WHEN THE MONEY RUNS OUT: EVALUATING THE LONGER-TERM IMPACTS OF A TWO YEAR CASH TRANSFER PROGRAM*

Sarah Baird
George Washington University

Craig McIntosh
University of California, San Diego

Berk Özler
World Bank

February 2016

Abstract

We study a group of young women in rural Malawi who were exposed to a cash transfer experiment as adolescents. More than two years after the end of transfers, we find that the substantial short-term benefits of the program have largely evaporated. Unconditional cash transfers (UCT) caused a short-term delay of marriage, fertility, and HIV infection, but the ending of the program is immediately followed by a wave of marriages and pregnancies, accompanied with a catch-up to the control group in HIV prevalence. It does however appear that children born to UCT mothers during the program are significantly taller for their age. For those who had already dropped out when the study began, two years of conditional transfers produced a meaningful long-term increase in educational attainment and a more educated pool of husbands. Even in this group we see no increase in employment rates, wages, real-life capabilities, or empowerment, suggesting that schooling itself has not improved the medium-term labor market prospects of young women in this context

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

JEL Codes: C93, I21, I38, J12

* Thanks to .. The study design was approved by ethical review committees at the National Health Sciences Research Council (Malawi, Protocol #569) the University of California at San Diego (USA, Protocol #090378), and George Washington University (USA, Protocol #061037).

1. INTRODUCTION

The past decade has witnessed an impressive growth in the number, volume, and types of transfer programs in developing countries. A rigorous evidence base has shown that cash transfers can have surprisingly large effects on a range of concurrent outcomes, even if the poor receive these transfers with few (or no) strings attached.[1] However, extant evidence relies mostly 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.[2] As these programs grow to be major vehicles for social protection it becomes increasingly important to understand what happens when we transfer people *off* of them. Hence the longer-term question: can these programs actually help individuals to transition out of poverty, or do benefits evaporate when the money runs out?

For any intervention with a fixed duration to have a sustained effect well after the cessation of support, it needs to lead to an increase in the stock of some asset that keeps producing a stream of returns. For example, a program that provides cash grants to groups of unemployed youth who were required to submit plans for income generating activities may have lasting effects on earnings through the accumulation of physical (productive assets) and human (vocational skills) capital (Blattman, Fiala, and Martínez 2014). Alternatively, large unconditional cash grants to poor households may increase future earnings by increasing investments in productive assets, such as livestock (Haushofer and Shapiro 2015); while small monthly conditional cash transfers over a long period of time may lead to increased consumption after beneficiaries exit the program by increasing savings and investments in small-scale agriculture (Gertler, Martínez, and Rubio-Codina 2012) or by stimulating entrepreneurial activity (Bianchi and Bobba 2013). Thinking about transfer programs targeting younger people, the causal pathway to improved welfare over the long-run is more likely to be human capital accumulation, either in the form of education and skills or health – especially reproductive and sexual health for adolescent females.

---

[1] See, for example, Banerjee et al. (2015) and Bandiera et al. (2015a) for asset transfers combined with complementary activities for the ultra-poor in six countries and Bangladesh, respectively; Beaman et al. (2015) for cash grants to farmers in Mali; Blattman et al. (2015) for cash grants combined with skills training and supervision to young, marginalized villagers in a post-conflict setting in Northern Uganda; de Mel, McKenzie, and Woodruff (2008) for cash and in-kind grants to households with existing microenterprises in post-tsunami Sri Lanka.

[2] Generally speaking, traditional cash transfer programs that provide small, monthly, and often conditional transfers, typically have 12- to 24-month follow-ups. An exception is Gertler, Martínez, and Rubio-Codina (2012), which compares consumption among households that have been in Mexico's Oportunidades program for 5.5 years vs. four years. Behrman, Parker, and Todd (2009) do the same for schooling impacts. Evaluations of programs that provide larger lump-sum grants usually report longer-term results. For example, Blattman, Fiala, and Martínez (2014) and de Mel, McKenzie, and Woodruff (2012) reporting impacts four to six years after the initial grants; Banerjee et al. (2015) and Bandiera et al. (2015a) report impacts one and two years after the cessation of all support, respectively; Haushofer and Shapiro (2015) report outcomes, on average, less than a year after cash transfers.

In this paper, we aim to add to this growing literature by examining the welfare of adolescent girls and young women more than two years after they stopped being beneficiaries of a cash-transfer experiment. As our earlier work has demonstrated short-term effectiveness of the transfers in improving school attendance and test scores, as well as reducing the incidence of pregnancy, marriage, and sexually transmitted infections, it indicated the possibility of finding longer-term improvements in welfare as young adults (Baird, McIntosh, and Özler 2011; Baird et al. 2012).[3] Here, we make three main contributions: First, we examine longer-term outcomes to assess the persistence of effects in the context of a cluster-randomized controlled trial that provided conditional (on school attendance) and unconditional cash transfers. Second, as the study sample is older, we cover a much broader set of outcomes that include school attainment, skills, labor market outcomes, empowerment, and health. Third, as many of the study participants are now married with children, we expand our study sample to include their children and husbands and report on their characteristics.

We find that the short-term improvements observed during the program failed to translate into increased welfare in the longer-run. Substantial reductions in teen marriages, total live births, and sexually transmitted infections among those that received unconditional cash transfers (UCTs) completely disappeared two years after the end of the intervention. Among those that received conditional cash transfers (CCTs), only baseline dropouts, who returned to school in very large numbers during the program, saw sustained effects on school attainment, marriage, pregnancy, total live births, age at first birth, and desired fertility two years later. Even in this group, however, there were no improvements in important longer-term outcomes, such as individual earnings, per capita household consumption, subjective wellbeing, health, or empowerment. In short, the welfare effects of this two-year cash transfer program were largely transient, mainly because the transfers, especially UCTs, did not cause a significant accumulation of capital – physical, human, or social. Even when

[3] A recent review of interventions targeted at adolescents in low- and middle-income countries that reported effects on childbearing-related outcomes lists five types of interventions: cash transfer programs, communications, peer education, school-based interventions or workshops, and health services or counseling (McQueston, Silverman, and Glassman 2013). Rigorous evaluations of these interventions generally indicate positive, albeit modest, effects. 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 long-run (Duflo, Dupas, and Kremer 2015). Empowerment and Livelihood for Adolescents (ELA) program in Uganda showed significant declines in childbearing, marriage, and having had sex unwillingly after two years, as well as increases in self-employment activities and expenditures on private consumption goods (Bandiera et al. 2015b). A systematic review of cash transfer programs by Baird et al. (2013) indicates that both conditional and unconditional cash transfer programs improve school enrollment and attendance, with little effect on test scores – generally confirming findings from earlier reviews of conditional cash transfer programs (Fiszbein and Schady 2009; Saavedra and Garcia 2012).

there was some accumulation of human capital under CCTs, they did not translate into broader gains in important domains.

Our short-term (two-year) findings are consistent with effects found in other studies. Two other experiments compare CCTs with UCTs: Akresh, de Walque, and Kazianga (2013) finds that CCTs are more effective in improving school attendance and test scores than UCTs for "marginal children" who are initially less likely to attend school in Burkina Faso after two years. Benhassine et al. (2015) finds that cash transfers that are "labeled" as education support in Morocco perform as well as CCTs in increasing school participation – also after two years. Elsewhere, Haushofer and Shapiro (2015) find that large unconditional cash transfers to poor significantly improve psychological wellbeing approximately eight months after the transfers. In the short-run, we also found larger improvements in school enrollment and test scores for the CCT group and we detected significant improvements in mental health among adolescents in both CCT and UCT groups (Baird, McIntosh, and Özler 2011; Baird, de Hoop, and Özler 2013).

The longer-term (four-year) findings reported here are cause for concern regarding the promise of these programs for sustained poverty alleviation. With respect to UCTs, there is no doubt that they are very valuable as social protection, leading to improvements in a wide range of outcomes from teen pregnancies to nutritional intake; from mental health to sexually transmitted diseases – while the transfers are in place. However, our evidence suggests that once the money runs out, the reversion to the control group levels is very fast for almost all outcomes. Money is very effective in alleviating the multitude of problems that adolescent females face in this context, but we cannot safely transition them off cash transfers – at least not after two years – because they cause no discernible accumulation of any type of capital.[4]

With respect to CCTs, the story is more nuanced but still not very promising. The raison d'etre of CCTs that are conditional on school attendance is the supposed link between increased schooling and future welfare. Among baseline schoolgirls, almost 90% of whom completed primary school in

[4] Cash transfer programs can serve three types of policy goals: (i) as safety nets, providing a consumption floor for the poor; (ii) as programs to increase human capital accumulation among children to break the cycle of intergenerational poverty; and (iii) as programs that sustainably promote the currently poor out of poverty. In this paper, we are able to examine the effectiveness of CCTs and UCTs with respect to the first two goals among our target population of initially never-married young females: UCTs perform very well as safety nets but fail to improve welfare in the longer-run. CCTs perform less well as safety nets (by denying transfers to non-compliers) but do cause some longer-term changes among groups with substantial increases in school attainment. Our study cannot speak to the effectiveness of cash transfers for promoting households out of poverty, as examined in Gertler, Martínez, and Rubio-Codina (2012) for example, because we followed the adolescent girls and their households across time, rather than the original households of their parents or guardians at baseline.

the control group four years after baseline, the transfers were largely inframarginal, leading to small gains in enrollment and test scores that did not translate into higher attainment or skills at endline. Among baseline dropouts, there were large and lasting effects on school attainment, marriage, and childbearing, but these had not translated into improvements in other domains two years later.

Several reasons might explain the disconnect between higher attainment and no improvements outside of delayed marriage and reduced childbearing. First, it is possible that higher attainment does not provide one with the higher cognitive 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 measured competencies that relate to skills needed in farming and running small household enterprises and detected no effects in these domains. Second, task performance is dependent on not only improvements in cognitive skills, but also on character skills, and effort (Heckman and Kautz 2013). Hence, it is possible that CCTs, by providing incentives for formal schooling, improve only cognitive skills, which may not be sufficient. A mentoring program called "Empowerment and Livelihood for Adolescents" in Uganda that provided young females with "hard" vocational and "soft" life skills was, in contrast, effective in causing sustained improvements in welfare (Bandiera et al. 2015b).[5] Alternatively, if safe and well-paying jobs existed for women in Malawi, parents might invest in the necessary human capital for their daughters on their own – without the need for any outside intervention (Heath and Mobarak forthcoming; Jensen 2012; Munshi and Rosenzweig 2006; Oster and Steinberg 2013).

While the effects of UCTs may be transient for the adolescent beneficiaries, they may have lasting effects on their own children. Policies for child development often target the first 1,000 days from conception to the second birthday (Barham, Macours, and Maluccio 2013). In our experiment, more than 2,000 babies were born to study participants by the endline – with endogenous variation in exposure to the cash transfer program. As well-known channels for growth, such as maternal nutrition and stress, improved during the two-year program, there is reason to think that children of UCT beneficiaries who were exposed to it would also benefit from it.[6] In fact, this is what we find: anthropometric measurements at endline suggest that children of UCT recipients who were born during the two-year program are substantially taller for their age than children in both the control and the CCT group. Comparing our findings to studies evaluating CCT programs that impose explicit

[5] Heckman and Mosso (2014) suggests "The most effective adolescent interventions target formation of personality, socioemotional, and character skills through mentoring and guidance, including providing information."

[6] Unlike UCT recipients, CCT recipients would, by and large, drop out of school and stop receiving transfers after having a baby, so it's less likely that their children would benefit from cash transfers.

conditions on health behaviors, such as attending educational workshops and preventive health visits, we speculate that income support for prospective mothers may be largely responsible for the observed effects on child health (Barham, Macours, and Maluccio 2013; Gertler 2004; Macours, Schady, and Vakis 2008).

The remainder of this paper is structured as follows. Section 2 provides a brief background and describes the 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 children and husbands in Section 5. In Section 6, we provide some concluding remarks, present limitations of our study, and discuss a few policy recommendations.

2. BACKGROUND, STUDY DESIGN, AND DATA

2.1 Background

Adolescent girls in sub-Saharan Africa (SSA) face a multitude of hazards during their transition from childhood to adulthood. Net primary school enrollment for females in the region is lower than 75% -- with lower completion rates and much lower transition rates into secondary school, resulting in net enrollment rates in secondary schools of around 25% in countries like Malawi (World Bank 2013). Age at first marriage, while recently increasing, remains around 18 or 19 in many countries in the region (Marston et al. 2009; Garenne 2008) – with adolescent childbearing rates showing no decline since the 1990s and remaining higher than 100 births per 1,000 adolescent girls in nearly half of the countries (UNFPA 2012). Onset of sexual activity at an early age with older male partners exposes young women to sexually transmitted diseases, with HIV prevalence among females aged 15-24 at 22.7% in Swaziland in 2007, 13.6 in South Africa in 2009 and 5.2 in Malawi in 2010. Along with HIV/AIDS and abortion, depression makes up the leading contribution to disability adjusted life-years in SSA, with one recent study in Malawi showing that more than a third of school-aged girls suffering from psychological distress (Patel et al. 2007; Baird, de Hoop, and Özler 2013).

Governments and aid organizations have responded to this dire picture by designing a variety of interventions targeting school-aged girls and young women. For example, donor organizations like DFID have formed strategic collaborations with foundations like Nike and NoVo to create the Girl Hub and invest in policies and programs that benefit adolescent girls and will have lasting impacts (http://www.girleffect.org/about/girl-hub/). Such interventions are wide ranging in their approaches. A recent review of interventions targeted at adolescents in low- and middle-income countries that reported effects on childbearing-related outcomes lists five types of interventions: cash transfer programs, communications, peer education, school-based interventions or workshops, and health services or counseling (McQueston, Silverman, and Glassman 2013).

2.2. Study Setting

The primary study area is Zomba district, in Southern Malawi, a largely agricultural economy that is characterized by low educational attainment and high HIV rates. Zomba district includes both a large rural population and an urban center in Zomba City, one of Malawi's four large cities. As of 2009, Zomba district was the third poorest district in Malawi. This is also reflected in our sample where real monthly per-capita exchange rate comparable consumption in 2008 was $20.6USD/month. Secondary school completion rates are low—14.6% of our control group passed 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 District Assembly 2009), a number that is likely even lower for females.

2.3 Study Design

Our study began by listing all households within 176 Enumeration Areas (EAs) of the 550 EAs in Zomba District in order to identify those containing never-married adolescent girls between the ages of 13 and 22. This target population was then divided into two main strata: those who were out of school at baseline (baseline dropouts) and those who were in school at baseline (baseline schoolgirls). Baseline dropouts were relatively rare (15% of target population) and so were all recruited into the study. Baseline schoolgirls were sampled into the study at rates increasing in age and rural status.

Treatment was assigned first at the 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. The amount of money received by the household head was randomized between $4 and $10 at the EA level, and the core respondents selected their own individual transfer amounts from between $1 and $5 in a public lottery. Offer letters explaining treatment were distributed in December 2007, payments began in February 2008 and continued through the end of 2009. Survey waves were conducted in 2007, 2008, 2010, and 2012. Figure 1 presents an illustration of the study design, and a more detailed description of the treatments and the sub-experiments that were a part of the study can be found in Appendix A.

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 administratively removed from the program for failing to meet the monthly 80% attendance rate, meaning that if they subsequently had satisfactory attendance, then their payments would resume. Other design aspects of the intervention were kept identical so as to be able to isolate the effect of imposing a schooling conditionality on primary outcomes of interest.[7]

2.3 Data Sources and Outcomes

Data Sources. The focus of this paper is data collected in Round 4 which took place in 2012, two years after the end of the intervention. However, to provide context to these results, we also present results on the same outcomes, when available, for data collected in Round 1-Round 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 and early child development tests for children of core respondents under the age of 5 and a household survey and biomarker data on HIV for husbands of core respondents.

The annual household survey consisted of a multi-topic questionnaire administered to the households in which the sampled respondents resided. It consisted of two parts: one that was administered to the head of the household and the other administered to the core respondent, i.e. the sampled girl from our target population. The former collected information on the household roster, dwelling characteristics, household assets and durables, shocks and consumption. The survey administered to the core respondent provides detailed information about her family background, schooling status, health, dating patterns, sexual behavior, fertility, marriage, and the labor market. In addition to the household survey administered to the core respondent (and to her parents/guardian if

[7] 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 paid 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 satisfactory school attendance.

she still lived with them), the Round 4 survey included a similar module administered to the husbands of married core respondents.

The Round 4 household survey also consisted of a set of questions to try and measuring basic competencies of the core respondent. These competencies included reading and following instructions to apply fertilizer; making correct change during a hypothetical transaction; sending text messages and using a calculator on a mobile phone, and calculating profits for a hypothetical business scenario. They were designed to replace the achievements 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 and their husbands) 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 under 5.

Early childhood development (ECD) tests were administered for all 3-4 year-old children of the core respondent. These tests consisted of the Malawi Development Assessment Tool (MDAT) for fine motor skills, language, and hearing, which were administered directly to the child and the Strengths and Difficulties Test (SDQ) which was administered to the core respondent or the guardian responsible for the child.

Prior to the analysis of data from Round 4, a pre-analysis plan was drafted and registered at the AEA RCT Registry (AEARCTR-0000036; https://www.socialscienceregistry.org/trials/36) that described our analysis plan as well as our primary and secondary outcomes. Our outcomes cover six domains for the core respondent: education, marriage and fertility, health, empowerment and aspirations, employment and wages, and consumption, as well as looking at outcomes in these domains for husbands and children. Many of our outcomes are in the form of an index constructed using the following rubric. First, we ensured that all sub-questions are aligned so that higher numbers 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 normalize each sub-question by subtracting the mean and dividing by the standard deviation. Finally we construct (and then normalize) the raw mean of the normalized variables for all sub-questions within a family of variables to create the final index. We briefly discuss these outcomes below and provide additional detail on both outcome construction and data collection procedures in Appendix B.

Outcomes: Education and Competencies. The primary outcomes we examine for education are self-reported highest grade completed and indices on the overall score and time taken on the competency tests. 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 look at the separate components of the competency index.

Outcomes: Marriage and Fertility. Our primary outcomes look at self-reported data on whether or not the core respondent was ever-married or ever-pregnant. We also look at age at first marriage and age at first birth[8], as well as total live births. Desired fertility is measured as a secondary outcomes.

Outcomes: Health. Our primary health outcomes are HIV and anemia prevalence, both measured with biomarker data. We also look at HIV incidence as a secondary outcome. Additional secondary outcomes include psychological wellbeing measured with the General Health Questionnaire 12 (GHQ-12), the number of meals eaten in the last week that contained, meat, fish, or eggs, and use of reliable birth control.

Outcomes: Sexual behavior. All outcomes in the sexual behavior domain are secondary and self-reported. On the extensive margin, our sexual behavior outcomes include ever had sex, number of lifetime sexual partners, and being sexually active during the past 12 months. On the intensive margin, we look at age at first sex, having a sexual partner 5 years older or more, and condom use during the most recent sexual intercourse.

Outcomes: 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 includes sub-indices (all secondary outcomes) that measure self-efficacy, preferences for child education, an index of social participation, and aspirations. We also construct a super index of empowerment separately for the married sub-sample and the unmarried sub-sample, as well as a super index of economic control for the married sub-sample. These three indices are also primary outcomes and are described in detail in Appendix B.

Outcomes: Employment and Wages. In this domain we look at 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

[8] Our pre-analysis plan suggested we would use a hazard model. We instead simply use OLS to look at age at first marriage and age at first birth.

include whether the core respondent participated in any wage work in the past three months, labor income in the past five seasons, and an effective daily wage rate for work done in the past seven days.

Outcomes: Husband. Our analysis of husbands focuses on husband quality and husband gender 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, and his mental health measured through the GHQ-12. The gender empowerment index, the second primary outcome, includes sub-indices for gender empowerment, wife autonomy, justification for abuse, divorce prospects and desired fertility. All components of these two main indices are presented as secondary outcomes.

Outcomes: Child. The child outcomes sit under four domains: anthropometrics, health, parental practices and educational testing. For anthropometrics we construct height for age and weight for height z-scores (primary outcomes) for living children under 5, as well as the corresponding binary measures of stunting (below -2 standard deviations height for age) and wasting (below -2 standard deviations for weight for height), both secondary outcomes. It is important to note that despite the intended age correction of the anthropometric measures, the Malawian context displays sharp deterioration in height for age beginning immediately after birth (see Appendix Figure C1). Our health outcomes include neonatal and post-neonatal mortality (primary), birthweight, vaccinations, and whether or not the child usually sleeps under a bed-net (secondary). For parental practices we construct variables for exclusive breastfeeding in the past six months and an index of parenting practices (both primary). Finally, for educational testing we conduct the MDAT and SDQ for all 3-4 year olds (both primary outcomes).[9]

3. Estimation Strategy

We first discuss the experimental estimation strategy used to look at core respondent outcomes, before turning to identification challenges when analyzing the children and husband data, and our approach for looking at impact on these outcomes.

3.1. Core Respondents

[9] The pre-analysis plan also indicated that we would look at an indicator for child mortality, but we only have 22 deaths to children over both dropouts and schoolgirls, so we choose not to look at this outcome. We also do not look at nutritional status, as the measure was essential the same as weight for height.

The evaluation of the impact of the ZCTP 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 we employ a simple reduced-form linear probability model of the following form:

$$Y_i = \alpha + \gamma^c T_i^c + \gamma^u T_i^u + \beta X_i + \varepsilon_i \qquad (1)$$

where Y_i is an outcome variable for core-respondent *i,* T_i^C and T_i^U are binary indicators for offers in the CCT and the UCT arms, respectively, and X_i is a vector of baseline characteristics. Note that for baseline dropouts we only have the CCT binary indicator. The standard errors ε_i 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 variables as controls: a household asset index, highest grade attended, a dummy variable for having started sexual activity, and dummy variables for age. These variables were chosen because they are strongly predictive of schooling outcomes and, as a result, improve the precision of the impact estimates. We also include indicators for the strata used to perform block randomization – Zomba Town, within sixteen kilometers of the town, and beyond sixteen kilometers (Bruhn and McKenzie 2008). Age- and stratum-specific sampling weights are used to make the results representative of the target population in the study area.

Table 1 examines attrition in the core respondent sample first looking at baseline dropouts and then at baseline schoolgirls. Our attrition analysis focuses on whether the core respondent is part of our Round 4 household survey sample, the main data source used in our analysis. Appendix Table C1 replicates the core respondent impact analysis for the primary outcomes accounting for this differential attrition, and all results remain robust. Table 2 provides baseline descriptive statistics and shows baseline balance.

3.2 Identification Challenges with Children and Husbands of Core Respondents

We begin our analysis of child and husband outcomes by presenting simple comparisons between the experimental groups following the specification in equation (1). However, the study of impacts on the children and husbands of the core respondents introduces issues of selection and pathways to treatment that are hard to sidestep. Treatment of a set of prospective mothers may lead to impacts on their children either because of selection (which children get born) or because of direct causal changes in outcomes (for the children actually born). Thus, the simple experimental difference between treatment and control children does not isolate any standard direct treatment effect of the program on a specific sample of children. However, because the selection in the composition of children born is endogenous to treatment, any effort to deal with this issue econometrically must confront the well-established fact that conditioning on a post-treatment variable can introduce bias into experimental estimates (Rosenbaum 1984). To structure the analysis, we now present a simple conceptual framework for the ways in which selection and direct treatment effects combine to produce the observed differences in child outcomes between the treatment and control.

The root of the identification problem is that the sample of observed children is not pre-selected prior to the beginning of the experiment. If one wants to understand the casual impact of a maternal treatment program on children, an experiment targeted at the mothers of a pre-defined sample of children presents no special difficulties. The inherent challenge is that as we move towards trying to understand programs that target mothers earlier in the process (such as shortly after conception or even before it) there is no pre-defined sample of children on which to draw. Any such study must be performed on a sample of potential mothers, in which case the issue of selection of the children actually born will arise if the program also affects fertility or mortality. Hence, the child-related empirical issues dealt with in this paper are endemic to the experimental analysis of maternal interventions that begin prior to pregnancy.

The analysis of the effect of the program on husbands is in many ways simpler, since for most husband outcomes it is implausible that the program delivered any direct effect: the most obvious pathway for differences between the treatment and the control is differential husband selection. Even for husbands, however, outcomes such as attitudes towards female empowerment (given the nature of the program) or HIV prevalence (given the impact of the program on CRs) may potentially combine a direct treatment effect on a given husband with a selection effect. Furthermore, partner choice represents another pathway through which the program may potentially affect the children born.

We refer to the straightforward comparison of outcomes for children in the treatment and control as the 'basic experimental estimand.' There are some important ways in which this combined second-generation effect is a quantity of interest from a policy perspective. If we are asking the simple reduced-form question "how does a two-year program change the average child outcome two years after the program ends", then the basic experimental estimand provides an unbiased answer to this question.[10] There are other obvious and critical policy questions, however, which are not answered by

[10] Note, however, that this question is not the average of the individual-level treatment effects over any well-defined population. It is equivalent to analyzing an experiment with differential attrition and asking 'what is the impact on the sample that remains'.

this difference, such as "does the program confer a protective effect on children?" Selection effects can render the former quantity a useless guide to the latter: if for example the treatment causes a temporary fertility delay by a specific type of mother, then we may see a large basic experimental treatment effect that is both illusory and temporary, unwinding over the longer term as the delayed children are eventually born.

To use the terminology from the causal inference literature, selection of the sample of parents and children is a mediator for which we wish to control to isolate the direct effect of the treatment on child outcomes holding selection constant. The problem with controlling for an endogenous mediator is that, once we understand that the mediating variable is itself a product of treatment, comparisons of treatment and control observations with the same value of the mediating variable are "no longer apples-to-apples" (Angrist and Pischke 2008). Analysis that attempts to partial out selection effects in order to identify a direct causal impact on children must necessarily make functional form/ignorability assumptions that are not required for a standard experiment. The assumptions underlying such approaches are strong (Sobel 2008), and the analysis of mechanisms is best approached by experiments specifically designed for the purpose (Bullock, Green, and Ha 2010). On the other hand, the standard casual effect on a given set of children is not what the experiment on mothers delivers for us, and so in this case in order to move towards a causal effect on a given child, we need to attempt to control for selection. What may make mediation analysis less objectionable in this case is that we never have an apples-to-apples comparison, since the treatment and control sample of children may differ in age, maternal age, and parental attributes.

To do this as flexibly as possible, we pursue two approaches simultaneously. First, we consider the sample selection of children present in each birth epoch as a form of attrition, and implement the standard inverse propensity weighting intended to make the observed sample of mothers be representative of the full sample of core respondents. This achieved by regressing an indicator for 'had a child in this epoch' on baseline covariates, treatment, and (covariate * treatment) interactions. The analysis of children is then weighted by the standard sampling weight divided by this estimated epoch- and treatment-specific fertility probability of the mother. Under the assumption that the probit functional form is correct (meaning that the treatment induces selection only on observable variables), this approach adjusts for treatment-induced selection on the extensive margin and makes treatment effects within each group representative of the entire sample of core respondents rather than those who happen to have given birth within any period.

Then, using the weighted data, we add a sequence of endogenous control variables that remove observable forms of heterogeneity via OLS regression adjustment on the intensive margin. This approach to mediation analysis is widely used in social science (Barron and Kenny 1986) and medicine (MacKinnon 1994).[11] An advantage of this approach is that it allows us to examine how treatment effect estimates change as we add sequentially more rigorous controls to attempt to isolate the impact of treatment on a given sample of children.

Figure 2 presents a schematic of the set of pathways thorough which maternal treatment status may change the observed outcome for the child. These pathways naturally break into two groups; the selection channels that alter the identity of the children born to the mother, and direct channels that affect the children actually born. Partitioning the set of channels in this way is useful because it helps to delineate the cases in which controlling for specific endogenous covariates may produce an estimand that has a clearer causal interpretation.

A first obvious example of such an endogenous covariate is child age. A program that alters the timing of fertility will produce a treatment distribution of child age that differs from the control, and so child age is a mediator for the treatment effect, although because it is endogenous it is also a "bad control" in the language of Angrist and Pischke (2009). Nonetheless, the raw difference between the treatment and control is certainly not informative of a meaningful long-term impact if the treatment has only improved the outcome relative to the control because children of the latter group are older on average.[12] Controlling for child age allows us to take a step towards an apples-to-apples comparison between treatment and control by estimating the average of the child age-specific impacts. We denote the component of the total treatment effect arising from differences in child age as a 'trivial' effect, and control for it using a flexible (cubic) functional form of the child's age in months.

A second, related case is maternal age. Even a temporary delay in fertility among adolescent mothers can give rise to an improvement in child outcomes because of greater gynecological maturity at the time of birth. This effect will lead to a permanent improvement in child outcomes even if the

[11] While a more recent literature has suggested means to conduct mediation analysis using principle stratification (Jo 2008) or non-parametric estimators (Flores and Lagunes-Flores 2008), we present simple OLS regressions to make the control structure as transparent as possible.

[12] Even variables that are supposed to be age-standardized, such as the HAZ, may face age-related problems in this context. While the children in our sample are born very close to the mean of the reference group, they decay steadily in height-for-age, ending up almost two standard deviations below the global distribution by the age of 3. This seems to be a common feature of HAZ in very poor countries (see, for example, Figure 1 in Barham, Macours, and Maluccio 2013), and is illustrated in the control group for our study in Appendix Figure C1.

fertility delay is temporary, and hence we refer to this as a 'non-trivial' effect, and control for it using dummy variables for the mother's age in years at birth.

Third, the composition of the sample of mothers introduces an important selection pathway to child outcomes. Permanent changes in fertility due to treatment that are differential across mother type can induce a non-trivial change in child outcomes by altering the composition of children born. These changes are meaningful and durable but do not arise from improvements in outcomes for any specific child. Temporary delays in fertility are more complex to categorize as trivial or non-trivial; the maternal age-driven effect outlined in the previous paragraph can have meaningful effects on child outcomes, but if this effect is weak and lifetime fertility is left unchanged then even very substantial differences in the composition of children born during the program can subsequently be completely unwound. In this case, the post-program period will be a mirror image to the program period in terms of (fixed) maternal quality differences, and so any medium-term treatment-control differences will disappear over time. To control for both permanent and dynamic selection in mother's attributes, we use an extensive battery of baseline maternal characteristics[13] plus the interaction between age at baseline and the most influential variables (location of household and highest grade completed at baseline).

Finally, selection effects can arise because the treatment changes the composition of fathers; we control for this effect using the ethnicity, age, relative wealth, relative landholdings, and health of the father.[14] Changes in partner selection are non-trivial in that their effects on child quality are durable.

These four selection-driven mechanisms work by altering the identity of the child who is actually born. If we were able to net out all of these mechanisms using regression controls, we would be left only with the direct effect that is the impact of the program on the child who is actually born. This latter quantity answers the most obvious policy question that might be asked about cash transfer

[13] Baseline values of: urban, near rural, mother alive, father in household, tested for HIV, self-assessed future HIV risk, household consumption and asset index, highest grade completed, female-headed, household size, never had sex, ever pregnant, total live births, total sexual partners, sexually active in past 12 months, whether certificate for completing primary, middle secondary, high secondary school. In general the maternal attributes are not strong predictors of child outcomes in our sample, and results remain robust to changes in this set of control variables.

[14] While father type selection appears to be non-trivial, the program could generate a 'diversionary' treatment effect (as in Crépon et al 2014) whereby the treatment changes the marital matches made. In this case, a large intent-to-treat effect could be accompanied with no total causal effect (Baird et al. 2014).

programs and children: "Do cash transfer programs confer protective effects on those children who are born to beneficiary households during the program?"[15]

We present six specifications for each epoch. First, the simple difference between treatment and control child outcomes. Then, we use the attrition weights to present the treatment/control difference corrected for differential maternal attrition on the extensive margin. We then add regression controls for each of the intensive-margin mediating selection mechanisms outlined above. In this way we are able to shed some light on the extent to which both trivial and non-trivial forms of selection are responsible for creating treatment-control differences in child outcomes.[16] The third specification (using the selection weights and controlling for child age) has removed mediators that are 'trivial'. The sixth specification, subject to selection on observables, has removed all selection-driven mediators and hence gives the impact of the program on the children actually born. Recognizing the somewhat speculative nature of this analysis, we suggest caution when interpreting these results, but feel that their inclusion is warranted (particularly because the longer-term impacts of the UCT intervention are difficult to interpret without them).

4. Core Respondent Results

This section focuses on the experimental impacts for the core respondents. We look at results across the following domains: education and competencies, marriage and fertility, sexual behavior and HIV, health, employment and consumption, and empowerment. It is worth noting that the focus of our analysis is on impacts at Round 4—more than two years after the intervention ended. We also present results from Round 2 and Round 3 when the same outcome is measured in earlier rounds to show the trajectory of the results over time. The majority of these results from earlier rounds have been previously published, and we cite these publications accordingly.

4.1 Education and Competencies

Table 3 presents program impacts on education and competency outcomes. Panel A looks at baseline dropouts and Panel B focuses on baseline schoolgirls—all tables in this section have this

[15] Gertler (2004) and Fernald et al. (2009) address this question in Mexico for children already born by the time PROGRESA/Oportunidades began in 1998.

[16] It is important to emphasize that even the direct effect is itself composed of many sub-channels, of which the most obvious might be nutrition, maternal stress, and improvements in parenting arising from better human capital. Because such strong assumptions are required even to attempt to isolate the aggregate direct effect, we make no effort to further decompose the direct effect into its component channels.

format. For baseline dropouts, the results suggest that the impacts on educational attainment seen during the program, whether in terms of highest grade completed or highest qualification achieved, have been largely sustained, and in some cases increased.[17] For example, baseline dropouts in the CCT arm were 5.8 and 8.1 percentage points more likely to have achieved a Primary School Leaving Certificate (PSLC) in Round 3 and Round 4 respectively (over a base of 0.351 in Round 3 and 0.366 in Round 4). Impacts on primary school completion are still increasing 17% at the end of the program and now 21% two years later. On the other hand, we find no impact on total competencies (Table 3, Column 13) nor on any of the components (Appendix Table C2). Improved education achievement and short-term gains in cognitive, English and math skills at the two-year follow-up (Baird et al. 2015) did not translate to improved performance on basic competencies in Round 4. We return to a discussion of this result after looking at the impacts for baseline schoolgirls.

The results for baseline schoolgirls suggest little, if any effects, on these outcomes for either treatment group (Table 3, Panel B). The significant and modest increases in education outcomes seen immediately at the end of the two-year intervention for the CCT arm (Baird et al. 2011) get smaller and become statistically insignificant by Round 4. The reader should note that the mean number of years completed in the control group in Round 4 is 10.4 and that 88% of the control group has passed the PSLC exam, meaning that they successfully completed primary school. While these are encouraging numbers for Malawi, they also indicate that close to 90% of the cash transfer recipients in this stratum were inframarginal with respect to primary school completion. The potential effect of cash transfers on these outcomes is limited by definition and targeting students at risk of dropping out may prove to be more cost-effective.

In addition, the short-term impacts on cognitive ability, mathematics, and English reading comprehension in the CCT group (with the last effect being significantly higher than that in the UCT arm) (Baird et al. 2012) did not translate to better performance in practical competencies measured at the five-year follow-up. The index of competencies increased in both the CCT and UCT groups by less than 0.1 standard deviations (SD) and only one out of 10 coefficient estimates is significant at the 10% level (UCT group is more proficient in sending text messages successfully). The results suggest that the earlier improvements in learning were too small to make a difference in the longer-run; that learning decays quickly; or that improved test scores do not translate to more practical life skills, such

[17] A more detailed discussion of education results from earlier rounds for baseline dropouts can be found in Baird et al. (2010) and Baird et al. (2015)

as making correct change during a transaction, reading and following instructions, or calculating profits.

The consistent pattern (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 abilities in a useful way. Arguing against this is the strong observational correlation between test scores and competencies (for example, the correlation between the standardized math score and the standardized total competency for baseline schoolgirls is 0.488 ($p<0.000$) which implies that a one standard deviation increase in the math test leads to a 0.488 standard deviation increase in the total competency). It therefore appears that a more likely explanation for this is that the increased school learning engendered by the CCT programs did not amplify the kinds of practical abilities that the competencies were geared to measure. This is in and of itself an interesting result in terms of the specific types of human capital that we can hope to improve through the use of schooling CCT programs.

4.2 Marriage and Fertility

As with the education outcomes, CCTs had large effects on marriage and fertility for baseline dropouts that were sustained five years after baseline (Table 4, Panel A).[18] Baseline dropouts were 14.0, 15.7, and 10.7 percentage points less likely to have been ever married at one-, two-, and five-year follow-ups, respectively (all significant at 99% confidence). The corresponding reductions were 5.7, 8.1, and 3.8 percentage points for being ever pregnant (all significant the 90% confidence or higher). Furthermore, there was a negative gradient in the total number of live births, with a reduction of 0.143 children during as of Round 3 ($p<0.001$). This implies that two years after the cash transfers were discontinued; the number of children in the CCT group was 0.143 lower than the control mean of 1.35, corresponding to a 10% reduction. Similarly, there was a modest decline in desired fertility of 0.16 children ($p<0.1$). Age at first marriage and first birth were similarly higher by 0.41 and 0.31 years, respectively. To the extent that delaying marriage and fertility are important predictors of future welfare for these young women, these delays may lead to sustainable improvements even if they eventually experience catch-up in marriage and fertility.

[18] A more detailed discussion of marriage and fertility results from earlier rounds for baseline dropouts can be found in Baird et al. (2010) and Baird et al. (2015)

As with earlier rounds, we find no effects on any of the six outcomes presented in this table for the CCT group of baseline schoolgirls. By Round 4, 40% of the baseline schoolgirls in the control group were ever married, 50% ever pregnant, and with the average age of marriage (among those married) being 18.6. There are no changes for these variables in the CCT arm. Interestingly, desired fertility is also unchanged in this group – with the mean number of children desired approximately three. Switching to examining impacts in the UCT arm, the coefficient estimates confirm earlier findings that UCTs were effective in reducing marriage and pregnancy rates among baseline schoolgirls during and immediately after the program (Baird et al. 2011). However, we see an almost complete reversal of these outcomes at the five-year follow-up: there are no longer any differences in ever married, ever pregnant, or the total number of live births between the UCT group and either of the two other study arms. 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 were quickly married following the end of the intervention. As with the CCT group, desired fertility also remains unchanged in this group.

These findings suggest that in the UCT arm any effects of the program were temporary and stemmed from having the cash on hand (income effect). In the CCT arm, one possible explanation of the sustained effects in the baseline dropout strata, with no impact in the baseline schoolgirl strata, is that the human capital gains were only large enough among baseline dropouts to have knock on effects to marriage and fertility. This result again points to the limitations of a CCT when the majority of the transfers are inframarginal.

4.3 HIV and Sexual Behavior

As presented in Baird et al. (2012) we find no impact on HIV prevalence in Round 2 among baseline dropouts. Similarly, we find no long term impact on HIV prevalence nor an impact on HIV incidence (Table 5, Panel A). This result suggests that the improvements in education and delays in marriage and fertility are not translating to reduction in HIV. Appendix Table C3 and C4 look at risky sexual behavior on the extensive and intensive margin. While the program was ongoing, baseline dropouts in the CCT group were significantly less likely to be sexually active—consistent with the effects on schooling and fertility. But, as of Round 4, virtually the entire group of baseline dropouts (97%) is sexually active, with no significant impact on any measure of sexual activity.

The significant treatment impacts on HIV seen for baseline schoolgirls during the program (Baird et al. 2011) have dissipated by Round 4 (Table 5, Panel B), indicating complete catch up. While the temporarily delay in HIV infection is still likely beneficial for the adolescent girl, the trajectory of HIV prevalence shows that the benefits of the program disappeared once the program ended. Figure 3 plots HIV prevalence from Round 2-Round 4. One interesting finding from Figure 3 is that catch-up in the CCT arm occurs during the program, while catch up in the UCT arm is sustained until the program ends—in line with the impacts on marriage and fertility. Neither type of cash transfer program had any effect on sexual behavior during or after the program– either on the extensive margin (such as onset of sexual activity or number of partners) or on the intensive margin (age at first sex, condom use, age of partner).

4.4 Health and Nutrition

Even with the improvements in education and delays in fertility in marriage for baseline dropouts we find no impact on anemia, mental health or number of meals eaten with a source of protein (meat, fish or eggs) during or after the program (Table 6, Panel A). For baseline schoolgirls, the picture is one of significant and meaningful effects during the program disappearing by Round 4. Looking at nutritional intake, we find that both types of cash transfers led to a significant increase in the number of meals during which a source of protein (meat, fish, or eggs) was consumed during Rounds 2 and 3: the effects are on the order of half a meal over a control mean of approximately four such meals per week, or about a 10% improvement (Baird et al. 2015). These effects also disappeared two years after the cash transfers were stopped consistent with an income effect (Table 6, Panel B).

Second, examining the prevalence of psychological distress (using a binary indicator based on the GHQ-12 index), we confirm earlier findings (Baird, de Hoop, and Özler 2013) of large improvements in mental wellbeing during the program that disappeared as soon as the cash transfers stopped. The analysis here confirms that there were no further changes between Rounds 3 and 4. A small, but reliable source of income seems to reduce psychological distress (i.e. the possibility of suffering from mild anxiety and depressive disorders), but does not cause sustained improvements beyond the intervention period (Table 6, Panel B).

4.5 Employment

The opportunities for any sort of formal employment for the women in our sample are minimal. Only six percent of the control group baseline dropouts' time is spent in self-employment or paid work (Table 7, Panel A), although XX% of them are no longer in school. There is no treatment impact on opportunity cost of time or time spent in work, and a significant negative effect on typical wage (corresponding to about a 15 cent decrease). This negative effect may reflect the fact that individuals in the treatment group were in school longer, and thus have been working for less time. We also find no impact on effective daily wage, labor income in the past five seasons, and any wage work in the past three months (Appendix Table C5, Panel A).

The story for baseline schoolgirls is similar, with very limited opportunities for any sort of formal employment. In the control group, only three percent of baseline schoolgirls participate in self-employment or paid work, and there is no treatment effect for either the CCT or UCT intervention across any of the primary employment outcomes (Table 7, Panel B). Given the limited labor market opportunities for women (and men) in this setting, it is clear that demand side interventions are likely not going to be enough to promote increased labor market participation, and that improvements need to be made on the supply side.

4.6 Empowerment

Although the increased schooling, and delayed fertility and marriage did not lead to visible gains in the labor market, it is possible that they will show up in measures of empowerment. Table 8, Panel A looks at the impact of the CCT on measures of empowerment. We find no impacts on a super index of empowerment overall, nor on indices constructed separately for the never married and married sub-groups. We also find no impact on where they feel they stand in society. When we break down these super indices into sub-components (Appendix Table C6, Panel A), we again see no impact of the CCT program. This result suggests that impacts of the program are not translating into any observable measure of gender empowerment.

For baseline schoolgirls in the CCT group we similarly see no impact of the intervention on empowerment, although coefficients are generally positive. However, for UCTs, we see strong negative impacts of the program on the super empowerment index compared to both the control and the CCT groups (Table 8, Panel B). Baseline schoolgirls in the UCT arms are 0.159 standard deviations lower in terms of the super-empowerment index, and it looks like this is largely driven by self-efficacy (Appendix Table C6, Panel B). We also see that this negative impact shows up even stronger on the married super index of empowerment (-0.357 standard deviations). This disempowerment effect of taking the money away that shows up in the UCT arm but not in the CCT arm perhaps points to a difference in the perception of a UCT vs. a CCT. This may be a result specific to having a two year program and taking it away, but does make one think about the dynamics of these programs. We will discuss the results on married disempowerment more when we discuss husbands below.

5. CHILD AND HUSBAND RESULTS

5.1 Child Results

Before presenting the child analysis it is worth restating what we already know to be true of the child sample based on analysis of the mothers (who comprise the clean experimental sample). Among baseline dropouts, fertility was lower among the CCT group in every round of the study and, by Round 4, the total number of live births was lower than the control by about 0.15 children – indicating that roughly 10% of the children that would have been born to the CCT group are missing from the sample. Given the substantial decrease in desired fertility seen among this group, these 'missing' children may never be born.[19]

In the sample of baseline schoolgirls, we see the UCT treatment depressing fertility almost as strongly as the CCT treatment among baseline dropouts in Round 3, which was conducted right at the end of the cash transfer program. However, once the unconditional cash transfers stopped, the total number of live births rebounded roughly halfway by Round 4 (see Table 4, Panel B). Because this treatment temporarily delayed fertility without an effect on overall birth rates or desired fertility, it will have changed the age composition of the children – with more R4 children aged 0-2 born to UCT mothers and more 3-4 year-olds born to the control group.

We begin our analysis of child outcomes from the broadest sample possible, namely the universe of births that occurred during the course of the study. Using these outcomes we can analyze neo-natal mortality (defined for all births) and post-neo-natal mortality (defined for all children who were old enough at R4 to observe the outcome). Despite the possibility that the income transfers from the program or the additional schooling would have insulated children of against mortality, Columns (1) and (2) of Table 9 show no evidence that mortality rates were lower in the treatment group.

In Columns (3) – (6) we examine our most objective child outcomes, namely the anthropometric data collected on all children between the ages of 0 and 5. In Columns (3) and (5) we consider the continuous standard Z-score measures of height-for-age. Columns (4) and (6) we use the binary definitions of these measures, stunting and wasting. The Dropouts see an increase in weight for height that is significant at 10%, but no significant shift in stunting or wasting. For the SG group,

---

[19] It is interesting to note that the treatment effect on actual fertility (-.147***) is 85% of the treatment effect on desired fertility (-. 172***), which would suggest that most of the fertility decline that might be caused by the program has occurred already.

CCT treatment has no effect on the biometric outcomes. The UCT treatment is associated with a significant, .3 standard deviation decrease in weight for height and a very small increase in wasting, but the treatment group is also somewhat taller and so this measure is difficult to interpret.

Moving to outcomes that have a stronger direct parental role, Column (7) examines birthweight. This outcome is taken from birth records where possible, but is self-reported by the parent otherwise: this provides a measure that is observed for the very large majority of births that occurred in the sample but is potentially open to reporting bias. Strangely, the CCT program appears to have had a positive effect on birthweight in the Dropout sample and a negative one in the Schoolgirl sample. In terms of parenting practices, the UCT is associated with significant improvements in all three measures (an overall score, the number of vaccinations, and whether the child was exclusively breastfed for the first six months. Appendix Table C8 examines the impact of the program on a comprehensive set of indicators designed to capture child development, social problems, hyperactivity, and problems with peers, and finds no evidence that the program shifted these indicators.

This concludes the presentation of the basic experimental impacts on child outcomes, and we now move into a sequence of more exploratory analyses that impose increasing degrees of endogenous control on the data.

5.2 Exploiting Variation in Exposure to the Program

Because the program ran for only two years and then shut down two years before the R4 survey, there is substantial heterogeneity in the extent to which children were exposed to it. Furthermore, only the UCT arm continued to receive transfers once out of school, and in Malawi schooling and pregnancy are essentially mutually exclusive. Thus we expect that the direct effect of the cash transfers themselves on child outcomes should be largely localized to the group of children who a) were born during or shortly after the program, and b) are in the UCT arm. Maternal nutrition and mental health (both important drivers of child outcomes) in the UCT were significantly different during the program and not at R4 provides reinforcing mechanisms to support the hypothesis that children born while unconditional transfers were running would have benefited from them. However, in Baird et al. (2012) we showed that the fertility and marriage effects of the UCT arm were primarily driven by girls who had dropped out of school, and so we may expect substantial selection effects in this group as well.

To investigate how this differential exposure drives treatment effects, we divide the sample of children into three 'epochs'. The first epoch captures those **directly exposed** to the program, meaning those born prior to the end of the transfers. This cohort is exposed for a maximum of two years, with some combination of in utero and child exposure depending on the exact birthdate of the child. The second epoch covers those born within 9 months of the program, who were **exposed in utero** for a maximum of nine months (those born immediately after the program ended) with declining and earlier in utero exposure for later-born children. The third epoch covers those born more than 9 months after the end of the program, who were **not exposed.**[20]

Analysis of impacts within epochs, however, may exacerbate problems of dynamic selection bias. To see this, Figure 4a plots smoothed monthly birthrates for each of the SG treatment arms. The UCT treatment depresses fertility strongly in the six months on either side of the end of the program, but a pronounced UCT 'baby boom' is visible 11-13 months after the program ends. A large fraction of the UCT sample therefore became pregnant within 2-4 months of the cash running out; 1 in 12 UCT girls conceived a child in these three months while only 1 in 30 SG CCT or Control girls did. This certainly indicates that the age distribution of children will be different even within epochs for the UCT arm, and may also indicate the presence of differential maternal type selection when we examine heterogeneity in child outcomes by date of birth.

Table 10 presents the results of the analysis by epoch. We focus on the most objectively and universally measures available, namely mortality and the anthropometric outcomes. The DO CCT treatment shows no difference from the control for children born during the program, an increase in height for age among those born within 9 months, and then potential negative effects for children born after the program ended (mortality rates are higher, and while weight-for-height is improved, this is within a sample that is also somewhat shorter than the control). The SG CCT treatment shows a small decrease in mortality in the second and third epochs. The 'epoch' approach may be of the most interest for the UCT group, given the strong and temporary effects of the program on the mothers in this arm. Here we see extremely large effects of the unconditional transfers on height for age as long as the program was in effect, but these effects drop off quickly and become weakly negative for children born after the program ended.

[20] Timing of birth is endogenous for reasons described in Section 3.2, and so it is important to recognize that simply analyzing births that occurred within windows of time can exacerbate selection problems that arise from temporary fertility delays.

Appendix Figure C2 provides a visual representation of the temporal effects of the UCT program on HAZ 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 90% confidence interval. For this group in which impacts on the mothers were so substantial and ephemeral, we see significant increases of almost a full standard deviation in weight-for-height, but these effects begin to evaporate for children exposed to the program for less than a year, and go to zero for children born at the moment the program ended. This suggests child exposure to UCT is beneficial, but in utero exposure does not seem to have conferred these same benefits.

Several of the results presented so far, however, should make us concerned about the possibility that these results might arise from mother selection, rather than any real changes to the outcomes of given children. For the DO group we see large changes in education as well as desired fertility in R4, suggesting that the composition of mothers has changed in some durable way. While this is an effect we defined as non-trivial, it still would provide a very different causal interpretation to the increases in child mortality than an improvement for a given set of pregnancies. While the large effects of the UCT program on HAZ provide tantalizing evidence that it could provide meaningful protection to children born into households receiving UCTs, and yet we also see a large temporary decrease in fertility (Figure 4a) with no shift in desired lifetime fertility among the UCTs (Table 4, Panel B). This combination of factors appears particularly consistent with of the type of temporary composition shift that has a trivial effect on the medium-run basic experimental estimand. Hence it is difficult to conclude on the basic matter of whether the UCT protected exposed children without resolving the mechanism for this effect.[21]

To shed some (admittedly speculative) light on this question, we proceed in Tabld 11 with the methodology outlined in Section 3.2. Because HAZ is objectively measured and represents the most important potential benefit of the program on children, we focus our analysis on this outcome. We begin by presenting the epoch-level regressions with no controls at all except gender of the child (column 1).[22] Column 2 introduces the attrition propensity weights, whose purpose is to make the

[21] Appendix Table C7 examines maternal selection directly. The first column shows that while the Dropout child sample is selected overall (as we saw from the CR analysis) there is no evidence of the treatment having induced a systematically different type of CR to be mothers. In the UCT treatment, we see that the depressive effects of the treatment on fertility were largest in urban and peri-urban areas, and among girls who were in the lowest grades at baseline. We thus enter the analysis of children with clear evidence of selection: there are fewer overall children in the Dropout treatment, and the UCT treatment children are disproportionately likely to be born to mothers who are rural and close to completing their education at the time the program began .

[22] Note that the estimates in the first column do not include the basic core respondent controls used throughout the paper, and hence differ from the results presented in Table 11.

estimated impacts representative of the original sample of core respondents rather than those who actually give birth in each epoch. Column 3 adds the cubic in child age, column 4 the maternal covariates and their interactions with age, Column 5 controls for maternal age at birth, and Column 6 controls for the father's attributes.

For the DO children, controls for selection generally reinforce the sense that the beneficial effect of the program were concentrated among those born within 9 months of it ending. The selection controls do not substantially change point estimates but reinforce significance of impacts within this second epoch. The SG CCT group sees no significant impact in any epoch under any specification. For the SG UCT group, when we focus on Column 3 of the 'During' epoch, which has removed the 'trivial' components of selection, we see an impact that is still almost three quarters of a standard deviation increase in height for age, significant at the 99% level. Controlling for mother attributes and age, and finally father attributes in Column 6, we see an impact that is now just under 10% significant but still suggests a treatment effect of almost a half a standard deviation. Recognizing this analysis as being speculative, we find confirmation here of the idea that cash transfers do provide a protective effect on children born into households receiving them. For the UCTs, however, the impact of the program ending appears to be detrimental, rather than positive as in the DO group. This impact appears to have been masked by the unequal age composition of children caused by the 'baby boom' after the ending of the program. Overall the contradiction in impacts on children born shortly after the program ended are consonant with the overall results of the paper: the CCT beneficiaries acquired other benefits from the program that had a more lasting set of impacts on outcomes, while life deteriorates rapidly for the UCT beneficiaries when the money runs out.

It is also possible to represent the heterogeneity in UCT height for age treatment effects visually as we change the control structure. To do this, we start by smoothing the weighted raw outcome data by month of birth for each treatment arm and subtracting the control from the treatment to provide a treatment effect at each age. We then predict the outcome for each observation, using the 'leave-one-out' prediction technique recommended by Abadie et al. (2014) to prevent overfitting. We predict child outcomes in the control group using coefficients from a regression model that begins only with gender, and then adds on the sequence of control variables as in the table above. By subtracting these predicted outcomes using different controls of off the true observed outcome, we net out the channels used in the regression and can estimate the magnitude of the direct treatment effect that does not pass through the mechanisms used as predictors. The resulting image, as shown in Figure C3, illustrates the relative imperviousness of the HAZ results to any selection controls we can include. The suggestion is thus that the mediator of selection is not driving this result, and that the simple treatment/control average gives a number close to the direct causal impact of the UCT on a given set of children.

Finally, to get a sense of significance across the birth timing distribution we can re-run the Fan regression using the full set of controls included in Column 5 of the regressions. When we do this, Figure 5 shows that the UCT program appears to confer a significant protective effect on the height for age of children born up to within a few months of the end of the program, even when we control for the obvious mechanisms of selection.

5.3 Husband Results

As with the analysis of children, we begin the discussion of the marital impacts of the program by recapping the relevant results for the core respondent herself. For the DO group, we know that there were large sustained decreases in marriage (suggesting the potential for substantial selection on the extensive margin), and that those marriages that did occur took place later than they otherwise would have. The CCT SG group has few impacts that suggest changes in marriage outcomes. The UCT SG group, on the other hand, saw a purely transitory delay in marriage and a large decrease in gender empowerment arising entirely from the married core respondents.

The literature on marriage markets in developing countries points to several primary mechanisms through which we should expect a delay in taking a husband to effect outcomes for women. First, by waiting to take a husband an adolescent girl may expand her opportunity to pursue education (Herrera and Sahn 2014). Second, environments in which adolescent marriage is common may feature a preference for young brides (Foster & Khan 2000), and hence a girl who waits longer may face worse prospects, resulting in lower husband quality (Field and Ambrus 2008). The cultural environment of Zomba is largely lacking in brideprice[23], and so we do not expect this to be an operative mechanism for reinforcing the benefits of expanded access to girls' education as shown in Ashraf et al. 2014, and nor does the presence of this 'market' mechanism complicate general equilibrium effects of large-scale changes in marriage patterns as illustrated by Anderson and Bidner 2015. Consequently, for the DO group who both delayed marriage and acquired additional education

[23] The word 'brideprice' is used to refer to payment from the husband's family to the wife's at the time of marriage, as is common in most of the Islamic world and Sub-Saharan Africa. 'Dowry' refers to payment from the bride's family to the husband's, as is common South and East Asia. Zomba is a predominantly matrilocal environment and while males are expected to contribute to the female's household on an ongoing basis, there is typically no lump-sum brideprice paid at the time of marriage.

we expect the net effect on husband quality may be positive, while for the UCT SG group who temporarily delayed marriage but acquired no additional education, we expect husband quality to be lower.

Turning to the results on husbands, Table 12 tests for differences between the treatment and control samples.[24] Among DOs, husbands are more likely to have received a secondary certificate (MSCE), and while insignificant the increase in average years of education of .56 years is very similar to the improvement seen in the DO core respondents. This is evidence of assortative matching in education whereby dropouts causally induced to return to school are marrying more educated husbands with than they would otherwise have found. These husbands are not otherwise different in terms of wages, intelligence, marital fidelity, mental health, HIV (Appendix Table C10), or attitudes towards gender empowerment (Appendix Table C11). For the SG CCT group, just as there is no evidence of impacts on marriage rates, there is no evidence that the types of men married has been strongly altered by the program (despite generally positive coefficients on husband quality and a 99% significant positive impact on marital fidelity, the number of outcomes examined is large and only one is significant). So, as with the core respondent schooling results, the SG CCT impacts look like scaled-down versions of the DO CCT impacts which are not significant given our sample size.

For the SG UCT group, the picture is darker and more complex. The general pattern is that UCT husbands are inferior; they have somewhat lower scores in the overall quality index and lower education by half a year (t-statistic: 1.07). More revealingly, on the objectively measured cognitive test they score lower than control group husbands by .36 standard deviations, significant at 95%. Appendix Table C11 shows that the UCT husbands have a less positive attitude towards girl's empowerment, significant at the 90% level. A possible mechanism for these differences can be seen by comparing Figure 4a, on the relative timing of births, with Figure 4b, on the relative timing of marriages. The large 'baby boom' apparent in the UCT group 10-12 months after the end of the program is preceded by only a few months by a similarly-sized 'marriage boom'; the fact that these marriages come 2-4 months before birth of a child indicates that a large share of these late UCT marriages may have been forced by a pregnancy. Thus, consistent with the broader literature on the effect of marriage delay in the absence of either education improvements or a tradition of brideprice, it appears that the UCT girls have suffered to some extent on the marriage market by waiting to get married.

---

[24] Appendix Table C9 shows that the SG husbands were not differentially likely to attrite, while the DO husbands are 6% less likely to do so, significant only at the 90% level.

6. Conclusion

In this paper, we present a variety of outcomes – educational attainment, learning and skills, marriage and fertility, health and nutrition, and sexual behavior – for a group of school-aged girls in Malawi who were part of a cash transfer experiment. The outcomes we examine were measured approximately one, two, and five years after baseline data collection. Our findings indicate that while CCTs and UCTs had significant effects on many of these outcomes during and immediately after the program, these effects had mostly dissipated two years after the experiment ended. We argue that the lack of medium-term effects, especially among baseline schoolgirls, is due to the fact that there was an insufficient amount of human capital accumulation. In fact, the only large and sustained effects (in marriage, fertility, and desired fertility) was among baseline dropouts, which was accompanied by large increases in school attainment in this more 'at risk' group of adolescent girls and young women, many of whom returned to school in response to offers of CCTs.

The baseline dropout group does see large and meaningful changes in marriage, fertility, and even desired lifetime fertility. While the total numbers of live births for the control and CCT groups were very similar at Round 2, girls in the CCT group had approximately 0.1 less births at Round 3 and 0.15 births at R4 – meaning that the fertility gap relative to the control actually expanded by about 50% during the two years after the program ended. Marriage rates remain 10pp lower at round 4 (a decrease of 25% on the control group rate) and it seems clear that the program and the associated surge in the probability of returning to school will lead to substantial long-term changes in the marriage and fertility patterns of this group.

Nonetheless, even in this (baseline dropout) group where real improvements in education and teen marriage were seen, these did not translate into increases in employment rates, wages, or applied competencies within two years. This result is particularly discouraging as it seems to suggest that schooling itself is ineffective at improving economic outcomes in this context. This would appear to be a fundamental challenge to CCT programs, for which the connection between improved schooling and higher adult incomes is a critical long-term link in the long-term causal chain. Perhaps we should not be too surprised at this result given the 6% formal employment rate prevailing in Zomba district, but even our 'capabilities' measure failed to see any improvement under the CCT. Applied skills such as using a calculator, making change, and being able to read instructions on fertilizer application are highly correlated with socio-economic status and overall years of education, but were not improved by the program. So this suggests that the schooling induced by our program has not, at least within two years, proven productive.

Adolescents receiving UCTs simply delayed pregnancies and marriages in a very temporary way: the cash transfer program was akin to pushing a pause button for these school-aged girls for two years, but once that button was released they engaged in these behaviors at rates higher than the counterfactual. There is a pronounced marriage and baby boom in the months after the program ends, and HIV within two years has caught up with the trajectory on which it would have been had the program never taken place. The UCT does not seem to have enabled beneficiaries to accumulate physical or human capital that could translate into a different trajectory in life.

Our results on child impacts are intriguing but not as empirically sound as the simple experimental results for their mothers, our core respondents. We provide evidence that income support averaging \$10 per month to mothers of newborns produced large improvements in Height for Age Z-Scores that persist into older children. This result is complicated by the fact that the experiment altered the pattern of fertility for mothers, meaning that the sample of children born may be endogenously selected. We provide a framework in which to think through the potential selection effects, and show that even once we have controlled for these mechanisms the increase in HAZ for children born during the program is three quarters of a standard deviation. Is this effect credible? It is based on controlling for endogenous covariates and thus is not a purely experimentally identified parameter. Nutritional supplementation programs of similar duration typically return benefits of roughly $.2\sigma$, suggesting that the consumption of the mother alone would not be enough to justify such a large effect. On the other hand, these results must be seen in the context of the precipitous drop in HAZ by age for the entire sample (as in Appendix Figure C1). Seen through this lens, the UCT program (once we control for selection) prevented just over a third of the decline seen in Malawian children relative to the international norm. Given the broad improvements in consumption, mental health, and sexual partnerships among the UCT sample, this magnitude may be more credible.

The strong heterogeneity of impacts across strata reinforces the importance of targeting in CCT programs. For those in school at baseline many of the transfers were infra-marginal,[25] and CCTs really have no durable effects. Those out of school at baseline received less money on average (because their compliance rate was lower) but the conditionality drove much larger improvements in schooling. In this group there are substantial decreases in teen marriage, desired and actual fertility, as well as a

[25] By Round 4, 88% of control baseline schoolgirls had completed primary school and 54% had obtained a junior secondary school certificate

more educated set of husbands. These results reinforce a point that has been made many times in the CCT literature: the logic of the condition compels us to target these programs towards groups that would not otherwise have complied with the conditions. Only here can the 'price effect' play a causal role in driving long-term outcomes. Since we find the effects of the money itself to be very fleeting, this only emphasizes the importance of targeting schooling CCTs to those most likely to leave school otherwise.

Our study environment has several limitations in terms of informing us about the longer-term effects of cash transfers more generally. First, the data capture outcomes two years after the money ended, and so we cannot speak to the truly long-term effects of the program. Particularly, the improved human capital obtained by the CCT dropout group may take longer than 2 years to bear fruit in terms of employment or wages. Second, the program itself only ran for two years, which is not the typical environment in the broader world of cash transfer programs.[26] So while we find the benefit of the transfers to be fleeting, this does not mean that a program that had been in place for a longer period of time might not have created more long-lasting benefits. For example, had the UCT program been in place until these women had decided to marry free of financial pressures, might it have avoided the precipitous group of marriages occurring right after the benefits ran out? And particularly, if the CCT program had been in place long enough to create a more meaningful boost in education (such as widespread secondary completion) then employment outcomes might have improved. Also, we are limited by the fact that we track the original adolescent beneficiary of the household as she marries and forms her own household, and hence cannot speak to medium term benefits that may occur in the parental household that received the majority of the transfers.

Another caveat is that much of the recent excitement about UCTs has come from programs such as GiveDirectly that distribute large lump-sums of money, which is not what we did here. Large cash transfers may be more useful for enabling real asset investment, which is an obvious channel to longer-run impacts. Our results should not be taken to indicate that any UCT would have fleeting effects, but rather that in this context a more palliative monthly payment UCT did not lead to the accumulation of durable human capital for the beneficiaries (although it does seem to have done so for their children).

The strength of the study is its focus on the longer-term fate of adolescent girls, and the ways in which cash transfer programs may affect them across a range of outcomes. Because we went to great

[26] Indeed, it is precisely the very long-term eligibility that Mexican households have now had for Oportunidades that drives the development of Prospera, which is intended to segue people off of cash transfer programs.

lengths to track this girls as they grow into women, leave their birth households and form households and families of their own, we are in a strong position to speak to the ways in which their lives evolve. Our results speak to the strong relationships that exist between income and marriage in this context, and the extent to which extending schooling and delaying childbearing may (or may not) create opportunities down the road.

It is important to reiterate that the core message from the overall study is that cash transfers can generate a wide variety of benefits for adolescent girls at the time they are receiving them. What we show here is that many of these benefits do not outlast the program, and there is some evidence that the cessation of transfers (particularly if unconditional) may actually be harmful in the short term. Sustained gains require not only behavior change in the short-run, but the accumulation of some sort of capital – physical, human, or social – during and after the program. Without a significant accumulation of some combination of skills, health, knowledge, information, and networks, sustained future gains from programs targeting school-aged children and young women are unlikely.

REFERENCES (INCOMPLETE)

Abadie, A., Chingos, M. M., & West, M. R. (2013). Endogenous stratification in randomized experiments (No. w19742). National Bureau of Economic Research.

Akresh, R., de Walque, D. & Kazianga, H. 2013, “Cash Transfers and Child Schooling: Evidence from a Randomized Evaluation of the Role of Conditionality”, The World Bank, Policy Research Working Paper No. 6340.

Angrist, J. D., & Pischke, J. S. (2008). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Baird, S., de Hoop, J. and Özler, B. 2013, “Income Shocks and Adolescent Mental Health,” Journal of Human Resources, vol. 48, no. 2, pp. 370-403.

Baird, S., Garfein, R., McIntosh, C., and Özler, B. 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, Vol. 379(9823), pp. 1320-1329.

Baird, S., McIntosh, C. & Özler, B. 2011, "Cash or Condition? Evidence from a Cash Transfer Experiment", Quarterly Journal of Economics, vol. 126, no. 4, pp. 1709-1753.

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., Sulaiman, M., 2012, “Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda”, Unpublished Manuscript.

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." American Economic Review, 103(3): 467-71.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of personality and social psychology, 51(6), 1173.

Behrman, J. R., & Hoddinott, J. (2005). Programme evaluation with unobserved heterogeneity and selective implementation: The Mexican PROGRESA impact on child nutrition. Oxford bulletin of economics and statistics, 67(4), 547-569

Behrman, J.R., Parker, S.W. & Todd, P.E. 2011, "Do Conditional Cash Transfers for Schooling Generate Lasting Benefits? A Five-Year Follow-up of PROGRESA/Oportunidades", Journal of Human Resources, vol. 46, no. 1, pp. 93-122.

Benhassine, N., Devoto, F., Duflo, E., Dupas, P. and Poliquen, V. 2013, “Turning a Shove into a Nudge? A “Labeled Cash Transfer” for Education”, NBER Working Paper No. 19227.

Blattman, C., Fiala, N. & Martinez, S. 2013, “Generating Skilled Self-Employment in Developing Countries: Experimental Evidence from Uganda”, Quarterly Journal of Economics, Forthcoming.

Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism?(don't expect an easy answer). Journal of personality and social psychology, 98(4), 550.

Case, A., Hosegood, V. & Lund, F. 2005, "The reach and impact of Child Support Grants: Evidence from KwaZulu-Natal", Development Southern Africa, vol. 22, no. 4, pp. 467-482.

Duflo, E., Dupas, P., and Kremer, M. 2012, "Education, HIV, and Early Fertility: Experimental Evidence from Kenya", Working Paper.

Fan, J. (1992). Design-adaptive nonparametric regression. Journal of the American statistical Association, 87(420), 998-1004.

Fernald, L. C., Gertler, P. J., & Neufeld, L. M. (2009). 10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study. The Lancet, 374(9706), 1997-2005.

Field, E. and Glennerster, R. 2013, "Empowering Girls in Rural Bangladesh", (accessed May 6, 2013), http://www.povertyactionlab.org/evaluation/empowering-girls-rural-bangladesh.

Fiszbein, A. & Schady, N. 2009, Conditional Cash Transfers: Reducing Present and Future Poverty, With Francisco H. G. Ferreira, Margaret Grosh, Nial Kelleher, Pedro Olinto, and Emmanuel Skoufias; Washington, D.C.:World Bank.

Flores, C. A., & Flores-Lagunes, A. (2008). Nonparametric partial and point identification of net or direct causal effects.

Garenne, M. 2008, Fertility Changes in Sub-Saharan Africa. DHS Comparative Reports No. 18. Calverton, Maryland, USA: Macro International Inc.

Government of Malawi., 2012, "2012 Global AIDS Response Progress Report: Malawi Country Report for 2010/2011.

Haushofer, J. & Shapiro, J., 2013 "Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya", Working Paper.

Jo, B. (2008). Causal inference in randomized experiments with mediational processes. Psychological Methods, 13(4), 314.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention and intervention research. NIDA research monograph, 139, 127-127.

Marston M, Slaymaker E, Cremin I, et al. 2009, "Trends in marriage and time spent single in sub-Saharan Africa: a comparative analysis of six population-based cohort studies and nine Demographic and Health Surveys." Sexually Transmitted Infections 85(Suppl. 1), i64–i71.

Miller, C., Tsoka, M., Reichert, K. 2008, Impact evaluation of the Mchinji Social Cash Transfer. Report to the Government of Malawi and other stakeholders including UNICEF and USAID. Center for Global Health and Development. Boston, MA.

National Statistical Office (NSO) [Malawi] and ORC Macro. 2005, Malawi Demographic and Health Survey 2004. Calverton, Maryland: NSO and ORC Macro.

Patel, V., Fisher, A.J., Hetrick, S. and McGory, P. 2007, "Mental Health of Young People: a Global Public Health Challenge." Lancet, 369, 1302-1313.

Pearl, J. (1995). Causal diagrams for empirical research. Biometrika, 82(4), 669-688.

Rosenbaum, P. R. (1984). The consquences of adjustment for a concomitant variable that has been affected by the treatment. Journal of the Royal Statistical Society. Series A (General), 656-666.

Saavedra, J.E. & Garcia, S., 2012, "Impacts of Conditional Cash Transfers on Educational Outcomes in Developing Countries: A Meta-analysis." RAND Corporation Working Papers, WR-921-1.

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

UNAIDS. 2013, "Global report: UNAIDS report on the global AIDS epidemic 2013," UNAIDS: Geneva. Available from: http://www.unaids.org/en/media/unaids/contentassets/documents/epidemiology/2013/gr2013/UNAIDS_Global_Report_2013_en.pdf

United Nations Population Fund (UNFPA). 2012,"Status Report Adolescents and Young People in Sub-Saharan Africa Opportunities and Challenges", South Africa.

World Bank. 2010. World Development Indicators. Washington, DC: World Bank.

World Bank. 2013. World Development Indicators. Washington, DC: World Bank.

Tables

Table 1: Attrition

Table 1: Attrition

	Baseline Schoolgirl	Baseline Dropout
	HH Round 4	HH Round 4
	(1)	(3)
=1 if Conditional	0.055***	-0.007
	(0.019)	(0.031)
=1 if Unconditional	0.058***	
	(0.023)	
p-value UCT vs. CCT	0.896	N/A
p-value Treatment	0.004	N/A
Mean in Control Group	0.875	0.843
Number of observations	2,273	885

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

Table 2: Baseline Means and Balance

	Baseline Schoolgirl				Baseline Dropout	
	Mean (s.d.)			p-value (CCT-UCT)	Mean (s.d.)	
	Control group	Conditional group	Unconditional Group		Control group	Conditional group
Urban Household	0.348	0.475	0.427	0.783	0.181	0.126
	(0.477)	(0.500)	(0.496)		(0.385)	(0.333)
Mother Alive	0.841	0.798	0.834	0.304	0.786	0.754
	(0.366)	(0.402)	(0.373)		(0.410)	(0.431)
Father Alive	0.71	0.716	0.767	0.238	0.659	0.651
	(0.454)	(0.451)	(0.424)		(0.475)	(0.477)
Household Size	6.38	6.349	6.664	0.168	6.118	6.138
	(2.265)	(2.145)	(2.070)		(2.403)	(2.623)
Asset Index	0.637	1.063	1.342*	0.563	-0.806	-0.722
	(2.579)	(2.709)	(2.433)		(2.246)	(2.487)
Age	15.219	14.911*	15.433	0.004	17.622	17.188
	(1.897)	(1.826)	(1.918)		(2.385)	(2.493)
Highest Grade Attended	7.498	7.242	7.906**	0.005	6.142	5.955
	(1.646)	(1.599)	(1.580)		(2.857)	(2.877)
Never Had Sex	0.803	0.806	0.786	0.604	0.305	0.293
	(0.398)	(0.395)	(0.411)		(0.461)	(0.456)
Ever Pregnant	0.021	0.030	0.030	0.981	0.447	0.417
	(0.144)	(0.170)	(0.170)		(0.498)	(0.494)

Notes: Mean differences statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence. Stars on the coefficients in columns (2) and (3) indicate significantly different than the control group for baseline schoolgirls. Stars on the coefficients in column (6) indicate significantly different than the control group for baseline dropouts. Means are weighted to make them representative of the target population in the study EAs.

Table 3. Education and Competency Outcomes

Panel A: Baseline Dropouts

	Highest Grade Completed			**=1 if Passed Primary School (PSLC)**			**=1 if Passed Junior Secondary School (JCE)**			**=1 if Passed Secondary School (MSCE)**			**Total Competency (Standardized)**
	Round 2	**Round 3**	**Round 4**	**Round 2**	**Round 3**	**Round 4**	**Round 2**	**Round 3**	**Round 4**	**Round 2**	**Round 3**	**Round 4**	**Round 4**
=1 if Conditional Schoolgirl	0.579***	0.558***	0.621***	0.030	0.058**	0.081***	0.012	0.049**	0.034	0.004	0.003	0.016	0.064
	(0.073)	(0.102)	(0.125)	(0.025)	(0.025)	(0.026)	(0.019)	(0.021)	(0.022)	(0.008)	(0.010)	(0.011)	(0.057)
Mean in Control Group	6.345	6.967	6.997	0.328	0.351	0.366	0.085	0.123	0.136	0.008	0.025	0.026	0.000
Sample Size	697	718	744	697	718	744	697	718	744	697	718	744	742
Panel B: Baseline Schoolgirls													
=1 if Conditional Schoolgirl	0.078	0.126*	0.120	0.030	0.013	-0.014	-0.013	0.055**	0.033	-0.004*	0.005	0.006	0.065
	(0.090)	(0.069)	(0.080)	(0.039)	(0.024)	(0.019)	(0.022)	(0.028)	(0.028)	(0.002)	(0.011)	(0.021)	(0.058)
=1 if Unconditional Schoolgirl	0.122	0.103	0.095	0.046	0.030	0.017	0.002	0.016	0.010	-0.006*	-0.009	-0.065**	0.098
	(0.109)	(0.121)	(0.129)	(0.038)	(0.026)	(0.016)	(0.022)	(0.045)	(0.035)	(0.003)	(0.015)	(0.027)	(0.067)
p-value UCT vs. CCT	0.708	0.854	0.850	0.755	0.600	0.166	0.546	0.439	0.565	0.325	0.385	0.022	0.630
p-value Treatment	0.469	0.174	0.309	0.386	0.488	0.359	0.797	0.148	0.486	0.150	0.683	0.045	0.297
Mean in Control Group	8.590	9.677	10.415	0.496	0.776	0.879	0.144	0.337	0.537	0.004	0.054	0.170	0.000
Sample Size	1,965	2,019	2,049	1,967	2,019	2,047	1,967	2,019	2,047	1,967	2,019	2,047	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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Note that in Round 2 and Round 3 highest grade completed is actually highest grade attended. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 4: Marriage and Fertility Outcomes

	Ever Married			Age First Marriage	Ever Pregnant			Number of Live Births			Age First Birth	Desired Fertility
	Round 2	Round 3	Round 4	Round 4	Round 2	Round 3	Round 4	Round 2	Round 3	Round 4	Round 4	Round 4
Panel A: Baseline Dropouts												
=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. 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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 5: HIV

	HIV Prevalence			HIV Incidence
	Round 2	Round 3	Round 4	R4-R3
Panel A: Baseline Dropouts				
=1 if Conditional Schoolgirl	0.032	0.024	0.016	-0.004
	(0.027)	(0.025)	(0.029)	(0.014)
Mean in Control Group	0.098	0.122	0.156	0.034
Sample Size	417	769	738	699
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	-0.020**	-0.005	-0.001	0.005
	(0.009)	(0.011)	(0.019)	(0.013)
=1 if Unconditional Schoolgirl	-0.018	-0.021*	-0.006	0.015
	(0.012)	(0.012)	(0.024)	(0.017)
p-value UCT vs. CCT	0.818	0.235	0.850	0.610
p-value Treatment	0.080	0.218	0.966	0.656
Mean in Control Group	0.034	0.042	0.061	0.020
Sample Size	1,287	2,145	2,000	1,958

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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 6: Health and Nutrition

Panel A: Baseline Dropouts	**=1 if Anemic**	**=1 if Suffers from Psysological Distress**			**Number of Meals Eaten**		
	Round 4	**Round 2**	**Round 3**	**Round 4**	**Round 2**	**Round 3**	**Round 4**
=1 if Conditional Schoolgirl	0.037	-0.002	0.010	0.038	0.326	0.224	0.228
	(0.034)	(0.039)	(0.036)	(0.042)	(0.202)	(0.192)	(0.181)
Mean in Control Group	0.255	0.463	0.314	0.424	3.678	3.989	3.741
Sample Size	714	698	715	743	698	718	744
Panel B: Baseline Schoolgirls							
=1 if Conditional Schoolgirl	0.012	-0.068**	-0.037	-0.030	0.385**	0.596***	0.072
	(0.031)	(0.032)	(0.047)	(0.032)	(0.195)	(0.174)	(0.141)
=1 if Unconditional Schoolgirl	-0.065*	-0.139***	-0.026	-0.002	0.445**	0.338**	-0.043
	(0.033)	(0.035)	(0.054)	(0.046)	(0.199)	(0.153)	(0.240)
p-value UCT vs. CCT	0.074	0.068	0.860	0.552	0.814	0.215	0.672
p-value Treatment	0.123	0.000	0.677	0.627	0.023	0.001	0.858
Mean in Control Group	0.243	0.372	0.313	0.369	3.967	4.052	4.134
Sample Size	1,979	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. Baseline values of the following variables are included as controls in the regression analyses: age indicators, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Note that in Round 2 and Round 3 highest grade completed is actually highest grade attended. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 7: Employment Outcomes (Primary)

	Opportunity Cost of Time	Typical Wage in Past Three Months	Proportion of Hours Spent in Self-Employment or Paid Work in Past Week
Panel A: Baseline Dropouts			
=1 if Conditional Schoolgirl	-11.077	-41.927**	-0.011
	(23.718)	(20.336)	(0.009)
Mean in Control Group	212.324	112.661	0.061
Sample Size	718	743	744
Panel B: Baseline Schoolgirls			
=1 if Conditional Schoolgirl	-15.207	-3.387	0.003
	(30.374)	(17.502)	(0.005)
=1 if Unconditional Schoolgirl	-34.577	10.948	0.002
	(22.126)	(31.185)	(0.008)
p-value UCT vs. CCT	0.550	0.665	0.842
p-value Treatment	0.297	0.910	0.784
Mean in Control Group	269.565	63.566	0.029
Sample Size	2,002	2,048	2,045

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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 8: Empowerment Outcomes (Primary)

Panel A: Baseline Dropouts	**Super-index of Overall Empowerment**	**Change in Ladder from Five Years Ago to Today**	**Super-Index of Unmarried Empowerment**	**Super-Index of Married Empowerment**
=1 if Conditional Schoolgirl	-0.083	-0.032	0.018	-0.130
	(0.074)	(0.232)	(0.112)	(0.098)
Mean in Control Group	0.000	1.120	0.000	0
Sample Size	744	744	289	455
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	0.049	0.276	0.111	-0.005
	(0.082)	(0.187)	(0.098)	(0.099)
=1 if Unconditional Schoolgirl	-0.159**	0.176	-0.094	-0.357**
	(0.081)	(0.190)	(0.109)	(0.173)
p-value UCT vs. CCT	0.052	0.650	0.120	0.068
p-value Treatment	0.101	0.306	0.287	0.121
Mean in Control Group	0.000	0.906	0.000	0.000
Sample Size	2,049	2,049	1,271	776

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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table 9 Child Overall

Basic Child Outcomes.

Child Outcomes, Dropouts

	Neo-natal Mortality	Post-neo-natal Mortality	Height for Age	Stunting	Weight for Height	Wasting	Birthweight	Parenting Score	Total Number of Vaccinations	Exclusively Breastfed in First 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
=1 if Treatment Dropout	0.013	-0.009	-0.013	0.018	0.147*	-0.010	0.111*	-0.003	-0.091*	0.030
	(0.011)	(0.013)	(0.091)	(0.034)	(0.083)	(0.009)	(0.065)	(0.018)	(0.052)	(0.026)
Number of observations	958	707	742	742	742	742	841	861	950	971

Child Outcomes, Schoolgirls

	Neo-natal Mortality	Post-neo-natal Mortality	Height for Age	Stunting	Weight for Height	Wasting	Birthweight	Parenting Score	Total Number of Vaccinations	Exclusively Breastfed in First 6 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
=1 if Conditional Schoolgirl	-0.014	0.005	0.096	-0.046	0.070	-0.007	-0.084**	0.012	0.022	0.029
	(0.009)	(0.012)	(0.109)	(0.039)	(0.077)	(0.009)	(0.042)	(0.018)	(0.071)	(0.033)
=1 if Unconditional Schoolgirl	-0.012	0.001	0.065	0.048	-0.318**	0.006	-0.078	0.050*	0.183**	0.126***
	(0.012)	(0.010)	(0.176)	(0.061)	(0.136)	(0.018)	(0.060)	(0.029)	(0.077)	(0.039)
Number of observations	1,167	756	1,032	1,032	1,032	1,032	1,076	1,090	1,144	1,169

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs.

Table 10 Child by Epoch.

By Epoch Dropouts

	Neonatal Mortality	Height for Age	Weight for Height
DURING			
=1 if Treatment Dropout	-0.005	-0.011	-0.007
	(0.015)	(0.119)	(0.103)
Number of observations	545	367	367
WITHIN 9 MONTHS			
=1 if Treatment Dropout	0.034	0.580*	0.018
	(0.033)	(0.309)	(0.249)
Number of observations	100	88	88
AFTER			
=1 if Treatment Dropout	0.039**	-0.199	0.411***
	(0.019)	(0.161)	(0.132)
Number of observations	313	287	287

By Epoch Schoolgirls

	Neonatal Mortality	Height for Age	Weight for Height
DURING			
=1 if Conditional Schoolgirl	0.009	0.103	-0.098
	(0.022)	(0.173)	(0.131)
=1 if Unconditional Schoolgirl	0.018	0.833**	-0.184
	(0.031)	(0.389)	(0.188)
Number of observations	389	315	315
WITHIN 9 MONTHS			
=1 if Conditional Schoolgirl	-0.026*	0.403	0.229
	(0.016)	(0.286)	(0.169)
=1 if Unconditional Schoolgirl	-0.034	0.148	-0.020
	(0.022)	(0.393)	(0.171)
Number of observations	234	211	211
AFTER			
=1 if Conditional Schoolgirl	-0.019**	-0.031	0.146
	(0.009)	(0.165)	(0.141)
=1 if Unconditional Schoolgirl	-0.022**	-0.410*	-0.418**
	(0.009)	(0.211)	(0.193)
Number of observations	543	506	506

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs.

Table 11 Child HAZ Results, Varying Control Structure.

Panel A. Dropouts

	Born During Program						Born Within 9 Months						Born After Program					
	Gender	Selection weighted, Gender	+ Child Age	+ Mother Attributes	+ Mother Age	+ Father Attributes	Gender	Selection weighted, Gender	+ Child Age	+ Mother Attributes	+ Mother Age	+ Father Attributes	Gender	Selection weighted, Gender	+ Child Age	+ Mother Attributes	+ Mother Age	+ Father Attributes
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	-0.015	-0.200	-0.193	-0.220	-0.096	-0.100	0.353	0.405	0.385	0.475*	0.642**	0.596**	-0.269	-0.175	-0.186	-0.208	-0.251	-0.219
	(0.128)	(0.174)	(0.155)	(0.150)	(0.140)	(0.140)	(0.296)	(0.316)	(0.286)	(0.263)	(0.266)	(0.247)	(0.168)	(0.173)	(0.164)	(0.155)	(0.162)	(0.161)
Number of observations	367	367	367	364	364	364	88	88	88	86	86	86	287	287	287	284	284	284

Panel B. Schoolgirls

	Born During Program						Born Within 9 Months						Born After Program					
=1 if Conditional Schoolgirl	0.155	-0.038	-0.015	0.070	0.087	0.124	0.251	0.156	0.099	0.125	0.127	0.163	-0.011	0.186	0.298	0.228	0.257	0.265
	(0.162)	(0.190)	(0.200)	(0.164)	(0.149)	(0.157)	(0.279)	(0.280)	(0.218)	(0.194)	(0.203)	(0.193)	(0.187)	(0.245)	(0.243)	(0.175)	(0.161)	(0.162)
=1 if Unconditional Schoolgirl	0.953**	0.658***	0.716***	0.524*	0.500*	0.483	0.177	0.130	-0.483*	-0.469**	-0.447**	-0.513**	-0.351**	-0.365*	-0.102	-0.100	-0.098	-0.109
	(0.476)	(0.219)	(0.263)	(0.313)	(0.290)	(0.320)	(0.514)	(0.348)	(0.257)	(0.231)	(0.216)	(0.202)	(0.174)	(0.200)	(0.170)	(0.157)	(0.151)	(0.168)
Number of observations	315	315	315	315	315	315	214	211	211	210	210	210	507	506	506	503	503	503
Control Structure:																		
Uses attrition propensity weight		X	X	X	X	X		X	X	X	X	X		X	X	X	X	X
Cubic in child age in months			X	X	X	X			X	X	X	X			X	X	X	X
Mother baseline control battery, excepting maternal age				X	X	X				X	X	X				X	X	X
Dummies for maternal age in years, age interactions					X	X					X	X					X	X
Father characteristics						X						X						X

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. note: *** p<0.01, ** p<0.05, * p<0.1

Table 12 Husband Overall

Panel A: Dropouts	Husband Quality Index (1)	Highest Grade Completed (2)	PSLC (Primary certificate) (3)	JCE (Middle Secondary certificate) (4)	MSCE (Secondary Completion certificate) (5)	Daily Wage (6)	Currently Employed (7)	Cognitive Test (8)	Marital Fidelity (9)	Mental Health (10)
=1 if Treatment Dropout	0.084	0.561	0.032	0.029	0.074**	-24.5	-0.024	-0.049	0.032	0.014
	(0.106)	(0.348)	(0.054)	(0.046)	(0.037)	(67.582)	(0.040)	(0.110)	(0.106)	(0.126)
Number of observations	326	326	326	326	326	325	326	323	325	326
Control Group Mean	0.000	7.806	0.526	0.314	0.097	358.809	0.246	0.000	0.000	0.000

Panel B: Schoolgirls	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
=1 if Conditional Schoolgirl	0.141	0.046	0.024	0.012	0.059	4.3	0.045	0.014	0.284***	0.154
	(0.096)	(0.271)	(0.043)	(0.049)	(0.053)	(78.656)	(0.051)	(0.109)	(0.091)	(0.126)
=1 if Unconditional Schoolgirl	-0.186	-0.454	0.005	0.017	-0.088	-122.0	-0.091	-0.357**	0.013	0.016
	(0.180)	(0.425)	(0.068)	(0.086)	(0.054)	(103.321)	(0.093)	(0.163)	(0.219)	(0.194)
Number of observations	543	543	543	543	543	540	543	539	542	541
Control Group Mean	0.000	9.743	0.699	0.541	0.258	426.537	0.352	0.000	0.000	0.000
F test: CCT=UCT	3.025	1.391	0.081	0.003	4.227	1.487	1.899	4.119	1.669	0.441
p-value on F-test	0.084	0.240	0.776	0.954	0.042	0.225	0.170	0.044	0.199	0.508

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. Regressions include the standard battery of core respondent-level baseline controls.

FIGURES

Figure 1. Research Design

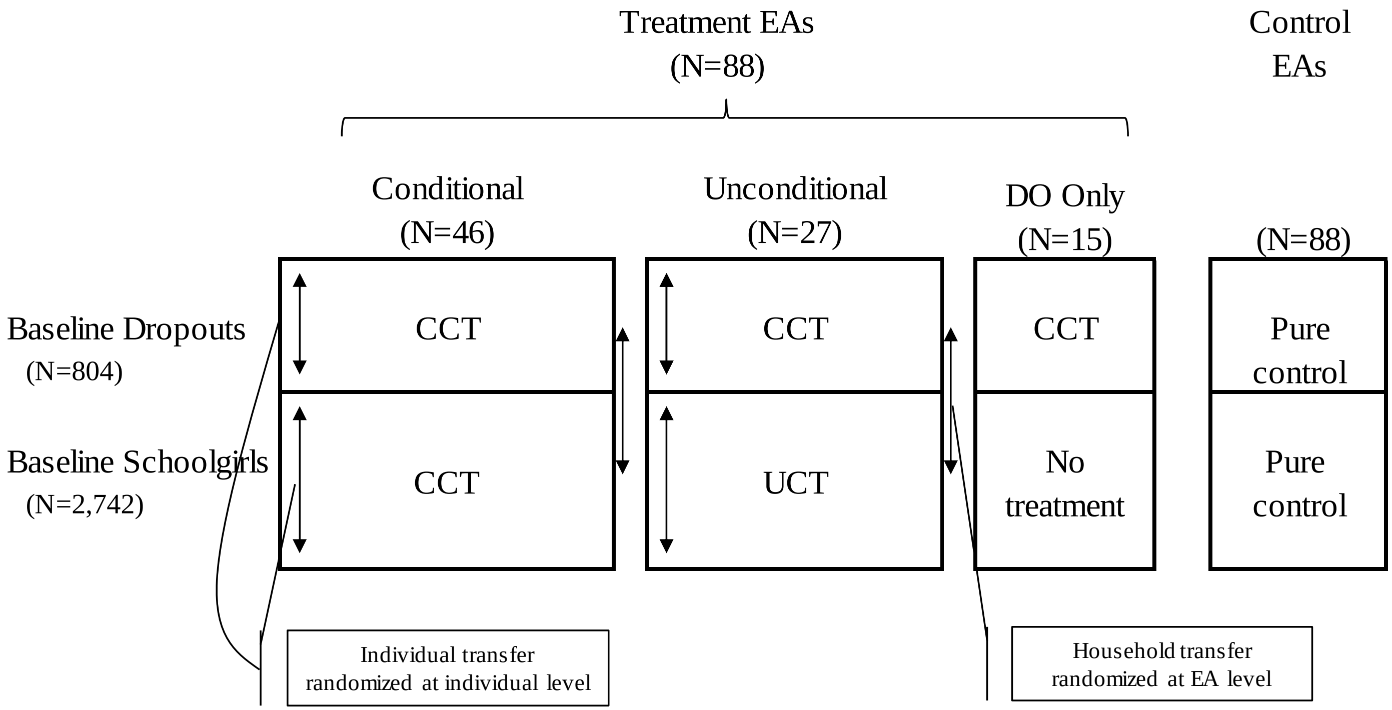


Figure 2. Selection and Direct Treatment Effects on Child Outcomes

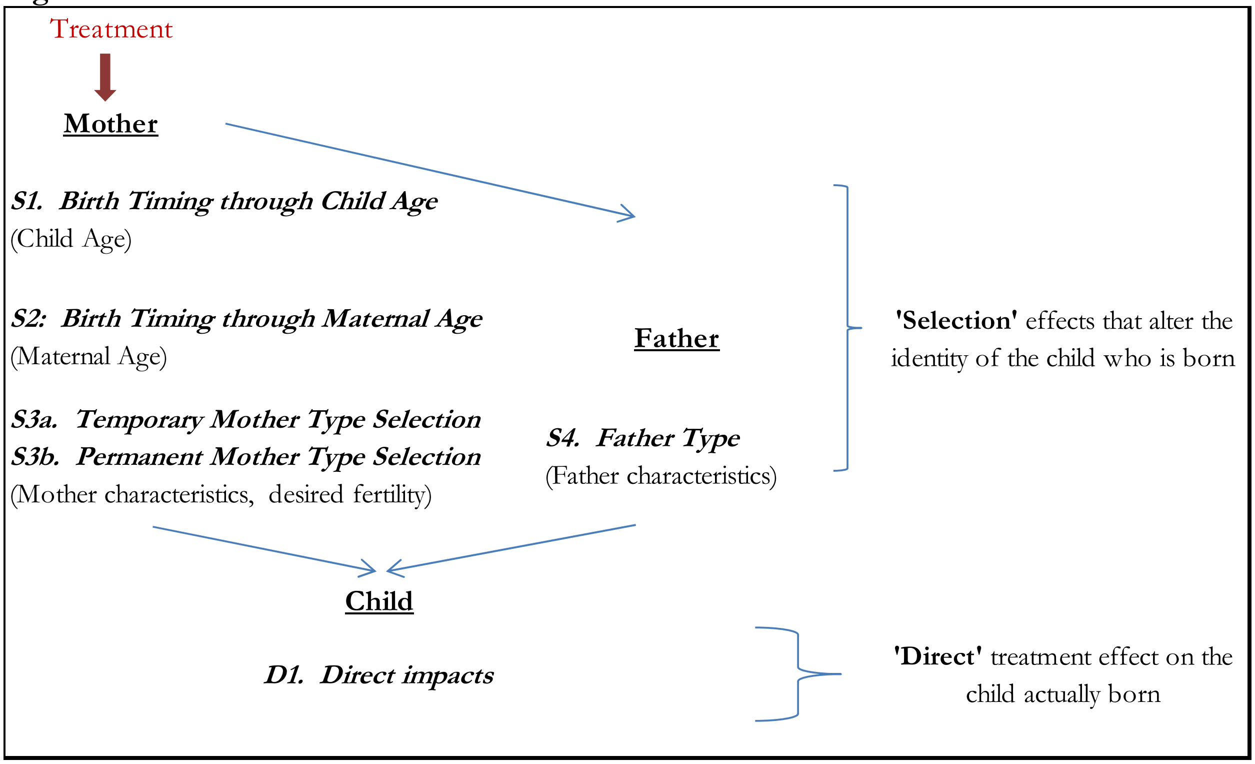


Figure 3: HIV Prevalence Over time (by treatment arm, baseline schoolgirls)

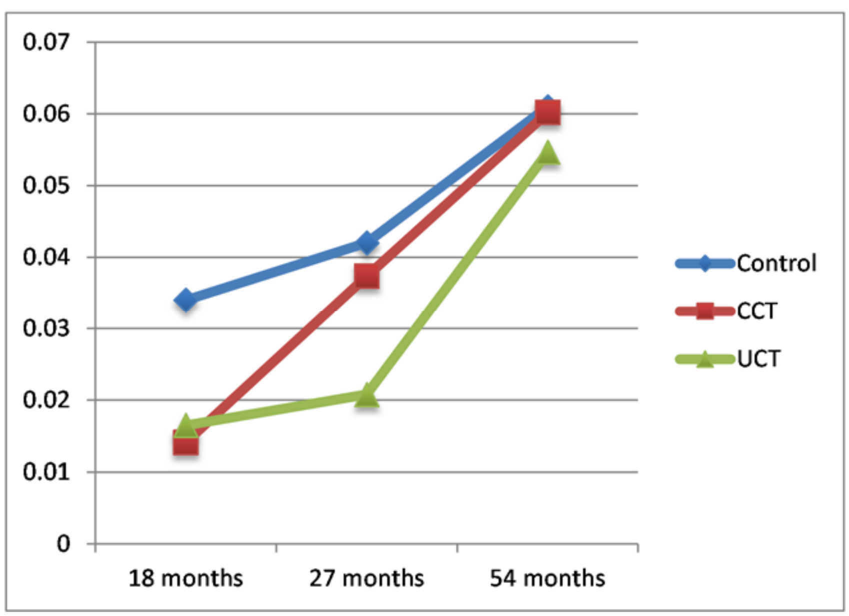


Figure 4a. Monthly Fertility Rates among the Baseline Schoolgirls.

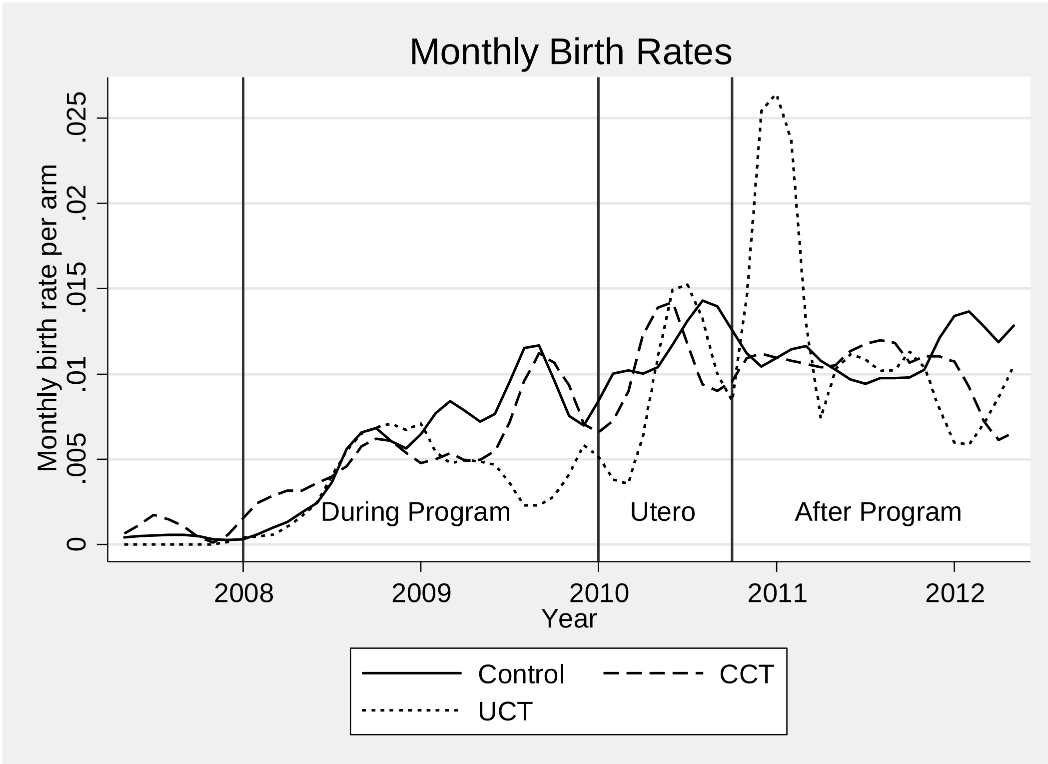


Figure 4b. Monthly Marriage Rates among the Baseline Schoolgirls.

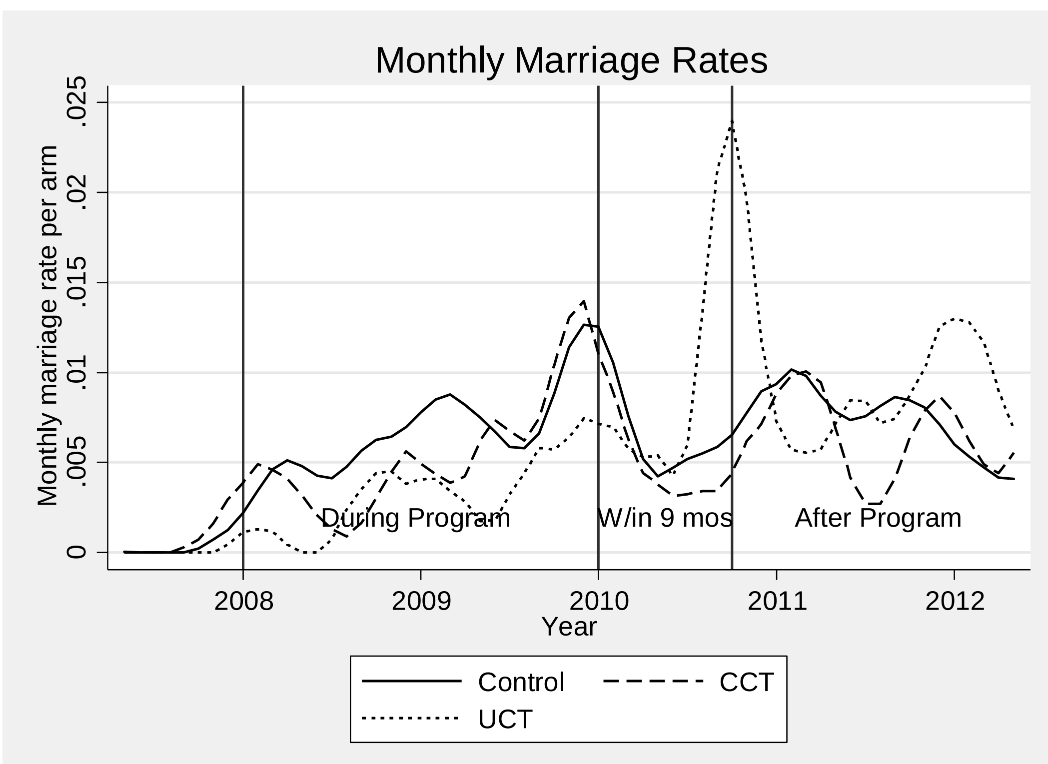
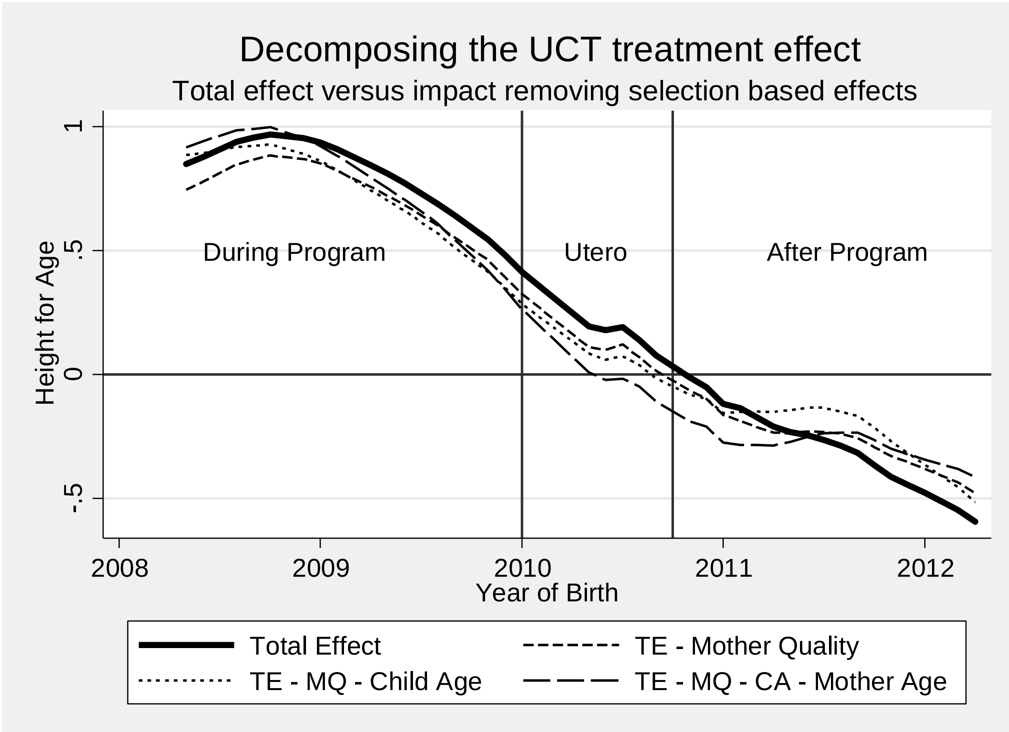


Figure 5. Controlling for selection effects to decompose total HAZ treatment differential:



APPENDIX

Appendix A: Study Design (TO BE ADDED)

Appendix B: Construction of Variables (TO BE ADDED)

Appendix C: Tables and Figures

Appendix Table C1: Replication of CR Impacts Accounting for Differential Attrition.

Appendix Table C2: Competency Components

Panel A: Baseline Dropouts	**Total Competency**	**Fertilizer**	**Change Given**	**Text Message**	**Calculator**	**Profit**	**Total Time**
=1 if Conditional Schoolgirl	0.064	-0.044	-0.014	0.101	0.065	0.094	-0.007
	(0.057)	(0.069)	(0.062)	(0.072)	(0.071)	(0.076)	(0.091)
Mean in Control Group	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	742	742	741	741	741	742	742
Panel B: Baseline Schoolgirls							
=1 if Conditional Schoolgirl	0.065	0.015	0.048	0.077	0.060	-0.006	-0.113
	(0.058)	(0.071)	(0.071)	(0.070)	(0.054)	(0.076)	(0.085)
=1 if Unconditional Schoolgirl	0.098	0.096	-0.017	0.161**	0.098	-0.045	-0.118
	(0.067)	(0.092)	(0.057)	(0.079)	(0.064)	(0.090)	(0.085)
p-value UCT vs. CCT	0.630	0.378	0.389	0.364	0.584	0.636	0.963
p-value Treatment	0.297	0.570	0.685	0.105	0.249	0.862	0.258
Mean in Control Group	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	2,048	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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. All outcome variables are standardized. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Appendix Table C3: Sexual Behavior (Extensive Margin)

Panel A: Baseline Dropouts	Ever Had Sex			# Sexual Partners			Sexually Active Past 12		
	Round 2	Round 3	Round 4	Round 2	Round 3	Round 4	Round 2	Round 3	Round 4
=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 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. Baseline values of the following variables are included as controls in the regression analyses: age indicators, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. we correct ever had sex for discrepencies across rounds. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Appendix Table C4: Sexual Behavior (Intensive Margin)

Panel A: Baseline Dropouts	Age at First Sex			Older Partner			Condom Use	
	Round 2	Round 3	Round 4	Round 2	Round 3	Round 4	Round 3	Round 4
=1 if Conditional Schoolgirl	-0.064	-0.061	0.110				0.046	0.030
	(0.137)	(0.144)	(0.133)				(0.037)	(0.030)
Mean in Control Group	16.250	16.578	16.790				0.159	0.156
Sample Size	525	625	723				446	600
Panel B: Baseline Schoolgirls								
=1 if Conditional Schoolgirl	0.220	0.136	0.147				-0.006	0.015
	(0.146)	(0.130)	(0.146)				(0.055)	(0.041)
=1 if Unconditional Schoolgirl	-0.152	-0.039	-0.207				0.102	0.057
	(0.179)	(0.189)	(0.127)				(0.086)	(0.048)
p-value UCT vs. CCT	0.064	0.404	0.052				0.268	0.482
p-value Treatment	0.143	0.536	0.128				0.483	0.479
Mean in Control Group	15.731	16.393	17.199				0.247	0.268
Sample Size	522	893	1,494				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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. we correct ever had sex for discrepencies 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. It is missing for those who were not sexually active in the past 12 months. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Appendix Table C5: Employment Outcomes (secondary)

	Effective Daily Wage (Past 7 Days)	Labor Income (past 5 seasons)	=1 if Any Wage Work in Past 3 Months
Panel A: Baseline Dropouts			
	Round 4	**Round 4**	**Round 4**
=1 if Conditional Schoolgirl	-68.456	1,240.350	-0.020
	(44.550)	(2,589.541)	(0.037)
Mean in Control Group	226.22	15873.06	0.366
Sample Size	263	744	744
Panel B: Baseline Schoolgirls			
=1 if Conditional Schoolgirl	36.359	2,245.830	-0.010
	(127.237)	(2,242.769)	(0.030)
=1 if Unconditional Schoolgirl	-164.929*	3,210.538	0.001
	(85.559)	(3,821.432)	(0.055)
p-value UCT vs. CCT	0.278	0.829	0.838
p-value Treatment	0.121	0.420	0.939
Mean in Control Group	270.876	10,003.970	0.250
Sample Size	465	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. Baseline values of the following variables are included as controls in the regression analyses: age indicators, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Appendix Table C6: Empowerment Outcomes (Secondary)

	Super-index of Overall Empowerment	Change in Ladder from Five Years Ago to Today	Super-Index of Unmarried Empowerment	Super-Index of Married Empowerment
Panel A: Baseline Dropouts				
=1 if Conditional Schoolgirl	-0.083	-0.032	0.018	-0.130
	(0.074)	(0.232)	(0.112)	(0.098)
Mean in Control Group	0.000	1.120	0.000	0
Sample Size	744	744	289	455
Panel B: Baseline Schoolgirls				
=1 if Conditional Schoolgirl	0.049	0.276	0.111	-0.005
	(0.082)	(0.187)	(0.098)	(0.099)
=1 if Unconditional Schoolgirl	-0.159**	0.176	-0.094	-0.357**
	(0.081)	(0.190)	(0.109)	(0.173)
p-value UCT vs. CCT	0.052	0.650	0.120	0.068
p-value Treatment	0.101	0.306	0.287	0.121
Mean in Control Group	0.000	0.906	0.000	0.000
Sample Size	2,049	2,049	1,271	776

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, strata indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed in Round 4. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Appendix Table C7 Child Selection effects by Epoch:

How does treatment Alter the types of CRs having children?

Outcome: Total Live Births in R4

Baseline covariates:	Baseline Dropouts	Schoolgirl CCT vs Control	Schoolgirl UCT vs Control
	(1)	(2)	(3)
Treatment	-0.142***	0.065*	-0.026
	(0.054)	(0.036)	(0.037)
T * Age	-0.006	0.006	-0.050*
	(0.029)	(0.023)	(0.029)
T * Near-Urban	-0.038	-0.123	-0.186***
	(0.183)	(0.100)	(0.065)
T * Urban	0.160	-0.267**	-0.210
	(0.249)	(0.120)	(0.132)
T * Highest Grade	-0.026	0.026	0.090***
	(0.026)	(0.023)	(0.033)
T * Asset Index	-0.016	0.011	-0.009
	(0.026)	(0.012)	(0.021)
T * Never had Sex	-0.076	-0.043	0.017
	(0.125)	(0.098)	(0.116)
Age	0.059***	0.090***	0.090***
	(0.021)	(0.010)	(0.010)
Near-Urban	0.007	-0.004	-0.004
	(0.135)	(0.052)	(0.052)
Urban	-0.236	-0.082	-0.082
	(0.188)	(0.061)	(0.061)
Highest Grade	0.015	-0.051***	-0.051***
	(0.018)	(0.012)	(0.012)
Asset Index	-0.030	-0.028***	-0.028***
	(0.019)	(0.007)	(0.007)
Never Had Sex	-0.401***	-0.348***	-0.348***
	(0.077)	(0.049)	(0.049)
Constant	1.369***	0.563***	0.563***
	(0.039)	(0.016)	(0.016)
Number of observations	744	1,791	1,585
R-Squared	0.096	0.2	0.162
F-test: Interacted Coeffs Jointly 0	0.73	1.49	3.54
Prob > F	0.63	0.188	0.003

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. All baseline variables demeaned so that uninteracted treatment effect is at mean of all coviates.

Appendix Table C8

CHILD BEHAVIORS.

DROPOUTS

	Fine Motor Skills	Language Hearing	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Problems	Total Difficulties	Total Pro Social
	(1)	(2)	(3)	(4)	coef/se	coef/se	coef/se	coef/se
=1 if Treatment Dropout	-0.059	-0.146	0.117	0.084	-0.020	0.320*	0.167	0.149
	(0.137)	(0.113)	(0.169)	(0.191)	(0.163)	(0.167)	(0.190)	(0.164)
Number of observations	215	213	224	224	224	224	224	224

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs.

SCHOOLGIRLS

	Fine Motor Skills	Language Hearing	Emotional Symptoms	Conduct Problems	Hyperactivity	Peer Problems	Total Difficulties	Total Pro Social
	(1)	(2)	(3)	(4)	coef/se	coef/se	coef/se	coef/se
=1 if Conditional Schoolgirl	-0.296	-0.351*	-0.013	-0.001	0.150	-0.091	0.016	-0.401
	(0.185)	(0.182)	(0.176)	(0.196)	(0.161)	(0.198)	(0.188)	(0.285)
=1 if Unconditional Schoolgirl	0.417	0.252	-0.020	0.105	-0.093	-0.100	-0.028	-0.115
	(0.328)	(0.242)	(0.228)	(0.202)	(0.318)	(0.310)	(0.224)	(0.306)
Number of observations	188	185	196	196	196	196	196	196

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs.

Appendix Table C9 Husband Attrition:

Panel A: Dropouts	Currently Married (1)	Husband Data Present (2)	Husband Refused Survey (3)	Failure to Track Husband (4)
=1 if Treatment Dropout	-0.097***	0.023	-0.063*	0.040
	(0.035)	(0.046)	(0.033)	(0.037)
Number of observations	744	455	455	455
Control Group Mean	0.654	0.700	0.168	0.132

Panel B: Schoolgirls	(1)	(2)	(3)	(4)
=1 if Conditional Schoolgirl	-0.012	0.075	-0.038	-0.035
	(0.027)	(0.048)	(0.035)	(0.036)
=1 if Unconditional Schoolgirl	0.039	0.002	-0.006	0.005
	(0.046)	(0.065)	(0.044)	(0.046)
Number of observations	2,049	776	776	776
Control Group Mean	0.342	0.706	0.128	0.164
F test: CCT=UCT	1.133	1.100	0.429	0.632
p-value on F-test	0.289	0.296	0.514	0.428

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. Regressions include the standard battery of core respondent-level baseline controls.

Appendix Table C10: Husband HIV:

Panel A: Dropouts

	Entire Sample		Married		Married + Husband Data	
	HIV	HIV alternate	HIV	HIV alternate	HIV	HIV alternate
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Treatment Dropout	-0.019	0.005	-0.021	0.005	-0.005	0.018
	(0.029)	(0.032)	(0.029)	(0.032)	(0.035)	(0.038)
Number of observations	350	355	328	333	265	269
Control Group Mean	0.074	0.079	0.072	0.077	0.055	0.062

Panel B: Schoolgirls	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	0.000	-0.001	-0.005	-0.006	0.001	-0.001
	(0.026)	(0.026)	(0.028)	(0.028)	(0.033)	(0.033)
=1 if Unconditional Schoolgirl	0.023	0.022	0.024	0.022	0.010	0.008
	(0.036)	(0.036)	(0.038)	(0.038)	(0.041)	(0.041)
Number of observations	614	615	567	568	457	458
Control Group Mean	0.045	0.046	0.046	0.048	0.052	0.054
F test: CCT=UCT	0.330	0.320	0.460	0.450	0.040	0.030
p-value on F-test	0.567	0.538	0.497	0.505	0.845	0.854

note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. Regressions include the standard battery of core respondent-level baseline controls.

Appendix Table C11 Husband Attitudes towards Gender Empowerment:

Panel A: Dropouts

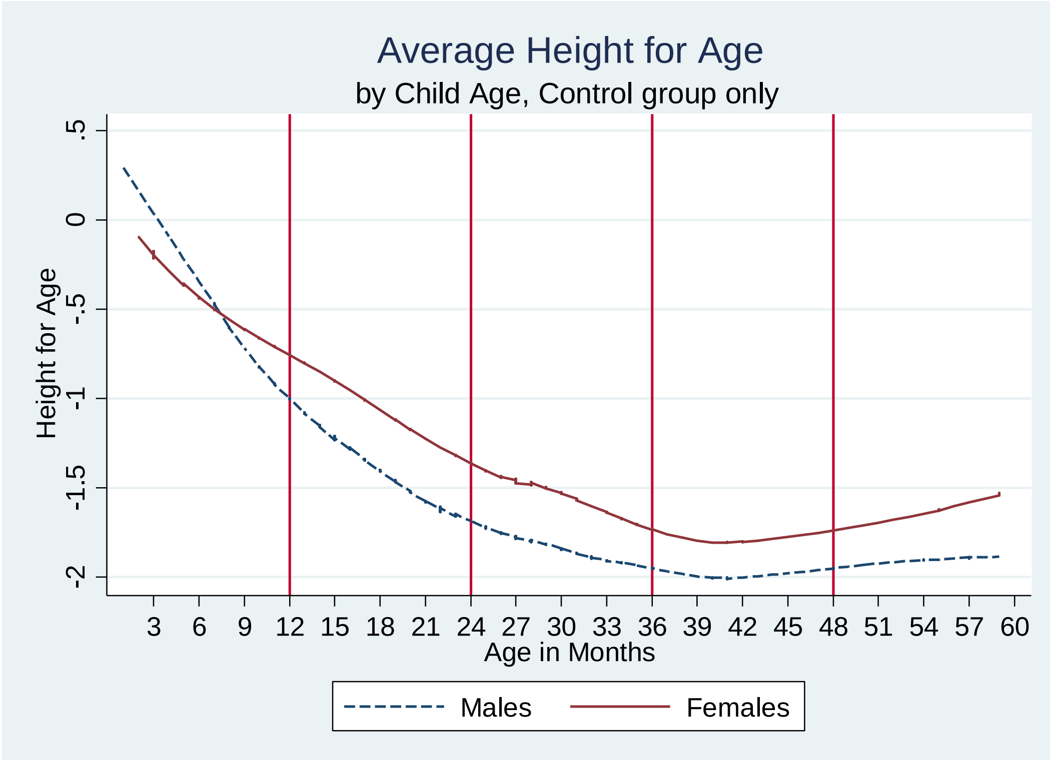
	Super Index of Gender Empowerment	Index of Gender Empowerment	Index of Wife Autonomy	Index of Non Abuse	Index of Divorce Prospects for Wife	Desired Fertility (negative of, standardized)
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Treatment Dropout	0.145	-0.000	0.189	0.162*	-0.162	0.155
	(0.100)	(0.117)	(0.129)	(0.091)	(0.129)	(0.133)
Number of observations	326	326	326	325	325	324
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000

Panel B: Schoolgirls	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	0.069	0.013	0.125	0.123	-0.048	-0.052
	(0.108)	(0.095)	(0.117)	(0.103)	(0.167)	(0.123)
=1 if Unconditional Schoolgirl	0.254	-0.315*	0.462	0.175	0.171	0.069
	(0.199)	(0.183)	(0.389)	(0.109)	(0.123)	(0.210)
Number of observations	543	543	543	543	542	541
Control Group Mean	0.000	0.000	0.000	0.000	0.000	0.000
F test: CCT=UCT	0.796	3.161	0.738	0.167	1.323	0.299
p-value on F-test	0.374	0.078	0.392	0.683	0.252	0.586

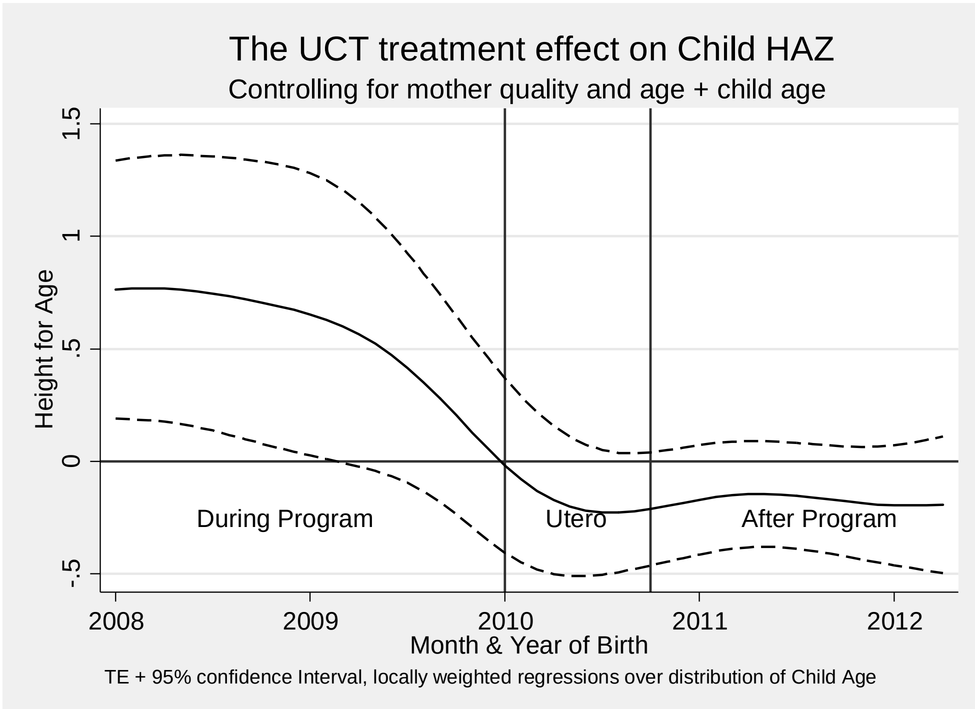
note: *** p<0.01, ** p<0.05, * p<0.1

OLS regressions with standard errors clustered at the EA level. Observations are weighted to make results representative of the target population in the study EAs. Regressions include the standard battery of core respondent-level baseline controls.

Appendix Figure C1. HAZ by age in the Control Group



Appendix Figure C2. Raw Treatment Effects by Month of Birth.



Appendix Figure C3. Treatment Effects by Month of Birth, Controlling for Confounds