pdf.

Manley, James, Seth Gitter, and Vanya Slavchevska. 2013. How effective are cash transfers at improving nutritional status? *World Development* 48:133-155.

Milligan, Kevin and Mark Stabile. 2009. Do child tax benefits affect the wellbeing of children?

Evidence from Canadian child benefit expansions. *Am Econ Rev* 99 (2): 128-132.

Molina-Millan, Teresa, Tania Barham, Karen Macours, John A Maluccio, and Marco Stampini. 2016. Long-Term Impacts of Conditional Cash Transfers in Latin America: Review of the Evidence Inter-American Development Bank Technical Note No. IDB-N-923.

Molyneux, Maxine, With Nicola Jones, and Fiona Samuels. 2016. Can Cash Transfer Programmes Have 'Transformative' Effects? *The Journal of Development Studies* 52 (8): 1087-1098.

Munshi, Kaivan D and Mark R Rosenzweig. 2006. Traditional institutions meet the modern world: Caste, gender and schooling choice in a globalizing economy. *American Economic Review* 96 (4): 1225-1252.

National Statistical Office and ICF Macro. 2005. Malawi Demographic and Health Survey 2010

Naudeau, Sophie, Rifat Hasan, and Anne Bakilana. 2015. Adolescent Girls in Zambia: Introduction and Overview. Policy Brief: Zambia:

http://elibrary.worldbank.org/doi/abs/10.1596/24597.

Oster, Emily and Bryce Millett Steinberg. 2013. Do IT service centers promote school enrollment? Evidence from India. *Journal of Development Economics* 104:123-135.

Ozier, Owen W. 2015. The impact of secondary schooling in Kenya: a regression discontinuity analysis. *World Bank Policy Research Working Paper*, no. 7384.

Saavedra, Juan E and Sandra Garcia. 2013. Educational impacts and cost-effectiveness of conditional cash transfer programs in developing countries: A meta-analysis. *CESR Working Paper*, no. 2013-007.

Thomas, Duncan, John Strauss, and Maria-Helena Henriques. 1991. How does mother's education affect child height? *Journal of Human Resources* 26 (2): 183-211.

Woerner, Wolfgang, Bacy Fleitlich-Bilyk, Rhonda Martinussen, Janet Fletcher, Giulietta Cucchiaro, Paulo Dalgalarrondo, Mariko Lui, and Rosemary Tannock. 2004. The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *European Child and Adolescent Psychiatry* 13 (2): ii47-ii54.

Zomba City Assembly. 2009. Zomba district socio economic profile 2009-2012.

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

Panel A: Baseline Dropouts

	Highest Grade Completed			English Test Score (Standardized)	TIMMS Math Score (Standardized)	Non-TIMMS Math Score (Standardized)	Cognitive Test Score (Standardized)	Competencies Score (Standardized)
	During Program	End of Program	Two Years After Program		End of Program			Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Conditional Schoolgirl	0.579***	0.558***	0.621***	0.079	0.147***	0.116	0.163**	0.064
	(0.073)	(0.102)	(0.125)	(0.071)	(0.056)	(0.072)	(0.070)	(0.057)
Mean in Control Group	6.345	6.967	6.997	0.000	0.000	0.000	0.000	0.000
Sample Size	697	718	744	704	704	704	704	742
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	0.078	0.126*	0.120	0.148***	0.136**	0.068	0.181***	0.065
	(0.090)	(0.069)	(0.080)	(0.056)	(0.069)	(0.063)	(0.050)	(0.058)
=1 if Unconditional Schoolgirl	0.122	0.103	0.095	-0.068	-0.027	0.026	0.094	0.098
	(0.109)	(0.121)	(0.129)	(0.090)	(0.106)	(0.090)	(0.129)	(0.067)
p-value UCT vs. CCT	0.708	0.854	0.850	0.035	0.157	0.657	0.514	0.630
p-value Treatment	0.469	0.174	0.309	0.021	0.118	0.560	0.002	0.297
Mean in Control Group	8.590	9.677	10.415	0.000	0.000	0.000	0.000	0.000
Sample Size	1,965	2,019	2,049	2,000	2,000	2,000	2,000	2,048

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

	=1 if Ever Married			Age First Marriage	=1 if Ever Pregnant			Number of Live Births			Age at First Birth	Desired Fertility
	During Program	End of Program	Two Years After Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	Two Years After Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
=1 if Conditional Schoolgirl	-0.140***	-0.157***	-0.107***	0.431***	-0.057*	-0.081***	-0.040*	-0.005	-0.095**	-0.147***	0.272*	-0.172*
	(0.029)	(0.037)	(0.032)	(0.155)	(0.030)	(0.027)	(0.021)	(0.033)	(0.044)	(0.054)	(0.164)	(0.087)
Mean in Control Group	0.291	0.575	0.809	19.644	0.610	0.784	0.924	0.520	0.819	1.380	18.499	3.217
Sample Size	698	718	744	500	698	718	744	698	718	744	634	744
Panel B: Baseline Schoolgirls												
=1 if Conditional Schoolgirl	0.000	-0.010	-0.035	-0.011	0.008	0.027	-0.024	0.023*	0.003	0.020	-0.144	-0.072
	(0.012)	(0.024)	(0.027)	(0.148)	(0.015)	(0.027)	(0.034)	(0.014)	(0.022)	(0.036)	(0.136)	(0.064)
=1 if Unconditional Schoolgirl	-0.033***	-0.083***	-0.010	0.486**	-0.013	-0.063**	-0.001	0.013	-0.055*	-0.024	0.001	-0.017
	(0.012)	(0.024)	(0.046)	(0.200)	(0.017)	(0.028)	(0.042)	(0.017)	(0.030)	(0.046)	(0.168)	(0.056)
p-value UCT vs. CCT	0.026	0.018	0.613	0.032	0.314	0.009	0.614	0.641	0.075	0.410	0.436	0.477
p-value Treatment	0.023	0.004	0.448	0.050	0.600	0.025	0.760	0.209	0.151	0.705	0.547	0.533
Mean in Control Group	0.047	0.180	0.402	18.651	0.092	0.247	0.501	0.055	0.199	0.511	18.718	2.974
Sample Size	1,967	2,018	2,049	821	1,966	2,019	2,049	1,966	2,019	2,049	998	2,048

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

	=1 if HIV Positive			=1 if Anemic
	During Program	End of Program	Two Years After Program	Two Years After Program
	(1)	(2)	(3)	(4)
=1 if Conditional Schoolgirl	0.022	0.020	0.012	0.039
	(0.024)	(0.023)	(0.026)	(0.035)
Mean in Control Group	0.06	0.094	0.135	0.255
Sample Size	373	694	715	711
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	-0.020**	-0.003	-0.001	0.012
	(0.009)	(0.011)	(0.019)	(0.031)
=1 if Unconditional Schoolgirl	-0.015	-0.019*	-0.002	-0.065*
	(0.012)	(0.012)	(0.023)	(0.033)
p-value UCT vs. CCT	0.616	0.237	0.980	0.068
p-value Treatment	0.112	0.249	0.996	0.122
Mean in Control Group	0.026	0.035	0.055	0.243
Sample Size	1,192	2,002	1,977	1,979

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 12d/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

	Labor Market Outcomes			Empowerment				
	Opportunity Cost of Time	Typical Wage in Past Three Months	Proportion of Hours Spent in Self-Employment or Paid Work in Past Week	Super-Index of Overall Empowerment (Standardized)	Change in Subjective Wellbeing from Five Years Ago to Today	Super-Index of Unmarried Empowerment (Standardized)	Super-Index of Married Empowerment (Standardized)	Married Index of Economic Control (Standardized)
				Two Years After Program				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Conditional Schoolgirl	-0.037	-0.140**	-0.011	-0.083	-0.032	0.018	-0.113	-0.118
	(0.079)	(0.068)	(0.009)	(0.074)	(0.232)	(0.112)	(0.102)	(0.096)
Mean in Control Group	0.707	0.375	0.061	0.000	1.120	0.000	0.000	0.000
Sample Size	718	743	744	744	744	289	455	455
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	-0.051	-0.011	0.003	0.049	0.276	0.111	0.068	-0.107
	(0.101)	(0.058)	(0.005)	(0.082)	(0.187)	(0.098)	(0.095)	(0.108)
=1 if Unconditional Schoolgirl	-0.115	0.036	0.002	-0.159*	0.176	-0.094	-0.342***	0.147
	(0.074)	(0.104)	(0.008)	(0.081)	(0.190)	(0.109)	(0.099)	(0.307)
p-value UCT vs. CCT	0.550	0.665	0.842	0.052	0.650	0.120	0.001	0.406
p-value Treatment	0.297	0.910	0.784	0.101	0.306	0.287	0.001	0.484
Mean in Control Group	0.897	0.212	0.029	0.000	0.906	0.000	0.000	0.000
Sample Size	2,002	2,048	2,045	2,049	2,049	1,271	776	774

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% (***), 95% (**), and 90% (*) confidence.

Table V: Program impacts marriage market outcomes (husband characteristics)

Panel A: Baseline Dropouts

	Husband Quality Super Index (Standardized)	Highest Grade Completed	=1 if Passed Primary School (PSLC)	=1 if Passed Junior Secondary School (JCE)	=1 if Passed Secondary School (MSCE)	Cognitive Test Score (Standardized)	Typical Wage in Past Three Months	=1 if Currently Employed	Sexual Activity and Marital Fidelity (Standardized)	=1 if Does Not Suffers from Psychological Distress	=1 if HIV Positive
					Two Years After Program						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
=1 if Conditional Schoolgirl	0.084	0.561	0.032	0.029	0.074**	-0.049	-0.081	-0.024	0.032	0.007	-0.005
	(0.106)	(0.348)	(0.054)	(0.046)	(0.037)	(0.110)	(0.225)	(0.040)	(0.106)	(0.061)	(0.035)
Mean in Control Group	0.000	7.806	0.526	0.314	0.097	0.000	1.194	0.246	0.000	0.634	0.055
Sample Size	326	326	326	326	326	323	325	326	325	326	265

Panel B: Baseline Schoolgirls

=1 if Conditional Schoolgirl	0.141	0.046	0.024	0.012	0.059	0.014	0.014	0.045	0.284***	0.074	0.001
	(0.096)	(0.271)	(0.043)	(0.049)	(0.053)	(0.109)	(0.262)	(0.051)	(0.091)	(0.060)	(0.033)
=1 if Unconditional Schoolgirl	-0.186	-0.454	0.005	0.017	-0.088	-0.357**	-0.406	-0.091	0.013	0.008	0.010
	(0.180)	(0.425)	(0.068)	(0.086)	(0.054)	(0.163)	(0.344)	(0.093)	(0.219)	(0.093)	(0.041)
p-value UCT vs. CCT	0.084	0.240	0.776	0.954	0.042	0.044	0.225	0.17	0.196	0.508	0.845
p-value Treatment	0.145	0.490	0.845	0.964	0.118	0.087	0.432	0.358	0.006	0.471	0.971
Mean in Control Group	0.000	9.743	0.699	0.541	0.258	0.000	1.42	0.352	0.000	0.647	0.052
Sample Size	543	543	543	543	543	539	540	543	542	541	457

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 Questionaire-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

	Height-for-Age z-score	Neonatal Mortality	Postneonatal Mortality	Parenting Practices Percentage Score	Exclusively Breastfed for First 6 Months	Malawi Developmental Assessment Tool (3-4 year-olds) (Standardized)	Reported Child Difficulties (3-4 year-olds) (Standardized)	Reported Pro-Social Behaviors (3-4 year-olds) (Standardized)
				Two Years After Program				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Treatment Dropout	-0.013	0.013	-0.009	-0.003	0.030	-0.086	0.104	0.123
	(0.091)	(0.011)	(0.013)	(0.018)	(0.026)	(0.112)	(0.190)	(0.157)
Mean in Control Group	-1.351	0.015	0.026	0.496	0.804	0.000	0.000	0.000
Sample Size	742	958	707	861	971	213	223	223
Panel A: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	0.096	-0.014	0.005	0.012	0.029	-0.294*	-0.011	-0.357
	(0.109)	(0.009)	(0.012)	(0.018)	(0.033)	(0.176)	(0.180)	(0.282)
=1 if Unconditional Schoolgirl	0.065	-0.012	0.001	0.050*	0.126***	0.213	0.035	-0.132
	(0.176)	(0.012)	(0.010)	(0.029)	(0.039)	(0.376)	(0.173)	(0.309)
p-value UCT vs. CCT	0.872	0.901	0.734	0.229	0.014	0.172	0.835	0.568
p-value Treatment	0.666	0.302	0.912	0.215	0.006	0.145	0.974	0.434
Mean in Control Group	-1.410	0.028	0.013	0.484	0.771	0.000	0.000	0.000
Sample Size	1,032	1,167	756	1,090	1,169	185	196	196

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**)

Panel A: Born During Program	Raw Effect				Direct Effect
	Gender	+ Maternal Selection weights	+ Paternal Selection Controls	+ Child Age	+ Mother Age
	(1)	(2)	(3)	(4)	(5)
=1 if Conditional Schoolgirl	-0.015	-0.174	-0.139	-0.154	-0.051
	(0.128)	(0.149)	(0.143)	(0.140)	(0.136)
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
Father selection controls			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**)

	Raw Effect				Direct Effect
Panel A: Born During Program	Gender	+ Maternal Selection weights	+ Paternal Selection Controls	+ Child Age	+ Mother Age
	(1)	(2)	(3)	(4)	(5)
=1 if Conditional Schoolgirl	0.155	-0.050	-0.054	0.023	0.124
	(0.162)	(0.192)	(0.186)	(0.177)	(0.155)
=1 if Unconditional Schoolgirl	0.953**	0.525**	0.549*	0.666**	0.523*
	(0.476)	(0.221)	(0.306)	(0.315)	(0.299)
p-value UCT vs. CCT	0.091	0.022	0.028	0.024	0.115
p-value Treatment	0.123	0.040	0.089	0.072	0.218
Sample Size	315	315	315	315	315
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		X	X	X	X
Father selection controls			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.

	Treatment EAs (88 Clusters)			Control EAs (N=88)
	Conditional (N=61*)		Unconditional (N=27)	
Baseline Dropouts (N=889)	CCT		CCT	Pure control
Baseline Schoolgirls (N=2,907)	CCT	Within-village control	UCT	Within-village control	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
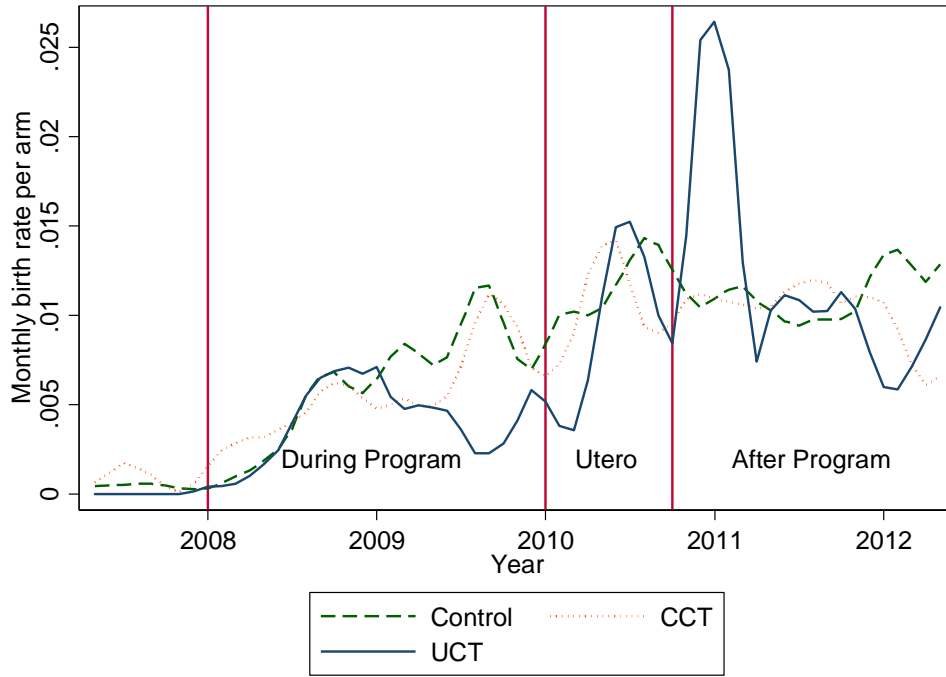
Panel B: Monthly Marriage Rates
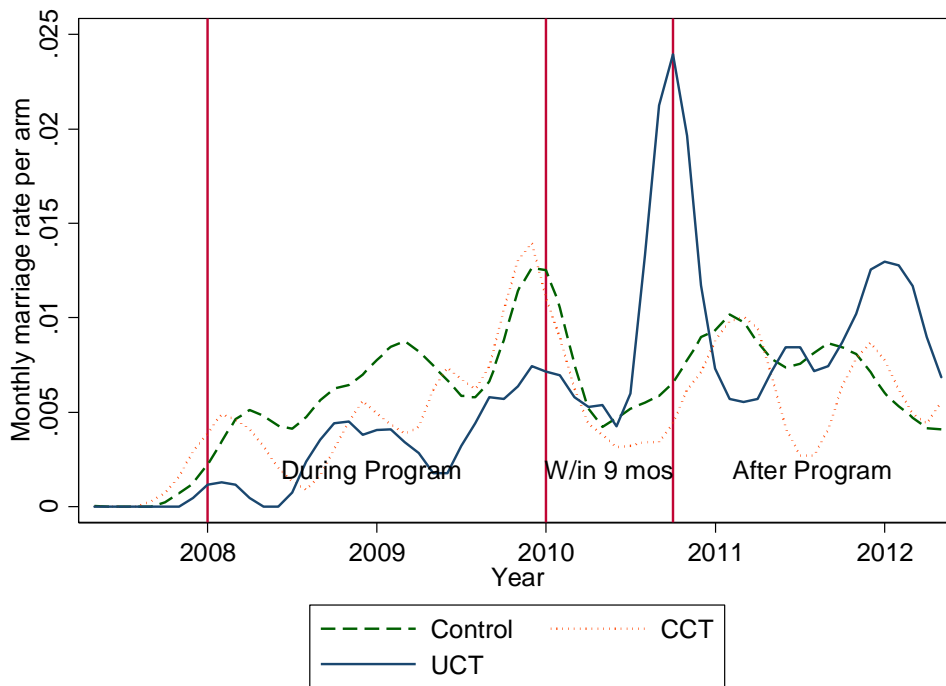
Notes: Figures illustrate the smoothed fraction of core respondents who give birth (Panel A) or get married (Panel B) in each month using retrospective information on the month of birth and marriage, respectively.

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

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

Panel B: Baseline Schoolgirls, CCT

Raw and Direct CCT Schoolgirl effect on Child HAZ
With 95 % Confidence Intervals

Adjustments include propensity weighting, child and maternal age, maternal and paternal attributes
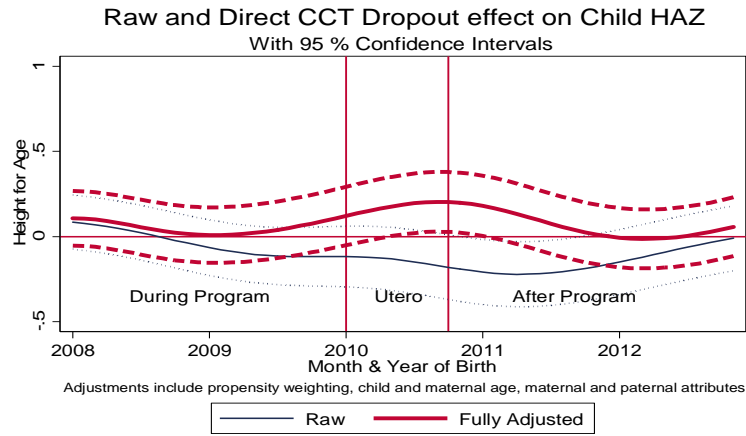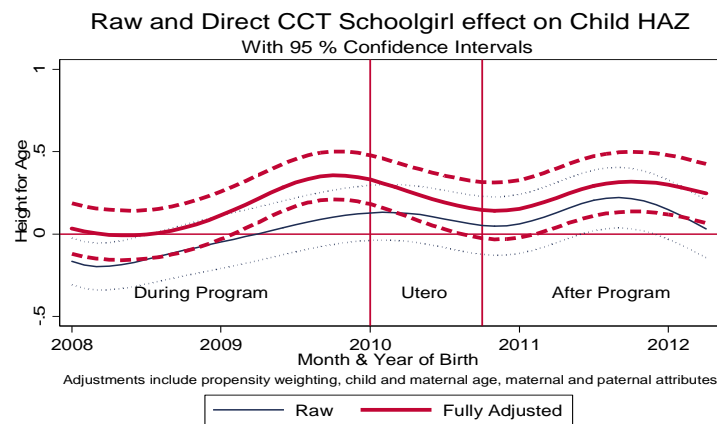
Panel C: Baseline Schoolgirls, UCT

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
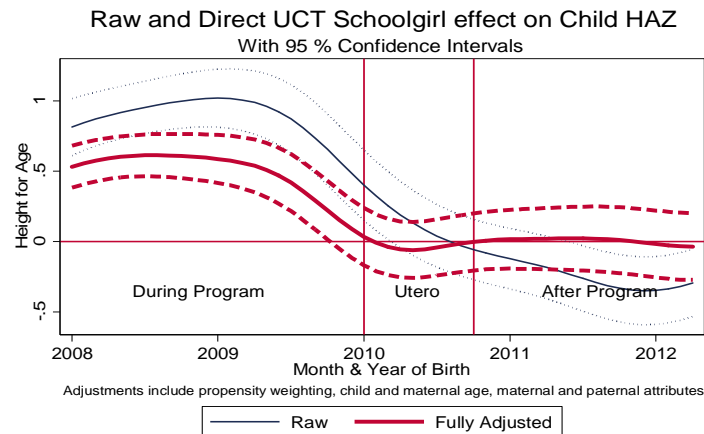
Supplemental Online Appendix for:

"WHEN THE MONEY RUNS OUT:
DO CASH TRANSFERS HAVE SUSTAINED EFFECTS?"

Sarah Baird, Craig McIntosh and Berk Özler*

*Corresponding author. E-mail: bozler@worldbank.org

November 30, 2016

This PDF file includes:

Appendix A: Pre-Analysis Plan for Round 4 Schooling, Income and Health Risk in Malawi (SIHR) Data (which can also be found at https://www.socialscienceregistry.org/trials/36)

Principal Investigators:

Sarah Baird, University of Otago
Ephraim Chirwa, Chancellor College
Craig McIntosh, UCSD
Berk Özler, World Bank and University of Otago

Analysis plan:

The core analysis will compare the impact of the Conditional Cash Transfer (CCT) and Unconditional Cash Transfer (UCT) treatment to control EAs for the baseline schoolgirl stratum, and will compare the CCT treatment to the control for the baseline dropout stratum. Most of the analysis will consist of Round 4 cross-sectional regression (using OLS unless not appropriate), although where possible we will also pursue panel difference-in-differences analysis for variables that have been consistently collected in multiple rounds. For consistency, the analysis will include the full set of controls used in the paper by Baird, McIntosh and Özler (2011). These controls include baseline values of the following: a household asset index, highest grade attended, a dummy variable for having started sexual activity, dummy variables for age, and strata dummies. Standard errors will be clustered at the EA level, and results will be weighted to make them representative of the target population in the study EAs.

Only if a significant impact is found in the core analysis will further heterogeneity of impact be explored. Heterogeneity will be explored along both experimental dimensions including the amount of the transfer and the split between the parent and the girl, as well as based on the age at which you got the program and differences between rural and urban.

Note on construction of indexes:

To construct indexes for classes of variables, we will adhere to the following rubric:

a) For each sub-question in a family of variables, first align answers so that higher numbers always have a consistent meaning (good or bad).
b) Calculate the mean and SD of the responses to each sub-question in the sample *in the control group – separately for baseline schoolgirls and baseline dropouts*.
c) Create normalized variables that have the mean subtracted off and are divided by the SD.
d) Calculate the raw mean of the normalized variables for all sub-questions within a family of variables. This mean is the 'index' for those variables. *This summary index can further be normalized if desired.*

For the core analysis we will not pursue the analysis of sub-variables within an index unless the index as a whole is significant.

Construction of Core Indexes: Primary outcomes are indicated in bold text.

1. <u>Core Respondent-level outcomes:</u> Can be analysis with simple cross-sections comparing UCT, CCT, and control. No extensive margin issue with any of these variables.
    a. Schooling and Marriage (replication of QJE results with age-appropriate dependent variables):
        i. **Highest grade completed (S7, Q7)**
        ii. Highest educational qualification achieved (S7, Q9)
        iii. **Achievement, replacing the test scores with the 'competencies' (see end of this appendix for further details on construction). We will show the components, as well as the index for the quality index; and show only an index for the quantity.**
        iv. **Ever married (Part II CS, Q2e), ever pregnant (S18, Q1, Q2), number of live births (S18, Q17).**
        v. **Hazard model of age of first marriage (S14, Q1 (and Round 3 data for those already married at Round 3)) and age at first birth (construct using age of respondent and DOB), with 'uncompleted spells' for those never married or never first birth.**
        vi. Sexual behavior: ABC, # partners ever, as in previous papers. Ever had sex (S12, Q2, Q3, Q4) age at first sex (S12 Q4), total number of partners ever (S12, Q5), sexually active in past 12 months (S12, Q7), condom use last sex with most recent partner (S12, Q23), having a partner with an age difference of more than 5 years (S12Q12)

    b. Health: Replication of Lancet results (Baird, Garfein, McIntosh and Özler 2012)
        i. **HIV prevalence in R4,** HIV incidence R3-R4.
        ii. **Anemia: construct binary measure based on definition of mild anemia from WHO and used in survey (with different thresholds if pregnant),** (with bednets, breastfeeding, timing of last meal, taking medical for anemia, and menstruating as secondary moderator channel) (VCT, S1).
        iii. Use of reliable birth control: S12 Q27 anything except withdrawal, periodic abstinence, other
        iv. Desired fertility S16 Q4 or Q10.
        v. Mental health, calculated as in Baird, de Hoop, Ozler (2013) (S9, Q9-20): binary
        vi. Number of meals eaten with meat, eggs, fish in past 7 days (S9, Q 6-8)

    c. Empowerment & aspirations:
        i. Index of self-efficacy: S11a Q1-10.
        ii. Index of preferences for child education: S11a Q17-25.
        iii. Index of social participation: S11a Q13,14,16.
        iv. Aspirations: Change in ladder from five years ago to five years from now (S9, Q23-Q21)
        v. **Change in ladder from five years ago to today (S9, Q22-Q21)**
    Super-index of overall empowerment i-iv.

d. Wages/Employment:
   i. Effective wage: S8 Q2 d-g, Q6/Q4 – converted into a daily wage rate.
   ii. **Opportunity cost of time: Minimum wage from S8 Q13-17.**
   iii. Labor income: S8 Q7.
   iv. **Typical wage: S8 Q10.**
   v. Any wage work in past 3 months (S8 Q9)
   vi. **Sector of employment: S8: Sum of Q4 a&c divided by the whole sum of Q4.**

e. Consumption:
   i. Impact on household-level consumption aggregate (Real Total comp. monthly consumption in market unit prices in USD per person), constructed as in previous rounds

2. <u>Married Core Respondent Outcomes:</u>   to be analyzed within the sample of married core respondents, contextualized by impacts on marriage rates and age at first marriage but no attempt at a 'Heckman-style' correction. Secondary analysis will delve more into intensive/extensive margin impacts as necessary.

   a. Core Respondent's Empowerment married:
      i. Index of financial decisionmaking: S15 Q1-5, 'Resp'=2. 'Joint'=1, other=0[1]
      ii. Index of marital satisfaction: S15 Q20-29.
      iii. Index of women's divorce prospects: S15 Q39-40, S15 Q48a-d
      iv. Index of fertility disempowerment: S16 Q6==3 & Q7==(1|2), Q39-40.
      v. Index of self-determination in marriage: S16 Q20-28
      vi. Index of frequency of social contact: S16 Q29+30+32
      vii. Index of spousal abuse: S16 Q41-46, 47, 50, 52
      viii. Age difference between wife and husband (S12, Q12-Part II CS, Q2c)
      ix. Female ag decisionmaking power: If the answer to any of the questions S3 Q12, Q16, or Q17 is the "CR" for any plot, the dummy variable for the entire HH is defined as a '1' and '0' otherwise.
      x. Female microenterprise participation: S6 Q11 Core Resp controlling any enterprise profits
      xi. Female livestock control: S3 Q22: Core Resp responsible for any livestock decisionmaking.
      xii. Ratio of female to male-specific consumption: calculate total spending for CR (per month or per 12 months) on all items asked and divide it by the same variable calculated for the husband from Section 27 (S10 Q3,4,7,8 normalized sum) /   (S27 Q3,4,7,8 normalized sum).

      Super-index of empowerment: i-viii.
      Super-index of economic control: iv-xii.

---

[1] Also, as a rule, indices for each individual include the items with no missing values. We do not impute values for item non-response, or exclude variables or individuals

    b. Husband Quality.
        i. Husband's highest grade completed, highest certificate attained. S25 Q2,4
        ii. Husband's wage rate S26 Q5
        iii. Currently employed S26 Q6..
        iv. Husband's score on cognitive test
        v. Husband HIV status.
        vi. Husband marital fidelity. Partners ever: S32 Q2, Partners 12 mo. S32
             Q3. Concurrence: S32 Q15 answer for spouse (column 1)
        vii. Husband's mental health (constructed in same manner as CR) and then
             standardized.
    Super-index of husband quality: i-vii.

    c. Husband Gender Empowerment.
        i. Husband Index of GE: S30 Q1-9
        ii. Husband Index of wife's autonomy: S30 Q13-21.
        iii. Husband Index of justification for abuse S30 Q26-28.
        iv. Husband Index of divorce prospects: S31 Q10-11, 15,16
        v. Husband desired fertility. S30 Q49
    Super-index of husband gender empowerment: i-v.

3. <u>Unmarried Core Respondent Outcomes:</u> to be analyzed within the sample of unmarried core respondents.

    a. Empowerment unmarried:
        i. Index of autonomy: S17 Q2a-d, Q4, Q6
        ii. Index of abuse: S17 Q14,16,17a-e.
    Super-index of unmarried empowerment: i-ii.

4. <u>Child-Level Outcomes:</u> Can be analyzed two different ways. First, unconditionally examining simple comparison between treatment and control. Secondly, conditionally including dummies for the age of child and the age of the mother so as to drop out effects coming from delayed fertility.
    a. Health Outcomes:
        i. Birthweight: S21 Q20
        ii. Vaccinations: S22 Q5, 7, 8, 10, 11 (index)
        iii. Neonatal/Infant/Child mortality: S18 Q16-21; S21 Q10 (use Round 3 data for children who had already passed away by Round 3)
        iv. Bednets: S22 Q 13

    b. Parental Practices:
        i. Breastfeeding (6 months exclusive or to death is died younger than 6 months): S21 Q22, Q23, Q25
        ii. Parenting: S23 Q3-16, 21, 24, 27 (index)

      c. Anthropometrics:
          i. **Height for age z-score, weight for height z-score, nutritional status (based on weight for height using Malawi standards)**
          ii. Stunting, wasting (binary, less than 2 std z-score)

      d. Educational Testing:
          i. **MDAT scores**
          ii. **SDQ scores on behavior towards others.** Use the following to score: http://www.sdqinfo.com/c3.html

Detailed Description of Construction of Competencies Index:

Moderator variable for fertilizer application: S11bQ22

Fertilizer (Q23-26):

Quantity index: time taken to complete (Q23), categorize it as 1 'below median' (in seconds); 2 'above median'; and 3 'did not complete/did not complete in time'. Median is calculated among those who completed under the allocated time.
Quality index: Each Q (24-26) coded as 1 if Yes 0 if No and then added up to create an index between 0-3 of the quality of the application of fertilizer.
Normalize each by subtracting the control mean and dividing by the control SD.

Making change (Q27-28):

Same as above: (quantity index, Q27) and quality index (Q28).
Use the same procedure for Q29-30, Q31-32. Then, add the quality indices (Q28, 30, and 32).
Add quantity indices (Q27, 29, and 31).
Normalize each by subtracting the control mean and dividing by the control SD.

Sending a text message (Q35-37) – moderator variables to be used for adjustment (Q33-34):

Same as above, then normalize each index.

Use the calculator on mobile phone (Q38-39):

Same as above, then normalize each index.

Calculate profits from trade (Q-40-42):

Same as above, then normalize each index.

Finally, average the normalized quantity indices and the quality indices separately to produce two final competency indices.

Appendix B: Detailed description of construction of outcome variables

This appendix provides additional detail on the full set of outcomes presented in the tables in the main text, as well as those in this supplementary online appendix. Additional detail can also be found in the pre-analysis plan described in Appendix A and found at https://www.socialscienceregistry.org/trials/36 or in the survey instruments which can be found at https://sites.google.com/site/decrgberkozler/datasets.

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 – as well as outcomes in these domains for their husbands and children. We briefly summarize these in the text below, and then provide full detail of each in Appendix Table B1.

Education and Competencies. The primary outcomes we examine for education are highest grade completed (self-reported) and the overall competencies score. Secondary outcomes include the highest qualification obtained which are separated into the Primary School Leaving Certificate (PSLC), Junior Certificate of Education (JCE) and the Malawi Secondary Certificate of Education (MSCE). We also present the components of the competencies index and the total time taken to complete these tests.

Marriage and Fertility. Our primary outcomes include self-reported data on whether or not the core respondent was ever married or ever pregnant. We also examine age at first marriage and at first birth,[2] as well as total live births. Desired fertility is a secondary outcome.

Health. Our primary health outcomes are HIV and anemia prevalence, both measured using biological data. Additional secondary outcomes include psychological wellbeing measured with the General Health Questionnaire 12 (GHQ-12), and the number of meals eaten in the last week that contained, meat, fish, or eggs.

Sexual behavior. All outcomes in the sexual behavior domain are secondary and self-reported. On the extensive margin, our sexual behavior outcomes include whether the core respondent has ever had sex, her number of lifetime sexual partners, and whether she was sexually active during the past 12 months. On the intensive margin, we look at the core respondent's age at first sex, whether she had a sexual partner five or more years older, and her condom use during her most recent sexual intercourse.

Empowerment and Aspirations. Our primary measures of empowerment include an indicator of changes in life satisfaction and a super index of overall empowerment. This super-index of empowerment includes sub-indices (all secondary outcomes) that measure self-esteem, preferences for children's education, an index of social participation, and aspirations. We also construct super indices of empowerment separately for the married and unmarried sub-samples, as well as a super index of economic control within marriage for the married sub-sample. These three indices are also primary outcomes.

Employment and Wages. In this domain, we examine the proportion of hours spent in self-employment or paid work, the typical wage rate for work done in the past three months, and the opportunity cost of time which is constructed by asking the core respondents a series of hypothetical questions regarding whether they would accept employment at a given wage rate. Secondary outcomes include whether the core respondent participated in any wage work in the

---

[2] Our pre-analysis plan suggested we would use a hazard model. We instead simply use OLS to examine age at first marriage and age at first birth in the intensive margin.

past three months, labor income in the past five seasons, and an effective daily wage rate for work done in the past seven days. Per capita consumption is reported as a secondary outcome.

Husbands. Our analysis of husbands focuses on spousal quality and their attitudes towards women's empowerment. The husband quality index, the first primary outcome, includes sub-components that measure the husband's highest grade completed and highest qualification obtained, his cognitive score on the Raven's Colored Progressive Matrices, his employment status and wage, his HIV status, his marital fidelity (self-reported), and his mental health measured through the GHQ-12. The index of spousal attitudes towards women's empowerment, also a primary outcome, includes sub-indices for attitudes towards their daughters' schooling and marriage, their wives' autonomy, domestic violence, as well as their divorce prospects and desired fertility levels. Components of these two indices are presented as secondary outcomes.

Children. The primary child outcomes fall under four domains: anthropometrics, health, parental practices and educational testing. For anthropometrics, we construct height-for-age z-score (HAZ) for living children younger than 60 months old. Our health outcomes include neonatal and post-neonatal mortality. For parental practices, we construct variables for exclusive breastfeeding in the past six months and an index of parenting practices. Finally, for educational testing we report MDAT and SDQ scores for all 36-59 month-olds.[3] The self-reported data on children come from complete birth and death histories collected at Rounds 3 and 4 from the mother or the primary caregiver.

---

[3] The pre-analysis plan also indicates that we would report impacts for an indicator for child mortality, but there are only 22 child deaths in our entire sample during the study period, so we exclude this outcome. We also exclude weight for height, as the prevalence of wasting (weight-for-height z-score <-2) is negligible in Malawi. Finally, due to space considerations, we do not show impacts on secondary child outcomes, which include birth weight, vaccinations, and whether or not the child usually sleeps under a bed-net.

Appendix Table B1: Detailed description of construction of outcome variables

Outcome	Table	Population	Primary?	Rounds	Description
Highest Grade Completed	I	Core Respondent	Yes	2,3,4	Highest grade completed is the self-reported highest grade attended/completed by the core respondent at the time of the household survey. In rounds 2 and round 3 this is the highest grade attended and in round 4 in is highest grade completed.
Passed Primary School (PSLC)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Primary School Leaving Certificate (PSLC) at the end of 8th grade.
Passed Junior Secondary School (JCE)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Junior Certificate of Education at the end of 10th grade
Passed Secondary School (MSCE)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Malawi School Certificate of Education at the end of 12th grade

Outcome	Table	Population	Primary?	Rounds	Description
TIMMS Math Score (Standardized)	I	Core Respondent	N/A	3	Total score from a professionally developed test of mathematics based on the Malawian school curricula for the grades the target population was attending. 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 4 years, were added to the Math test.
Non-TIMMS Math Score (Standardized)	I	Core Respondent	N/A	3	
Cognitive Test Score (Standardized)	I	Core Respondent	N/A	3	The cognitive test score is based on Raven's Colored Progressive Matrices.
Competencies Score (Standardized)	I	Core Respondent	Yes	4	Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The overall competencies score is a standardized index of the five competencies listed immediately below.
Fertilizer Application (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to follow instructions for applying fertilizer to maize.
Change Given (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to make change following hypothetical scenarios of market transactions.
Sending a Text Message (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to send a text message saying "hello" to a specified phone number.
Using a Calculator (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to use a cell-phone calculator to calculate 873*17.

Outcome	Table	Population	Primary?	Rounds	Description
Calculating Profits (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to calculate profits from hypothetical scenarios.
Total Time Spent on Competencies (Standardized)	S3	Core Respondent	No	4	Standardized index of average time taken across the five competencies
Ever Married	II	Core Respondent	Yes	2,3,4	=1 if core respondent ever married at the time of the household survey. Variable is corrected for inconsistences across rounds.
Age at First Marriage	II	Core Respondent	Yes	4	Age of core respondent at first marriage.
Ever Pregnant	II	Core Respondent	Yes	2,3,4	=1 if core respondent ever pregnant at time of the household survey. Variable is corrected for inconsistences across rounds.
Number of Live Births	II	Core Respondent	Yes	2,3,4	Number of live births reported by core respondent at time of the household survey.
Age of First Birth	II	Core Respondent	Yes	4	Age of core respondent at first live birth.
Desired Fertility	II	Core Respondent	No	4	Reported number of total children desired by the core respondent (including any they already have).
HIV Positive	III	Core Respondent	Yes	2,3,4	=1 if core respondent is HIV positive. Biomarker data for HIV were collected through home based voluntary counseling and testing (HCT).
Anemic	III	Core Respondent	Yes	4	=1 if core respondent is anemic (biomarker data). 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 12d/dL if non-pregnant based on WHO guidelines to define mild anemia.

11

Outcome	Table	Population	Primary?	Rounds	Description
Ever Had Sex	S4	Core Respondent	No	2,3,4	=1 if core respondent ever had sex at time of household survey. Variable is corrected for inconsistences across rounds.
Number of Sexual Partners (lifetime)	S4	Core Respondent	No	2,3,4	Number of lifetime sexual partners self-reported by core respondent.
Sexually Active During Past 12 Months	S4	Core Respondent	No	2,3,4	=1 if core respondent sexually active during 12 months prior to household survey
Age at First Sex	S5	Core Respondent	No	2,3,4	Core respondents age at first sexual activity, reported for the sub-sample that report having ever had sex at time of household survey.
Older Partner	S5	Core Respondent	No	2,3,4	=1 if core respondent reports an older partner. A core respondent is defined as having an older partner if she has had a partner who is 5 years older or more in the past 12 months. Variable is defined for the sub-sample that report having ever had sex at time of household survey.
Condom Use	S5	Core Respondent	No	2,3,4	=1 if core respondent uses a condom. 'Condom Use' is defined as using a condom at last sex with most recent sexual partner. Variable is defined for the sub-sample that report having ever had sex at time of household survey.
Psychological Distress	S6	Core Respondent	No	2,3,4	=1 if core respondent suffers from psychological distress. Psychological distress is equal to one if the summed General Health Questionnaire- 12 score is equal to three or higher, and is zero otherwise.

Outcome	Table	Population	Primary?	Rounds	Description
Number of Times Respondent Ate Protein Rich Foods During the Past 7 Days	S6	Core Respondent	No	2,3,4	Total number of times core respondent ate protein rich foods during the past 7 days. The variable takes on a value of 0-21. Protein rich foods are defined as those containing animal proteins, i.e. meat, fish, and eggs.
Opportunity Cost of Time	IV	Core Respondent	Yes	4	Opportunity cost of time is calculated by taking the minimum daily wage the respondent would take for one year of work in her village.
Typical Wage in Past Three Months	IV	Core Respondent	Yes	4	Typical wage the core respondent reports earning in the past three months. It takes on a value of zero if the core respondent earned nothing.
Proportion of Hours Spent in Self-Employment or Paid Work in Past Week	IV	Core Respondent	Yes	4	Total number of hours core respondent spent in self-employment or paid work during the past 7 days.
Effective Daily Wage (Past 7 Days)	S7	Core Respondent	No	4	Effective daily wage in the past 7 days in USD. The effective daily wage is calculated using total reported earnings and total hours worked in the past seven days.
Labor Income (Past 5 Seasons)	S7	Core Respondent	No	4	Labor income is calculated from total reported earnings over the past five seasons.
=1 if Any Wage Work in Past 3 Months	S7	Core Respondent	No	4	=1 if core respondent reports doing any wage work in the past three months (including any *ganyu*, or day labor)
Real Monthly Per Capita Household Consumption (USD)	S7	Core Respondent	No	2,3,4	Real monthly per-capita exchange rate comparable trimmed consumption aggregate using market unit prices.

Outcome	Table	Population	Primary?	Rounds	Description
Super-Index of Overall Empowerment (Standardized)	IV	Core Respondent	Yes	4	Standardized super index of overall empowerment for the core respondent. Includes the following four sub-components that are described below: index of self-efficacy, index of preferences for child education, index of social participation, and index of aspirations.
Index of Self-Esteem (Standardized)	S8	Core Respondent	No	4	Standardized index of self-esteem using the Rosenberg (1965) scale.
Index of Preferences for Child Education (Standardized)	S8	Core Respondent	No	4	Standardized index of preferences for child education. Includes nine questions regarding attitudes towards the importance of schooling for girls.
Index of Social Participation (Standardized)	S8	Core Respondent	No	4	Standardized index of social participation. Includes three questions on social participation: number of meetings attended in past year; number of times voted in past 5 years; and number of times in the past month core respondent has got together with friends for either food or drink.
Aspirations	S8	Core Respondent	No	4	The change in subjective wellbeing asks the respondent where she sees herself on a 10-step ladder comparing today to five years from now, where zero represents the worst possible life she could have and 10 represents the best possible life she could have.

Outcome	Table	Population	Primary?	Rounds	Description
Change in Subjective Wellbeing from Five Years Ago to Today	IV	Core Respondent	Yes	4	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.
Super-Index of Unmarried Empowerment (Standardized)	IV	Core Respondent	Yes	4	Standardized super index of unmarried empowerment, measured for core respondents who are not married at the time of the household survey. Includes two components described below: index of non-abuse and index of autonomy.
Index of Unmarried non-Abuse (Standardized)	-	Core Respondent	No	4	Standardized index of unmarried non-abuse, measured for core respondents who are not married at the time of the household survey. Includes question on whether the core respondent has been beaten or mistreated physically, whether the core respondent has been forced to have sex against her will, and a series of five questions on threats or physical violence. The variable is constructed so that higher values are an indication of not experiencing physical abuse.

Outcome	Table	Population	Primary?	Rounds	Description
Index of Unmarried Autonomy (Standardized)	-	Core Respondent	No	4	Standardized index of autonomy, measured for core respondents who are not married at the time of the household survey. Includes questions on whether the core respondent needs permission to do certain activities, whether the core respondent can travel alone, and whether the core respondent is allowed to have money set aside for her own use.
Super-Index of Married Empowerment (Standardized)	IV	Core Respondent	Yes	4	Standardized super index of married empowerment, measured for core respondents who are married at the time of the household survey. It includes 8 components described below: index of decision-making; index of marital satisfaction; index of women's divorce prospects; index of fertility empowerment; index of self-determination in marriage; index of frequency of social contact; index of spousal abuse; and age difference between husband and wife. All variables are constructed so that higher values are an indicator of increased empowerment.
Index of Decision-Making (Standardized)	-	Core Respondent	No	4	Standardized index on a series of questions on who makes decisions in the household related to food, clothing and children.
Index of Marital Satisfaction (Standardized)	-	Core Respondent	No	4	Standardized index on a series of questions related to how satisfied the core respondent is with her marriage.

Outcome	Table	Population	Primary?	Rounds	Description
Index of Women's Divorce Prospects (Standardized)	-	Core Respondent	No	4	Standardized index of ability of core respondent to divorce her husband and maintain household property.
Index of Fertility Empowerment (Standardized)	-	Core Respondent	No	4	Standardized index of differences in core respondent and her husband ideal degree of family planning.
Index of Self-Determination in Marriage (Standardized)	-	Core Respondent	No	4	Standardized index of core respondents need for permission from her husband to undertake certain activities.
Index of Frequency of Social Contact (Standardized)	-	Core Respondent	No	4	Standardized index of core respondents' frequency of travelling outside the community and sending and receiving phone calls and text messages.
Index of Spousal Abuse (Standardized)	-	Core Respondent	No	4	Standardized index of a series of questions both asking about when it is acceptable for the husband to beat his wife, as well as whether the husband is violent towards the core respondent.
Age Difference Between Wife and Husband	-	Core Respondent	No	4	Age difference between core respondent and her husband.
Married Index of Economic Control (Standardized)	VI	Core Respondent	Yes	4	Standardized super index of married economic control, measured for core respondents who are married at the time of the household survey. It includes 4 components described immediately below: agricultural decision-making power, microenterprise participation, livestock control, and ratio of wife to husband consumption.
Agricultural Decision-making Power	-	Core Respondent	No	4	=1 if the core respondent is involved in decision-making around any of the household plots.

Outcome	Table	Population	Primary?	Rounds	Description
Microenterprise Participation	-	Core Respondent	No	4	=1 if the core respondent controls the use of profits from any household microenterprise.
Livestock Control	-	Core Respondent	No	4	=1 if the core respondent is involved in decision-making around any of the household's livestock.
Ratio of Wife to Husband Consumption (Standardized)	-	Core Respondent	No	4	Standardized index of the ratio of wife's to husband's consumption focusing on gender specific goods.
Husband Quality Super Index (Standardized)	V	Husband	Yes	4	Standardized quality super index for husbands of core respondents. Includes nine components listed below: highest grade completed, passed PSLC, passed JCE, passed MSCE, cognitive test score, typical wage in the past three months, employment status, sexual activity and marital fidelity, and physiological distress.
Highest Grade Completed	V	Husband	No	4	Highest grade completed is the self-reported highest completed by the husband at the time of the household survey.
=1 if Passed Primary School (PSLC)	V	Husband	No	4	=1 if husband passed the Primary School Leaving Certificate (PSLC) at the end of 8th grade.
=1 if Passed Junior Secondary School (JCE)	V	Husband	No	4	=1 if husband passed the Junior Certificate of Education at the end of 10th grade
=1 if Passed Secondary School (MSCE)	V	Husband	No	4	=1 if husband passed the Malawi School Certificate of Education at the end of 12th grade
Cognitive Test Score (Standardized)	V	Husband	No	4	The cognitive test score is based on Raven's Colored Progressive Matrices.

Outcome	Table	Population	Primary?	Rounds	Description
Typical Wage in Past Three Months	V	Husband	No	4	Typical wage the husband reports earning in the past three months. It takes on a value of zero if the husband earned nothing.
=1 if Currently Employed	V	Husband	No	4	=1 if husband currently employed.
Sexual Activity and Marital Fidelity (Standardized)	V	Husband	No	4	The husband's sexual activity and marital fidelity standardized 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. Constructed so that higher values are better.
=1 if Does Not Suffer from Psychological Distress	V	Husband	No	4	=1 if husband does not suffer from psychological distress. Psychological distress is equal to one if the summed General Health Questionnaire- 12 score is equal to three or higher, and is zero otherwise.
=1 if HIV Positive	V	Husband	No	4	=1 if husband is HIV positive. Biomarker data for HIV was collected through home based voluntary counselling and testing (HCT).
Attitudes Towards Women's Empowerment Super Index (Standardized)	S9	Husband	Yes	4	Standardized super index of five indicators listed below: preferences for children's education, attitudes towards wife's autonomy, attitudes towards abuse, divorce prospects, and desired fertility.

Outcome	Table	Population	Primary?	Rounds	Description
Index of Preferences for Children's Education (Standardized)		Husband			Standardized index of preferences for child education. Includes nine questions regarding attitudes towards the importance of schooling for girls.
	S9		No	4	
Index of Attitudes Towards Wife's Autonomy (Standardized)	S9	Husband	No	4	Standardized index of the husband's attitude towards his wife's autonomy. Series of questions asking whether the wife needs permission to engage in day to day activities. This variable takes on a higher value if the wife does not need permission.
Index of Attitudes Towards (non)-Abuse (Standardized)	S9	Husband	No	4	Standardized index of husband's attitudes towards beating his wife. This variable uses a series of three questions asking when wife beating is justified. This variable takes on a higher value if the husband does not think beating is justified.
Index of Wife's Divorce Prospects (Standardized)	S9	Husband	No	4	Standardized index of wife's divorce prospects. This looks at a series of variables related to a man's ability to leave the marriage and keep things from the household. This variable takes on a higher value the less ability a man has to divorce his wife.
Desired Fertility	S9	Husband	No	4	Husband's desired fertility. Note that this variable is standardized and made negative when added to the super index.

Outcome	Table	Population	Primary?	Rounds	Description
Height-for-Age Z-Score	VI-VIII	Child	Yes	4	The height-for-age (length-for-age) z-score is calculated using the 2006 WHO child growth standards. See the following for more details: Leroy, Jef L (2011). zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards. http://www.ifpri.org/staffprofile/jef-leroy.
Neonatal Mortality	VI	Child	Yes	3,4	=1 if child died at or before 31 days.
Post-Neonatal Mortality	VI	Child	Yes	3,4	=1 if child died between the ages of one month and one year. Defined for those that survived the first month of life.
Parenting Practices Percentage Score	VI	Child	Yes	4	Percentage score across a series of 16 questions on parenting practices related to addressing behavior problems to interacting with the child (the total number of questions asked various by the age of the child). These questions are only asked of living children.
Exclusively Breastfed for First Six Months	VI	Child	Yes	3,4	=1 if mother exclusively breastfed for the first 6 months of life, is still exclusively breastfeeding for those under 6 months, or who breastfed until death.
Standardized Malawi Development Assessment Tool	VI	Child	Yes	4	The standardized Malawi Developmental Assessment Tool (MDAT) is a test of fine motor skills, language, and hearing administered directly to the child.

Outcome	Table	Population	Primary?	Rounds	Description
Reported Child Difficulties	VI	Child	Yes	4	The standardized reported child difficulties and reported pro-social behaviors are created using the Strengths and Difficulties Questionnaire (SDQ) and administered to the parent. Details on construction of these variables can be found at http://www.sdqinfo.com/c3.html.
Reported Pro-Social Behaviors	VI	Child	Yes	4	

Appendix C: Estimation of Treatment Effects on Children

This appendix provides an overview of the empirical issues involved in estimating treatment effects on child outcomes when the intervention under investigation targets prospective mothers and starts prior to their pregnancies. As the intervention may have altered the composition of children subsequently observed, we suggest a simple sequence of assumptions and steps in an attempt to move from the total reduced-form effect of the intervention to a more standard causal effect on the children actually born.

The natural experimental literature has recognized that maternal selection and differential mortality represent plausible causal pathways when analyzing child outcomes. Using data from the US, Buckles and Hungerman (2013) show that nearly half of the well-documented effects of season of birth on later-life outcomes can be explained by variation in the types of women giving birth across seasons. Aaronson, Lange and Mazumder (2014) show extensive margin selection also contributes to the quantity-quality tradeoff for children. This phenomenon is well documented in the developed world, and yet Currie and Vogl (2012) suggest that these extensive margin effects are likely to be more pronounced in the developing world where differential mortality, as well as differential fertility, is an operative channel. It has now become standard in the natural experiments literature to test for the presence of selection effects when studying child outcomes (see, for example, Almond 2006; Black et al. 2014; Adhvaryu et al. 2016). In many cases in this literature, the obvious selection effect suggests that the simple casual effects are lower bounds, but we present a context in which selection and treatment effects go in the same direction. It is therefore critical to attempt to control for the selection mechanism as a mediator, in order to isolate how much of the observed effect may have a simple causal interpretation.

To use counterfactual outcomes to represent impacts on subsequent children, we must define the universe as being the *potential children*: all those who might exist under either the treated or control state.[4] Outcome measured for potential child i at time t is Y_{it}. This outcome is only observed if the child is born and survives, a binary outcome denoted by $S_{it} = 1$. The probability that woman i has a surviving child at time t, as a function of her baseline characteristics, can be written as $S_{it} = S(X_i)$, where X_i is a set of pre-treatment maternal characteristics. For children who are born, we observe

$$Y_{it} = Y_{it}(x_{it}, a_{it}; X_i, A_{it}, Z_{it} \mid S_{it} = 1)$$

where A_{it} is the age of the mother at the time at which child data is collected, x_{it} are child-level determinants of the outcome, a_{it} is child age, and Z_{it} gives attributes of the father.

Critically, in a study tracking potential mothers, X_i is observed for all respondents (regardless of whether or not they had children) and hence can be used to predict fertility within the universe of these potential mothers. Controlling for selection into motherhood thus resembles standard attrition adjustment, and requires a (weaker) selection on observables assumption. Controlling for factors that are observed only among extant children, however, must be done on the intensive margin, and so resembles mediation analysis as in Baron and Kenny (1986), which

---

[4] This study tracks female respondents and their descendants, so it is concerned with the potential children of a fixed set of women. We do not attempt to capture all potential children born to the fathers in this study, and to do so would have needed to specify a group of males at baseline and tracked them. The control for father characteristics in this context therefore sits behind a layer of female selection, and so the assumptions needed to control for father type are stronger than those needed to control for mother type.

requires substantially stronger assumptions. Specifically, we must now assume that there is a globally correct functional form across both treatment and control, so that inclusion of the mechanism controls does not open a 'backdoor path' between the mediator and some other, unobserved determinant of outcomes (for more discussion of these assumptions, see Sobel 2008; Flores and Flores-Lagunes 2009; Bullock, Green and Ha 2010; Imai, Tingley and Yamamoto 2013; Huber 2014; and Heckman and Pinto 2015).

Given these strong assumptions, we now walk through a set of distinct treatment effects that we may wish to estimate, each of which has a different causal interpretation. We start with the simple reduced-form impact of the treatment on child outcomes and proceed to add successively stronger controls until we have isolated the *ceteris* paribus treatment effect that would have been observed had an experiment been conducted on the sample of children actually observed, rather than their mothers. Appendix Figure C1 presents a conceptual framework.

Total effect:

$$(1) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid S_{it} = 1\right).$$

This is the simple difference in outcomes between the children of those exposed to the treatment versus the control.

Correcting for Maternal Type Selection:

We can begin to control for the extensive margin effects of the treatment by modeling the probability that a child is born to a mother i during epoch t as: $\Pr(S_{it} = 1) = \Phi(X_i, T_i) + \varepsilon_{it}$. This problem is exactly analogous to attrition, and so we can exploit the familiar toolkit to test and correct for it (Hanson 1978, Rosenbaum and Rubin 1983). Observational correction can be conducted using a probit model that regresses a binary indicator for giving birth during an epoch on a rich set of baseline covariates, a treatment indicator, and the interaction between treatment and the covariates. We can use this regression to predict the probability of birth for all core respondents by epoch, and weight the subsequent analysis by the inverse of this probability. This is the application of standard attrition-based inverse probability weighting to the fertility problem. The required assumption is that there be no unobserved determinants of fertility that are correlated with the treatment or the treatment*covariate interactions. Regressions weighted by $1 \big/ \hat{\Pr}(S_{it} = 1)$, subject to this assumption, are now representative of the entire original sample of core respondents and hence not subject to selection effects arising from the decision to give birth or not. We can also use OLS to control for the same set of maternal baseline characteristics X_i to provide estimates that are "doubly robust" to the extensive margin selection controls (Robins and Rotnitzky 2001; Van der Laan and Robins 2003; Bang and Robins 2005). This then provides an estimate of the impact on children if the composition of women who gave birth was identical in treatment and control in each epoch:

$$(2) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid X = \bar{X}, \Pr(S = 1 \mid T) = \bar{S}\right)$$

Correcting for Paternal Type Selection:

The next selection margin is father type. In our data structure we do not observe the attributes of the universe of potential fathers, and rather have data on the fathers of children actually born. We therefore must control for paternal characteristics on the intensive margin rather than using the selection IPW approach that we use for mothers. The assumptions underlying these paternal controls are therefore the strong assumptions of mediation analysis, rather than those of attrition propensity weighting. Subject to these assumptions, the inclusion of paternal covariates gives us the expected treatment-control difference holding both maternal and paternal types constant across the treatment and control groups.

$$(3) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid X = \bar{X}, Z(T) = \bar{Z}, \Pr(S = 1 \mid T) = \bar{S}\right)$$

Correcting for Child Age:

Differences in the composition of child age across treatment and control can lead to large differences that are, in fact, completely trivial. If, for example, the treatment led to a delay in births, then the treatment children will be younger on average and hence may perform more poorly on a broad range of tasks than the control children.[5] We can recover a meaningful treatment effect by comparing children in treatment and control *at the same age*. We achieve this by including linear, quadratic, and cubic controls for child age in months in our regressions:

$$(4) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid a(T) = \bar{a}, X = \bar{X}, Z(T) = \bar{Z}, \Pr(S = 1 \mid T) = \bar{S}\right)$$

Direct Treatment Effect:

Maternal age represents a potentially important mechanism for improvements in child outcomes for the same set of mothers, even though it represents an extensive margin effect in that changes in maternal age must necessarily lead to a different set of children being born. It is possible, for example, that an intervention that, all else equal, simply delayed fertility from age 13 to age 18 would lead to improved child outcomes due to increased gynecological maturity. Effects driven only by changes in age therefore have a meaningful causal interpretation that can be seen as *ceteris paribus* for mothers even though it operates on the extensive margin for children. To control for age as a mediating variable, we can include age and $A*X$ interactions as covariates.

Subject to strong assumptions of (i) selection on observables in the fertility equation, and (ii) correct functional form and common support in the Barron-Kenny controls for the mechanisms, the resulting adjusted difference provides the 'direct' treatment effect of the program on a sample of children made homogeneous across treatment and control by reweighting and regression adjustments. The result is a suggestive answer to an obvious policy question: "Does the intervention confer a protective effect on a given child?"

---

[5] Conversely, as shown in Appendix Figure C2, the mean height-for-age z-score in our control group starts out very close to the mean of the reference group at birth, but declines steadily and rapidly as children get older, ending up almost two standard deviations below the global distribution by the time they are 36 months old. This seems to be a common pattern in poor countries (see, for example, Figure 1 in Barham, Macours and Maluccio 2013). Hence, comparing a younger cohort of children in the treatment group to an older cohort in control would spuriously show lower stunting in the treatment group in the absence of any meaningful effects on height.

$$(5) \qquad E\left(Y_{it}^{1} - Y_{it}^{0} \mid a(T) = \bar{a}, Z(T) = \bar{Z}, X = \bar{X}, A(T) = \bar{A}, \Pr(S = 1 \mid T) = \bar{S}\right)$$

Figure C1: Conceptual Framework for Causal Pathways to Child Treatment Effects

T_i :   Treatment

Mother:

$S_{it}(X,T) = 1$ Birth Outcome: Is a child born in this interval?

Maternal Selection effects that alter the identity of the woman who gives birth

$A_{it}^{C}(T)$: Birth Timing: Child Age   **Father**

$A_{it}^{M}(T)$: Birth Timing: Maternal Age $Z_i(T)$: Father Selection

Child Selection effects that alter the identity of the child who is born

Child

D1. Direct impacts

Direct Treatment effect on the child actually born

Figure C2. Height-for-age z-score by age in months (control group)
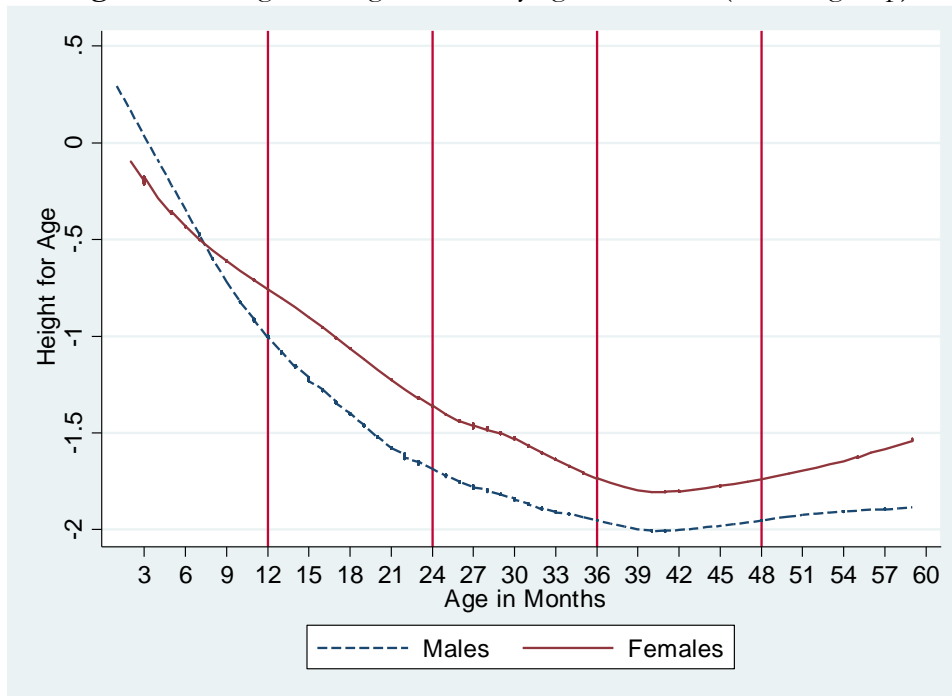
References (for the Appendix)

Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder. 2014. Fertility transitions along the extensive and intensive margins. *Am Econ Rev* 104 (11): 3701-3724.

Adhvaryu, Achyuta, Prashant Bharadwaj, James Fenske, Anant Nyshadham, and Richard Stanley. 2016. Dust and Death: Evidence from the West African Harmattan. *CSAE Working Paper WPS/2016-03.* Centre for the Study of African Economies, University of Oxford.

Almond, D. 2006. Is the 1918 influenza pandemic over? Long-term effects of *In Utero* Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy* 114 (4): 672-712.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. Cash or Condition? Evidence from a Cash Transfer Experiment. *Q J Econ* 126 (4): 1709-1753.

Baird, Sarah J, Richard S Garfein, Craig T McIntosh, and Berk Özler. 2012. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *The Lancet* 379 (9823): 1320-1329.

Baird, Sarah, Jacobus De Hoop, and Berk Özler. 2013. Income shocks and adolescent mental health. *Journal of Human Resources* 48 (2): 370-403.

Bang, Heejung and James M Robins. 2005. Doubly robust estimation in missing data and causal inference models. *Biometrics* 61 (4): 962-973.

Barham, Tania, Karen Macours, and John A Maluccio. 2013. Boys' Cognitive Skill Formation and Physical Growth: Long-Term Experimental Evidence on Critical Ages for Early Childhood Interventions. *Am Econ Rev* 103 (3): 467-471.

Baron, Reuben M and David A Kenny. 1986. The moderator--mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173.

Black, Sandra E, Aline Bütikofer, P Devereux, and K Salvanes. 2014. This Is Only a Test? Long-Run and Intergenerational Impacts of Prenatal Exposure to Radioactive Fallout. *Unpublished manuscript.*

Buckles, Kasey S and Daniel M Hungerman. 2013. Season of birth and later outcomes: Old questions, new answers. *Review of Economics and Statistics* 95 (3): 711-724.

Bullock, John G, Donald P Green, and Shang E Ha. 2010. Yes, but whats the mechanism?(dont expect an easy answer). *Journal of Personality and Social Psychology* 98 (4): 550.

Currie, Janet and Tom Vogl. 2012. Early-life health and adult circumstance in developing countries. *NBER Working Paper 18371.* National Bureau of Economic Research. National Bureau of Economic Research.

Flores, Carlos A. and Alfonso Flores-Lagunes. 2009. Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness. *IZA Discussion Paper No. 4237.*

Hanson, Robert Harold. 1978. The current population survey: design and methodology

Heckman, James J and Rodrigo Pinto. 2015. Econometric mediation analyses: Identifying the sources of treatment effects from experimentally estimated production technologies with unmeasured and mismeasured inputs. *Econometric Reviews* 34 (1-2): 6-31.

Huber, Martin. 2014. Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics* 29 (6): 920-943.

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. Experimental designs for identifying causal mechanisms. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (1): 5-51.

Robins, James M and Andrea Rotnitzky. 2001. COMMENTS. *Statistica Sinica* 920-936.

Rosenbaum, Paul R and Donald B Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70 (1): 41-55.

Sobel, Michael E. 2008. Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics* 33 (2): 230-251.

Van der Laan, Mark J and James M Robins. 2003. Unified methods for censored longitudinal data and causality. Springer Science & Business Media.

Table S1: Baseline means and balance

	Baseline Dropout		Baseline Schoolgirl			p-value
	Mean (s.d.)		Mean (s.d.)			p-value
	Control group	Conditional group	Control group	Conditional group	Unconditional Group	(CCT-UCT)
	(1)	(2)	(3)	(4)	(5)	(6)
Urban Household	0.181	0.129	0.346	0.478	0.418	0.726
	(0.385)	(0.335)	(0.476)	(0.500)	(0.494)	
Mother Alive	0.783	0.749	0.839	0.800	0.828	0.431
	(0.413)	(0.434)	(0.368)	(0.401)	(0.378)	
Father Alive	0.656	0.649	0.709	0.718	0.76	0.341
	(0.476)	(0.478)	(0.454)	(0.451)	(0.428)	
Household Size	6.120	6.104	6.375	6.341	6.659	0.156
	(2.388)	(2.617)	(2.262)	(2.134)	(2.063)	
Asset Index	-0.831	-0.743	0.632	1.100	1.373*	0.572
	(2.233)	(2.484)	(2.575)	(2.721)	(2.444)	
Age	17.579	17.162	15.228	14.919	15.466	0.002
	(2.397)	(2.478)	(1.904)	(1.828)	(1.926)	
Highest Grade Attended	6.105	5.940	7.506	7.262	7.928**	0.004
	(2.856)	(2.864)	(1.651)	(1.601)	(1.587)	
Never Had Sex	0.315	0.294	0.800	0.807	0.790	0.682
	(0.465)	(0.456)	(0.400)	(0.395)	(0.408)	
Ever Pregnant	0.445	0.420	0.021	0.029	0.029	0.964
	(0.498)	(0.494)	(0.144)	(0.169)	(0.168)	
Chi-squared joint test of orthogonality (p-value)		0.168		0.122	0.121	0.032

Notes: Mean differences statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence. Stars on the coefficients in columns (2) indicate significantly different than the control group for baseline dropouts.Stars on the coefficients in column (4) and (5) indicate significantly different than the control group for baseline schoolgirls. Means are weighted to make them representative of the target population in the study EAs.

Table S2: Attrition

	Baseline Dropout		Baseline Schoolgirl	
	=1 if Completed Household Survey Round 4			
	(1)	(2)	(3)	(4)
=1 if Conditional	-0.007	-0.008	0.055***	0.056***
	(0.031)	(0.029)	(0.019)	(0.018)
=1 if Unconditional			0.058***	0.061***
			(0.023)	(0.021)
p-value UCT vs. CCT	-	-	0.896	0.825
p-value Treatment	0.828	0.774	0.004	0.002
Baseline controls interacted with treatment?	NO	YES	NO	YES
p-value on joint F-test for interactions CCT	-	0.009	-	0.332
p-value on joint F-test for interactions UCT	-	-	-	0.101
p-value UCT interactions vs. CCT interactions	-	-	-	0.690
Mean in Control Group	0.843	0.843	0.875	0.875
Number of observations	885	885	2,273	2,273

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. All regressions include baseline centered values of the following variables: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex. Columns (2) and (4) interact the centered baseline controls with treatment. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S3: Program impacts on educational qualifications and competencies (beneficiaries)

Panel A: Baseline Dropouts

	Educational Qualifications									Competencies (Standardized)					
	=1 if Passed Primary School (PSLC)			=1 if Passed Junior Secondary School (JCE)			=1 if Passed Secondary School (MSCE)			Fertilizer Application	Change Given	Sending a Text Message	Using a Calculator	Calculating Profits	Total Time Spent
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	Two Years After Program					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
=1 if Conditional Schoolgirl	0.030	0.058**	0.081***	0.012	0.049**	0.034	0.004	0.003	0.016	-0.044	-0.014	0.101	0.065	0.094	-0.007
	(0.025)	(0.025)	(0.026)	(0.019)	(0.021)	(0.022)	(0.008)	(0.010)	(0.011)	(0.069)	(0.062)	(0.072)	(0.071)	(0.076)	(0.091)
Mean in Control Group	0.328	0.351	0.366	0.085	0.123	0.136	0.008	0.025	0.026	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	697	718	744	697	718	744	697	718	744	742	741	741	741	742	742

Panel B: Baseline Schoolgirls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
=1 if Conditional Schoolgirl	0.030	0.013	-0.014	-0.013	0.055*	0.033	-0.004*	0.005	0.006	0.015	0.048	0.077	0.060	-0.006	-0.113
	(0.039)	(0.024)	(0.019)	(0.022)	(0.028)	(0.028)	(0.002)	(0.011)	(0.021)	(0.071)	(0.071)	(0.070)	(0.054)	(0.076)	(0.085)
=1 if Unconditional Schoolgirl	0.046	0.030	0.017	0.002	0.016	0.010	-0.006*	-0.009	-0.065**	0.096	-0.017	0.161**	0.098	-0.045	-0.118
	(0.038)	(0.026)	(0.016)	(0.022)	(0.045)	(0.035)	(0.003)	(0.015)	(0.027)	(0.092)	(0.057)	(0.079)	(0.064)	(0.090)	(0.085)
p-value UCT vs. CCT	0.755	0.600	0.166	0.546	0.439	0.565	0.325	0.385	0.022	0.378	0.389	0.364	0.584	0.636	0.963
p-value Treatment	0.386	0.488	0.359	0.797	0.148	0.486	0.150	0.683	0.045	0.570	0.685	0.105	0.249	0.862	0.258
Mean in Control Group	0.496	0.776	0.879	0.144	0.337	0.537	0.004	0.054	0.170	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	1,967	2,019	2,047	1,967	2,019	2,047	1,967	2,019	2,047	2,048	2,046	2,047	2,047	2,048	2,048

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. 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). 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 competency components are standardized to have a mean of zero and a standard deviation of one in the control group. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S4: Program impacts on sexual behavior (beneficiaries: **extensive** margin)

Panel A: Baseline Dropouts

	=1 if Ever Had Sex			Number of Sexual Partners (lifetime)			=1 if Sexually Active During the Past 12 Months		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
=1 if Conditional Schoolgirl	-0.036*	-0.034	-0.004	0.004	-0.118	-0.023	-0.123***	-0.094**	-0.046
	(0.020)	(0.021)	(0.010)	(0.153)	(0.153)	(0.095)	(0.035)	(0.037)	(0.028)
Mean in Control Group	0.814	0.918	0.976	1.395	1.734	2.063	0.503	0.674	0.830
Sample Size	698	718	744	698	718	744	697	718	744

Panel B: Baseline Schoolgirls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
=1 if Conditional Schoolgirl	-0.009	-0.003	0.005	-0.023	0.005	0.005	-0.009	0.001	-0.030
	(0.017)	(0.024)	(0.035)	(0.040)	(0.048)	(0.061)	(0.023)	(0.029)	(0.035)
=1 if Unconditional Schoolgirl	-0.022	0.003	0.041	-0.044	-0.007	0.108	-0.021	-0.036	0.037
	(0.021)	(0.030)	(0.036)	(0.049)	(0.036)	(0.066)	(0.030)	(0.032)	(0.044)
p-value UCT vs. CCT	0.581	0.864	0.414	0.699	0.815	0.142	0.728	0.327	0.177
p-value Treatment	0.551	0.984	0.519	0.627	0.969	0.218	0.768	0.514	0.395
Mean in Control Group	0.303	0.455	0.701	0.335	0.559	1.045	0.175	0.308	0.563
Sample Size	1,965	2,016	2,048	1,964	2,016	2,047	1,965	2,015	2,048

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. We correct for inconsistencies in 'ever had sex' across rounds. 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 S5: Program impacts on sexual behavior (beneficiaries: **intensive** margin)

<u>Panel A: Baseline Dropouts</u>

	Age at First Sex			=1 if Older Partner			=1 if Use a Condom	
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
=1 if Conditional Schoolgirl	-0.064	-0.061	0.110	0.018	-0.005	0.015	0.046	0.030
	(0.137)	(0.144)	(0.133)	(0.054)	(0.045)	(0.037)	(0.037)	(0.030)
Mean in Control Group	16.250	16.578	16.790	0.230	0.300	0.309	0.159	0.156
Sample Size	525	625	723	303	427	578	446	600
<u>**Panel B: Baseline Schoolgirls**</u>								
=1 if Conditional Schoolgirl	0.220	0.136	0.147	-0.074	-0.006	-0.041	-0.006	0.015
	(0.146)	(0.130)	(0.146)	(0.050)	(0.044)	(0.038)	(0.055)	(0.041)
=1 if Unconditional Schoolgirl	-0.152	-0.039	-0.207	0.022	-0.081	0.018	0.102	0.057
	(0.179)	(0.189)	(0.127)	(0.103)	(0.057)	(0.049)	(0.086)	(0.048)
p-value UCT vs. CCT	0.064	0.404	0.052	0.351	0.258	0.248	0.268	0.482
p-value Treatment	0.143	0.536	0.128	0.291	0.367	0.422	0.483	0.479
Mean in Control Group	15.731	16.393	17.199	0.193	0.274	0.304	0.247	0.268
Sample Size	522	893	1,494	376	661	1,162	672	1,183

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. 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). We correct for inconsistencies in 'ever had sex' across rounds. 'Age at First Sex' is defined for those that had ever had sex. 'Older Partner' is defined as having a partner who is 5 years older or more in the past 12 months. 'Condom Use' is defined as using a condom at last sex with most recent sexual partner. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S6: Program impacts on health and nutrition (beneficiaries)

Panel A: Baseline Dropouts

	=1 if Suffers from Psychological Distress			Number of Times Respondent Ate Protein Rich Foods During the Past 7 Days (out of 21)		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	-0.002	0.010	0.038	0.326	0.224	0.228
	(0.039)	(0.036)	(0.042)	(0.202)	(0.192)	(0.181)
Mean in Control Group	0.463	0.314	0.424	3.678	3.989	3.741
Sample Size	698	715	743	698	718	744

Panel B: Baseline Schoolgirls

=1 if Conditional Schoolgirl	-0.068**	-0.037	-0.030	0.385**	0.596***	0.072
	(0.032)	(0.047)	(0.032)	(0.195)	(0.174)	(0.141)
=1 if Unconditional Schoolgirl	-0.139***	-0.026	-0.002	0.445**	0.338**	-0.043
	(0.035)	(0.054)	(0.046)	(0.199)	(0.153)	(0.240)
p-value UCT vs. CCT	0.068	0.860	0.552	0.814	0.215	0.672
p-value Treatment	0.000	0.677	0.627	0.023	0.001	0.858
Mean in Control Group	0.372	0.313	0.369	3.967	4.052	4.134
Sample Size	1,963	2,013	2,045	1,967	2,018	2,047

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. Psychological distress is equal to one if the summed GHQ- 12 score is equal to three or higher, and is zero otherwise. Protein rich foods are defined as those containing animal proteins, i.e. meat, fish, and eggs. The number of days each item was consumed over the past week are summed to create the outcome variable. 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 S7: Program impacts on labor market outcomes and consumption (beneficiaries: **secondary** outcomes)

Panel A: Baseline Dropouts

	Effective Daily Wage (Past 7 Days)	Labor Income (Past 5 Seasons)	=1 if Any Wage Work in Past 3 Months	Real Total Household Monthly Consumption (USD)		
	Two Years After Program			During Program	End of Program	Two Years After Program
=1 if Conditional Schoolgirl	-0.228	4.129	-0.020	-0.257	-1.941*	0.535
	(0.148)	(8.620)	(0.037)	(1.029)	(1.113)	(1.130)
Mean in Control Group	0.753	52.840	0.366	17.502	20.860	17.977
Sample Size	263	744	744	712	719	737
Panel B: Baseline Schoolgirls						
=1 if Conditional Schoolgirl	0.121	7.476	-0.010	3.192**	3.223**	2.804*
	(0.424)	(7.466)	(0.030)	(1.261)	(1.364)	(1.432)
=1 if Unconditional Schoolgirl	-0.549*	10.688	0.001	-0.586	-0.880	-0.817
	(0.285)	(12.721)	(0.055)	(1.441)	(1.524)	(1.876)
p-value UCT vs. CCT	0.278	0.829	0.838	0.032	0.034	0.127
p-value Treatment	0.121	0.420	0.939	0.030	0.044	0.137
Mean in Control Group	0.902	33.302	0.250	18.638	23.342	20.774
Sample Size	465	2,049	2,049	2,006	2,021	2,040

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. Effective daily wage is calculated using earnings and activities in the past seven days. 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 S8: Program impacts on empowerment (beneficiaries: **secondary** outcomes)

Panel A: Baseline Dropouts

	Empowerment			
	Index of Self-Efficacy (Standardized)	Index of Preferences for Child Education (Standardized)	Index of Social Participation (Standardized)	Aspirations
	(1)	(2)	(3)	(4)
=1 if Conditional Schoolgirl	-0.041	-0.020	-0.052	-0.221
	(0.076)	(0.079)	(0.068)	(0.225)
Mean in Control Group	0.000	0.000	0.000	3.267
Sample Size	744	744	744	744
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	0.059	-0.004	-0.026	0.235
	(0.079)	(0.076)	(0.068)	(0.228)
=1 if Unconditional Schoolgirl	-0.149	-0.106	-0.095	0.004
	(0.100)	(0.087)	(0.069)	(0.207)
p-value UCT vs. CCT	0.061	0.343	0.424	0.379
p-value Treatment	0.170	0.477	0.393	0.566
Mean in Control Group	0.000	0.000	0.000	3.352
Sample Size	2,049	2,049	2,049	2,049

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. Detail on the construction of the indices can be found in Appendix A and Appendix B. Aspirations asks the respondent where she sees herself on a 10-step ladder comparing today to five years from now, 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% (***), 95% (**), and 90% (*) confidence.

Table S9: Attrition Analysis for Primary Round 4 Outcomes (Baseline Dropouts)

| Outcomes | Dependent Variable: =1 if Conditional Schoolgirl | | | |
	Original	IPW	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
Highest Grade Completed	0.621***	0.606***	0.616***	0.632***
	(0.125)	(0.128)	(0.125)	(0.125)
Competencies Score (Standardized)	0.064	0.065	0.061	0.089
	(0.057)	(0.057)	(0.057)	(0.056)
=1 if Ever Married	-0.107***	-0.108***	-0.112***	-0.106***
	(0.032)	(0.033)	(0.032)	(0.032)
Age at First Marriage	0.431***	0.434***	0.418***	0.471***
	(0.155)	(0.154)	(0.156)	(0.153)
=1 if Ever Pregnant	-0.040*	-0.041**	-0.044**	-0.040*
	(0.021)	(0.021)	(0.021)	(0.021)
Number of Live Births	-0.147***	-0.145***	-0.153***	-0.127**
	(0.054)	(0.055)	(0.053)	(0.052)
Age at First Birth	0.272*	0.328**	0.251	0.329**
	(0.164)	(0.163)	(0.165)	(0.163)
= if HIV Positive	0.012	0.007	0.011	0.019
	(0.026)	(0.026)	(0.026)	(0.026)
=1 if Anemic	0.039	0.041	0.038	0.045
	(0.035)	(0.035)	(0.035)	(0.035)
Opportunity Cost of Time	-0.037	-0.041	-0.038	0.065
	(0.079)	(0.079)	(0.079)	(0.052)
Typical Daily Wage in Last Three Months	-0.140**	-0.132*	-0.141**	-0.074
	(0.068)	(0.069)	(0.068)	(0.055)
Proportion of Hours Spent in Self-Employment or Paid Work in Past	-0.011	-0.012	-0.011	-0.007
	(0.009)	(0.009)	(0.009)	(0.008)
Super Index of Overall Empowerment (Standardized)	-0.083	-0.073	-0.097	-0.065
	(0.074)	(0.075)	(0.074)	(0.072)
Change in Subjective Wellbeing from Five Years Ago to Today	-0.032	-0.050	-0.091	0.048
	(0.232)	(0.235)	(0.232)	(0.226)

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. 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). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S10: Attrition Analysis for Primary Round 4 Outcomes
(Baseline Schoolgirls: CCT vs. Control/UCT vs. Control)

Outcomes	Dependent Variable: =1 if Conditional Schoolgirl				Dependent Variable: =1 if Unconditional Scoolgirl			
	Original	IPW	Lower Bound	Upper Bound	Original	IPW	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Highest Grade Completed	0.120	0.122	0.208***	0.041	0.095	0.095	0.210*	0.052
	(0.080)	(0.080)	(0.075)	(0.084)	(0.129)	(0.129)	(0.123)	(0.132)
Competencies Score (Standardized)	0.065	0.066	0.129**	-0.045	0.098	0.101	0.176***	0.028
	(0.058)	(0.058)	(0.060)	(0.050)	(0.067)	(0.068)	(0.063)	(0.067)
=1 if Ever Married	-0.035	-0.035	0.208***	-0.065**	-0.010	-0.010	0.210*	-0.049
	(0.027)	(0.028)	(0.075)	(0.027)	(0.046)	(0.048)	(0.123)	(0.045)
Age at First Marriage	-0.011	-0.006	0.086	-0.136	0.486**	0.468**	0.650***	0.306
	(0.148)	(0.148)	(0.154)	(0.137)	(0.200)	(0.202)	(0.202)	(0.190)
=1 if Ever Pregnant	-0.024	-0.025	-0.001	-0.046	-0.001	0.000	0.027	-0.034
	(0.034)	(0.034)	(0.033)	(0.035)	(0.042)	(0.043)	(0.039)	(0.042)
Number of Live Births	0.020	0.020	0.044	-0.027	-0.024	-0.022	0.002	-0.094*
	(0.036)	(0.037)	(0.036)	(0.035)	(0.046)	(0.048)	(0.046)	(0.049)
Age at First Birth	-0.144	-0.144	0.028	-0.283**	0.001	0.002	0.123	-0.131
	(0.136)	(0.136)	(0.119)	(0.132)	(0.168)	(0.173)	(0.179)	(0.146)
= if HIV Positive	-0.001	-0.002	0.002	-0.057***	-0.002	-0.002	0.001	-0.055***
	(0.019)	(0.020)	(0.020)	(0.008)	(0.023)	(0.023)	(0.025)	(0.008)
=1 if Anemic	0.012	0.011	0.029	-0.031	-0.065*	-0.067**	-0.054	-0.105***
	(0.031)	(0.031)	(0.031)	(0.033)	(0.033)	(0.033)	(0.035)	(0.031)
Opportunity Cost of Time	-0.051	-0.049	-0.015	-0.343***	-0.115	-0.112	-0.083	-0.320***
	(0.101)	(0.100)	(0.102)	(0.060)	(0.074)	(0.071)	(0.079)	(0.062)
Typical Daily Wage in Last Three Months	-0.011	-0.009	-0.003	-0.131***	0.036	0.036	0.052	-0.177***
	(0.058)	(0.058)	(0.061)	(0.035)	(0.104)	(0.105)	(0.109)	(0.038)
Proportion of Hours Spent in Self-Employment or Paid Work in Past	0.003	0.004	0.005	-0.010***	0.002	0.001	0.004	-0.021***
	(0.005)	(0.005)	(0.005)	(0.003)	(0.008)	(0.008)	(0.008)	(0.003)
Super Index of Overall Empowerment (Standardized)	0.049	0.046	0.121	-0.079	-0.159*	-0.156*	-0.079	-0.260***
	(0.082)	(0.082)	(0.085)	(0.072)	(0.081)	(0.082)	(0.089)	(0.071)
Change in Subjective Wellbeing from Five Years Ago to Today	0.276	0.275	0.706***	-0.142	0.176	0.174	0.515***	-0.108
	(0.187)	(0.189)	(0.173)	(0.157)	(0.190)	(0.190)	(0.179)	(0.190)

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. 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). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S11: Attrition Analysis for Primary Round 4 Outcomes
(Baseline Schoolgirls: CCT vs. UCT)

Outcomes	Dependent Variable: =1 if Unconditional Schoolgirl			
	Original	IPW	Lower Bound	Upper Bound
	(1)	(2)	(3)	(4)
Highest Grade Completed	-0.021	-0.020	-0.005	-0.027
	(0.129)	(0.129)	(0.127)	(0.129)
Competencies Score (Standardized)	0.029	0.032	0.039	0.015
	(0.072)	(0.073)	(0.070)	(0.072)
=1 if Ever Married	0.011	0.009	0.013	0.006
	(0.049)	(0.050)	(0.049)	(0.049)
Age at First Marriage	0.536**	0.527**	0.536**	0.527**
	(0.220)	(0.223)	(0.220)	(0.221)
=1 if Ever Pregnant	0.010	0.010	0.013	0.006
	(0.046)	(0.047)	(0.046)	(0.047)
Number of Live Births	-0.053	-0.052	-0.051	-0.061
	(0.053)	(0.054)	(0.054)	(0.053)
Age at First Birth	0.088	0.086	0.088	0.042
	(0.186)	(0.189)	(0.186)	(0.176)
= if HIV Positive	0.003	0.003	0.004	-0.006
	(0.029)	(0.029)	(0.029)	(0.028)
=1 if Anemic	-0.074*	-0.075*	-0.072*	-0.086**
	(0.042)	(0.041)	(0.041)	(0.042)
Opportunity Cost of Time	-0.040	-0.035	-0.037	-0.119
	(0.106)	(0.103)	(0.107)	(0.095)
Typical Daily Wage in Last Three Months	0.079	0.077	0.080	-0.058
	(0.108)	(0.108)	(0.109)	(0.046)
Proportion of Hours Spent in Self-Employment or Paid Work in Past	-0.003	-0.003	-0.003	-0.013*
	(0.009)	(0.010)	(0.009)	(0.007)
Super Index of Overall Empowerment (Standardized)	-0.221**	-0.214**	-0.206*	-0.243**
	(0.106)	(0.107)	(0.108)	(0.104)
Change in Subjective Wellbeing from Five Years Ago to Today	-0.134	-0.134	-0.073	-0.203
	(0.221)	(0.221)	(0.202)	(0.219)

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. 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). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Outcomes	Baseline Schoolgirl						Baseline Dropout	
	CCT vs. Control		UCT vs. Control		CCT vs. UCT		CCT vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	p-value	q-value	p-value	q-value	p-value	q-value	p-value	q-value
Highest Grade Completed	0.136	1.000	0.465	1.000	0.850	1.000	0.000	0.007
Competencies Score (Standardized)	0.269	1.000	0.147	0.478	0.630	1.000	0.263	0.237
=1 if Ever Married	0.206	1.000	0.829	1.000	0.613	1.000	0.001	0.007
Age at First Marriage	0.940	1.000	0.016	0.289	0.032	0.465	0.006	0.022
=1 if Ever Pregnant	0.471	1.000	0.980	1.000	0.614	1.000	0.054	0.099
Number of Live Births	0.580	1.000	0.600	1.000	0.410	1.000	0.007	0.022
Age at First Birth	0.292	1.000	0.997	1.000	0.436	1.000	0.100	0.145
= if HIV Positive	0.955	1.000	0.938	1.000	0.980	1.000	0.649	0.504
=1 if Anemic	0.699	1.000	0.053	0.299	0.068	0.465	0.263	0.237
Opportunity Cost of Time	0.617	1.000	0.120	0.478	0.550	1.000	0.641	0.504
Typical Daily Wage in Last Three Months	0.847	1.000	0.726	1.000	0.665	1.000	0.041	0.090
Proportion of Hours Spent in Self-Employment or Paid Work in Past	0.502	1.000	0.831	1.000	0.842	1.000	0.221	0.237
Super Index of Overall Empowerment (Standardized)	0.551	1.000	0.051	0.299	0.052	0.465	0.890	0.504
Change in Subjective Wellbeing from Five Years Ago to Today	0.143	1.000	0.354	1.000	0.650	1.000	0.263	0.237

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. 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). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S13: Program impacts on marriage market outcomes (husband attitudes towards women's empowerment)

Panel A: Baseline Dropouts

	Attitudes Towards Women's Empowerment Super Index (Standardized)	Index of Preferences for Children's Education (Standardized)	Index of Attitudes Towards Wife's Autonomy (Standardized)	Index of Attitudes Towards (non)-Abuse (Standardized)	Index of Wife's Divorce Prospects (Standardized)	Desired Fertility
			Two Years After Program			
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	0.145	-0.000	0.189	0.162*	-0.162	-0.161
	(0.100)	(0.117)	(0.129)	(0.091)	(0.129)	(0.138)
Mean in Control Group	0.000	0.000	0.000	0.000	0.000	3.649
Sample Size	326	326	326	325	325	324
Panel B: Baseline Schoolgirls						
=1 if Conditional Schoolgirl	0.069	0.013	0.125	0.123	-0.048	0.050
	(0.108)	(0.095)	(0.117)	(0.103)	(0.167)	(0.118)
=1 if Unconditional Schoolgirl	0.254	-0.315*	0.462	0.175	0.171	-0.066
	(0.199)	(0.183)	(0.389)	(0.109)	(0.123)	(0.201)
p-value UCT vs. CCT	0.374	0.078	0.392	0.683	0.252	0.586
p-value Treatment	0.414	0.196	0.325	0.208	0.336	0.837
Mean in Control Group	0.000	0.000	0.000	0.000	0.000	3.194
Sample Size	543	543	543	543	542	541

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. All variables are constructed so that higher values are better. The husband gender empowerment super index is a standardized index of the other variables in this table. Additional detail on the construction of the indices can be found in in Appendix A and Appendix B. 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 husbands of 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.