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

Sarah Baird
George Washington University

Craig McIntosh
University of California, San Diego

Berk Özler
The World Bank

June 13, 2016

Abstract

We examine the medium-term effects of a two-year cash transfer program for adolescent girls on a broad range of outcomes. Just two years after the cessation of transfers, the substantial short-term impacts of unconditional cash transfers (UCTs) completely disappeared: in fact, UCT recipients were less empowered and had worse marriage market outcomes than the control group. However, children born to UCT beneficiaries during the program had significantly higher height-for-age z-scores. On the other hand, for those who had already dropped out of school at baseline, conditional cash transfers (CCTs) produced a large increase in educational attainment, a sustained reduction in the total number of births, and a more educated pool of husbands. Even in this group, however, there were no gains in health, labor market outcomes, or empowerment. These findings suggest that while UCTs provide adequate safety nets for families during their eligibility periods and CCTs can effectively contribute to sustained human capital accumulation, the current enthusiasm for the promise of cash transfer programs as a tool for transformative changes for their direct beneficiaries may be misguided.

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

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

---

* Corresponding author: Berk Özler, 1818 H St. NW, Washington, DC 20010; phone: (202)-458-5861; email: bozler@worldbank.org. We thank seminar and conference participants at the Center for the Study of African Economies, Columbia University, Middlebury College, Monash Development Workshop, Oregon State University, Otago Development Workshop, Otago International Health Research Network, Labor Econometrics Workshop, PopPov Annual Research Conference, University of Maryland, University of Oklahoma, University of Oregon, University of Southern California, Washington Area Development Economics Symposium and the World Bank. We thank everyone who provided this project with great fieldwork and research assistance and are too numerous to list individually. We gratefully acknowledge funding from the Global Development Network, the Bill and Melinda Gates Foundation, 3ie, NBER Africa Project, and the World Bank. Ethical review committees at the National Health Sciences Research Council (Malawi, Protocol #569) the University of California at San Diego (USA, Protocol #090378), and George Washington University (USA, Protocol #061037) approved the study design. The trial is registered at AEA RCT Registry (#AEARCTR-0000036). The findings, interpretations, and conclusions expressed in this article are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development or the World Bank.

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, even if the poor receive these transfers with few 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 cash transfer programs become major vehicles for social protection, it is increasingly important to understand what happens after beneficiaries "graduate:" can these programs actually help individuals transition out of poverty, or do the benefits evaporate when the money runs out?

For any intervention with a fixed duration to have a sustained effect, it needs to lead to an increase in the stock of some asset that produces a stream of returns in the future. For example, a program that provides cash grants to groups of unemployed youth for income generating activities may have lasting effects on earnings through the accumulation of physical (productive assets) and human (vocational skills) capital (Blattman, Fiala and Martinez 2014). Alternatively, large unconditional cash grants to poor households may increase future earnings by increasing

[1] See, for example, (Banerjee et al. 2015) and (Bandiera et al. 2016) 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, Jamison and Sheridan 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, Martinez 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 2011) 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 Martinez 2014) and (De Mel, McKenzie and Woodruff 2012) report impacts four to six years after the initial grants; (Banerjee et al. 2015) and (Bandiera et al. 2016) report impacts one and two years after the cessation of all support, respectively; (Haushofer and Shapiro 2016) report outcomes, on average, less than a year after cash transfers. A distinct and mostly U.S.-based literature, largely using quasi-experimental methods, has examined the very long-term effects of being exposed to cash, 'near cash,' or other safety net programs during childhood (e.g. (Aizer et al. 2016); (Currie and Almond 2011); (Hoynes, Schanzenbach and Almond 2016)) and has demonstrated beneficial effects on a host of outcomes as adults.

investments in productive assets, such as livestock (Haushofer and Shapiro 2016); while small monthly conditional cash transfers over a long period of time may lead to increased consumption after beneficiaries exit the program by increasing savings and investments in small-scale agriculture (Gertler, Martinez and Rubio-Codina 2012) or by stimulating entrepreneurial activity (Bianchi and Bobba 2013). For programs targeting younger people, the causal pathway to improved welfare over the long run is more likely to be human capital accumulation, either in the form of education and skills or health – especially reproductive and sexual health for adolescent females.

In this paper we report the effects of a cash-transfer experiment more than two years after it ended, tracking a broad range of outcomes for females aged 18-27. Our earlier work has demonstrated the short-term effectiveness of cash transfers in improving school attendance and test scores as well as reducing the incidence of pregnancy, marriage, and sexually transmitted infections during adolescence, indicating the possibility of finding longer-term improvements in the wellbeing of the adolescent beneficiaries as young adults ((Baird, McIntosh and Özler 2011); (Baird et al. 2012)).[3] Here, we first look at human capital accumulation, marriage and fertility, labor market outcomes, and empowerment among the beneficiaries to assess the persistence of the short-term effects in the context of a cluster-randomized controlled trial (RCT) that provided conditional and unconditional cash transfers. Then, as the majority of the study participants were

[3] A recent review of interventions targeted at adolescents in low- and middle-income countries generally indicate positive, albeit modest effects on childbearing-related outcomes (McQueston, Silverman and Glassman 2013). The evaluation of a school-based intervention in Kenya testing the effects of education subsidies found significant reductions in school dropout, pregnancy, and marriage among girls in the short- and medium-run, and school attainment, marriage, and childbearing by age 16 in the longer-run (Duflo, Dupas and Kremer 2015). 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. 2015). 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, Schady and Ferreira 2009); (Saavedra and Garcia 2013)).

married and/or had children, we examine their marriage market outcomes and their children's physical development using data we collected on their husbands and anthropometric measurements of their children. These rich data on the beneficiaries and their new families give us a comprehensive picture of their lives as young adults and allow us to contrast the medium-term impacts of conditional and unconditional cash transfers.

We find that the short-term improvements in the UCT arm observed during and at the end of the program failed to translate into increased welfare in the longer-run. Substantial reductions in teen marriages, total live births, and HIV infections, as well as improvements in psychological wellbeing and nutritional intake observed at the end of the program, completely disappeared two years after the end of the intervention. In this group, the end of the cash transfer program was immediately followed by a marriage and baby boom among the beneficiaries, who reported lower levels of empowerment and had husbands with lower cognitive ability compared with both the CCT and the control groups. However, consistent with improved health and nutrition during the program, we find evidence of improved physical development – measured by height-for-age z-scores – among children born to the beneficiaries during the program.

CCTs, on the other hand, caused sustained effects on school attainment, incidence of marriage and pregnancy, age at first birth, total number of births, and desired fertility – but only among those who had already dropped out of school at baseline. This result stems from the fact that the CCTs were highly effective in allowing a very large share of this group to return to school. In contrast with the marital outcomes in the UCT group, the increased educational attainment in this group led to assortative matching: their husbands were significantly more likely to have completed secondary school. However, even in this stratum, there were no improvements in important longer-term outcomes, such as individual earnings, per capita household consumption,

subjective wellbeing, health, or empowerment. Among those who were in school at baseline, CCTs did not have any lasting effects, positive or negative, mainly because the transfers were inframarginal with respect to school attainment: 88% of the control group in this stratum had completed primary school two years after the end of the program.

The lack of persistent effects two years after the end of the program is cause for concern regarding the promise of cash transfers for sustained improvements in wellbeing. With respect to UCTs, there is no doubt that they are very valuable as a tool for social protection, with beneficial effects on a wide range of outcomes while the transfers are in place. Money is very effective in alleviating a multitude of problems that adolescent females face in this context, but we cannot safely transition them off cash transfers – at least not after a two-year program – because they lead to no discernible capital accumulation of any type.[4] Ultimately, our evidence suggests that once the money runs out, the reversion to the control group levels is fast for almost all outcomes.

With respect to CCTs, the story is more nuanced but still not very promising. The raison d'etre of schooling CCTs is the link between increased human capital accumulation and future welfare. Among baseline schoolgirls, the largely inframarginal transfers led to modest gains in enrollment and test scores in the short-run that did not translate into higher attainment or skills. Among baseline dropouts, who came from more vulnerable households, there were large and lasting effects on school attainment, which had knock-on effects on desired and actual fertility as well as husband quality but did not translate into medium-term improvements in other domains.

[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, for example, in (Gertler, Martinez and Rubio-Codina 2012).

Several reasons might explain the disconnect between increased school attainment and no improvements in labor market outcomes, empowerment, or health. First, it is possible that increased schooling does not provide one with the skills needed to increase future welfare in this context. There are very few formal sector jobs for women in Malawi and most households depend on subsistence farming and a variety of informal sector activities. We measured competencies that relate to skills needed in farming and running small household enterprises and detected no effects in these domains. If safe and well-paying jobs existed for women in Malawi, households might invest in the necessary human capital of adolescent females on their own – perhaps even without the help of any outside interventions ((Heath and Mobarak 2015); (Jensen 2012); (Munshi and Rosenzweig 2006); (Oster and Steinberg 2013)). Second, task performance is dependent on not only improvements in cognitive skills, but also on character skills and effort (Heckman and Kautz 2013). Hence, it is possible that CCTs, by providing incentives for formal schooling, improved only cognitive skills, which may not have been sufficient.[5] In this sense, our results are pessimistic about the longer-term effects of formal schooling in this context.

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) and "…one of the more effective ways to improve children's long term outcomes might be to target women of child bearing age in addition to focusing on children after birth" (Currie and Almond 2011). As well-known channels for growth, such as maternal nutrition and stress, improved during the program, there is reason to think that children of beneficiaries who were exposed to cash transfers

[5] (Heckman and Mosso 2014) states "The most effective adolescent interventions target formation of personality, socioemotional, and character skills through mentoring and guidance, including providing information." (Bandiera et al. 2015) provides suggestive evidence that a mentoring program in Uganda (ELA) that provided young females with "hard" vocational and "soft" life skills may have led to longer-term improvements in welfare.

would also reap benefits. In fact, this is what we find: children of UCT recipients born during the program were substantially taller for their age than children in both the control and the CCT group two years after the program ended.[6] Does this finding arise from a treatment effect on the composition of mothers and children or the direct effect of cash transfers on a given composition of children? In exploratory analysis, using inverse propensity weighting and regression controls, we suggest that this impact on child height arises in roughly equal proportion from composition and direct treatment effects. Comparing our findings with other studies of cash transfer programs reporting effects on early childhood development, we suggest that income support for prospective mothers may be largely responsible for the observed effects on height-for-age z-scores ((Aguero, Carter and Woolard 2006); (Barham, Macours and Maluccio 2013); (Fernald, Gertler and Neufeld 2009); (Gertler 2004); (Macours, Schady and Vakis 2012)).

Our paper contributes to a number of distinct literatures. First, it adds to a growing literature that speaks to the medium- to long-term effects of cash transfer programs in developing countries: our finding that CCT programs can substantially increase school attainment among vulnerable populations without significant effects on test scores, cognitive skills, employment, or earnings is consistent with evaluations of CCT or scholarship programs in other developing countries (Baez and Camacho 2011); (Filmer and Schady 2014); (Molina-Millan et al. 2016)).[7] Other studies of cash transfers and skills training programs targeted to adolescent females have also reported significant reductions in teen pregnancies and marriages (Bandiera et al. 2015) and improvements in mental health and food intake (Haushofer and Shapiro 2016) in the short-run. Our longer-term

[6] Unlike UCT recipients, CCT recipients, 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. In this group, the potential benefits to child development may come from increased schooling of the mothers for children born after the program.

[7] (Barrera-Osorio, Linden and Saavedra 2015) also show longer-term effects of a CCT program in Colombia, but do not examine non-education outcomes. Our study also complements two other experiments that compare the relative effects of CCT and UCT programs in the short-run ((Akresh, De Walque and Kazianga 2013); (Benhassine et al. 2015)).

findings show a lack of persistence for such outcomes after the cessation of support and provide a note of caution for the current optimism about the promise of these programs for future welfare gains among their beneficiaries.

Second, our study conributes to the literature on the effects of human capital accumulation and increased age at marriage on marriage market outcomes ((Anderson and Bidner 2015); (Ashraf et al. 2014); (Field and Ambrus 2008)). Theory suggests that these two factors affect spousal quality in opposite directions with, *ceteris paribus*, increased education improving marital outcomes while delaying marriage worsening them. Our findings confirm these predictions: CCT beneficiaries who delayed marriage while substantially increasing their school attainment have more educated husbands, while UCT beneficiaries who similarly delayed marriage, but did not accumulate any additional schooling, are less empowered and have husbands with lower cognitive ability.

Third, our study speaks to a relatively large literature on the effects of programs that support families with pregnant women and young children. There is a sizeable and growing literature on the long-term effects of cash and "near cash" transfer programs in the U.S. ((Aizer et al. 2016); (Currie and Almond 2011); (Hoynes, Schanzenbach and Almond 2016)) and in developing countries (see, e.g., (Molina-Millan et al. 2016) for a recent review). Our study helps fill a gap in the literature by constrasting the effects of CCTs and UCTs for the wellbeing of adolescent girls during and after their transition to early adulthood and the development of their own children born during this period.[8] Moreover, because our program targeted adolescent girls of childbearing age, our findings provide evidence on the important policy question of how to time interventions to protect early childhood development.

[8] (Aizer et al. 2016) states: "Whether cash transfers are more or less cost-effective than in-kind transfers or conditional cash transfers is an important question for future research."

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

2. Study Setting, Design, and Data Sources

2.1 Study Setting

The 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 – only 17.0% of our control group had completed secondary school as of 2012. Although most adults 15 and over participate in some form of employment, the majority do not receive a formal income. In 2008, only 6% of the adult population received a formal income (Zomba City Assembly 2009), a number that is likely lower for females.

2.2 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 already out of school at baseline (baseline dropouts) and those who were still 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 enumeration area (EA) level; 88 to treatment and 88 to control. All baseline dropouts in treatment EAs received conditional cash transfers (CCTs), while a further experiment was performed within the larger cohort of baseline schoolgirls. For them, 46 EAs were assigned to CCTs, 27 were assigned to unconditional cash transfers (UCTs), and 15 were assigned to receive no transfers in order to study spillovers. The amount of money received by the household head was randomized between $4 and $10 at the EA level, and the core respondents were assigned their own individual transfer amounts 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 I presents an illustration of the study design, and a more detailed description of the experiment can be found in (Baird, McIntosh and Özler 2011).

Girls receiving UCTs simply had to show up at a local distribution point each month to pick up their transfers. Monthly school attendance for all girls in the CCT arm was checked and payment for the following month was withheld for any student whose attendance was below 80% of the number of days school was in session for the previous month. However, participants were never removed from the program for failing to meet the monthly 80% attendance rate, meaning that if they subsequently had satisfactory attendance, then their payments would resume. Other design aspects of the intervention were kept identical so as to be able to isolate the marginal effect of imposing a schooling conditionality on primary outcomes of interest.[9]

[9] 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.

2.3 Data Sources and Outcomes

Data Sources. The focus of this paper is data collected in Round 4, which took place in 2012, more than two years after the end of the intervention. However, to provide context to these results, we also present impacts on the same outcomes, when available, for data collected during Rounds 1 through 3. Focusing on the core respondent, the data sources include household surveys (all rounds), biomarker data collection on HIV (Round 2-4) and Anemia (Round 4), and competencies (Round 4). In Round 4, data collection also included anthropometric data (children under 60 months of age) and early child development tests (children 36-59 months old) for children of core respondents and a survey and biomarker data collection among husbands of married core respondents.

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

The Round 4 household survey also consisted of a set of questions to try and measuring basic labor market skills of the core respondent, termed "competencies." They included reading and following instructions to apply fertilizer; making correct change during a hypothetical market

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 reading comprehension, math, and cognitive skills tests utilized in Round 3, and serve as a measure of a more practical set of skills that might be influenced by increased schooling and needed in the labor market.

Home-based voluntary counseling and testing for HIV (for core respondents 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 aged 59 months or younger.

Early childhood development (ECD) tests were administered for all 36-59 month-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 (Gladstone et al. 2008) and the Strengths and Difficulties Test (SDQ), which was administered to the core respondent or the guardian responsible for the child ((Goodman 2001); (Woerner et al. 2004a).

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 the 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 outcomes in

these domains for their husbands and children.[10] We briefly discuss these outcomes below. A detailed description of all outcomes reported in this paper is provided in Appendix A.

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

Marriage and Fertility. Our primary outcomes include self-reported data on whether or not the core respondent was ever married or ever pregnant. We also examine age at first marriage and at first birth,[11] as well as total live births. Desired fertility is a secondary outcome.

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

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

[10] 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 scores always have a consistent meaning (good or bad). We then calculated the mean and standard deviation of the responses to each sub-question in the control group – separately for baseline schoolgirls and baseline dropouts. We then 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.

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

whether she had a sexual partner five or more years older, and her condom use during her most recent sexual intercourse.

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

Employment and Wages. In this domain, we examine the proportion of hours spent in self-employment or paid work, the typical wage rate for work done in the past three months, and the opportunity cost of time which is constructed by asking the core respondents a series of hypothetical questions regarding whether they would accept employment at a given wage rate. Secondary outcomes include whether the core respondent participated in any wage work in the past three months, labor income in the past five seasons, and an effective daily wage rate for work done in the past seven days. Per capita consumption is reported as a secondary outcome.

Husbands. Our analysis of husbands focuses on spousal quality and their attitudes towards women's empowerment. The husband quality index, the first primary outcome, includes sub-components that measure the husband's highest grade completed and highest qualification obtained, his cognitive score on the Raven's Colored Progressive Matrices, his employment status and wage, his HIV status, his marital fidelity (self-reported), and his mental health measured through the GHQ-12. The index of spousal attitudes towards women's empowerment, also a primary outcome, includes sub-indices for attitudes towards their daughters' schooling and

marriage, their wives' autonomy, domestic violence, as well as their divorce prospects and desired fertility levels. Components of these two indices are presented as secondary outcomes.

Children. The primary child outcomes fall under four domains: anthropometrics, health, parental practices and educational testing. For anthropometrics, we construct height-for-age z-score (HAZ) for living children younger than 60 months old. Our health outcomes include neonatal and post-neonatal mortality. For parental practices, we construct variables for exclusive breastfeeding in the past six months and an index of parenting practices. Finally, for educational testing we report MDAT and SDQ scores for all 36-59 month-olds.[12] The self-reported data on children come from complete birth and death histories collected at Rounds 3 and 4 from the mother or the primary caregiver.

3. ESTIMATION STRATEGY

In this section, we discuss the experimental estimation strategy used to examine program impacts on core respondents. The causal identification of program impacts on husband characteristics and children's outcomes is more challenging and the estimation strategies used to analyze those outcomes are discussed in Sections 4.5 and 4.6.

The evaluation of the impact of the Zomba Cash Transfer Program utilizes the experimental design of the intervention for causal identification. To estimate intention-to-treat effects of the program in each treatment arm on our primary outcomes by stratum, we employ a simple reduced-form linear model:

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

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

where Y_{ic} is an outcome variable for core-respondent i in cluster c, T_c^C and T_c^U are binary indicators for offers in the CCT and the UCT clusters, respectively, and X_{ic} is a vector of baseline characteristics. Note that for baseline dropouts we only have the CCT binary indicator. The standard errors ε_{ic} are clustered at the EA level, which account for both the design effect of our EA-level treatment and the heteroskedasticity inherent in the linear probability model.

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

Appendix Table S1 examines attrition in the core respondent sample first for baseline dropouts, then baseline schoolgirls. Our attrition analysis focuses on whether the core respondent is part of the Round 4 household survey sample, the main data source used in our analysis.[13] Attrition two years after the end of the cash transfer program is 15.7% among *baseline dropouts* and there is no difference in levels between the CCT and control groups (column 1), but CCT girls in urban areas were more likely to be lost to follow-up (column 2).[14] Attrition among the younger group of *baseline schoolgirls* is slightly lower at 12.5% and column 3 shows that it was

[13] To maximize sample size and statistical precision in our longer-term analysis, we use the sample interviewed in Round 4. Earlier impact findings from Rounds 2 and 3, including baseline balance, replicate in this sample.

[14] In the tables that follow, we do not make any corrections for differential attrition in the stratum of baseline dropouts, within which we show large, significant, and sustained effects on educational attainment, marriage, and fertility. All impact findings on primary outcomes among beneficiaries in this group are robust to correcting for selection on observables using inverse propensity weights to account for differential attrition between the CCT and control groups.

significantly lower in both treatment arms. However, joint F-tests for interactions of treatment with baseline characteristics are not significant at the 90% level. Furthermore, there is no differential attrition between the CCT and UCT arms – either in levels or by characteristics. Appendix Table S2 provides baseline descriptive statistics and shows balance.

4. RESULTS

We start by presenting the trajectory of program effects on outcomes in four domains, separately for baseline dropouts and baseline schoolgirls: education and competencies, marriage and fertility, health, and, finally, labor market participation and empowerment.[15]

4.1 Education and Competencies

Table I presents program impacts on highest grade completed and competencies. Among *baseline dropouts,* CCTs led to an increase in highest grade completed of approximately 0.6 years, which stayed more or less constant over time (Panel A), and represents a 0.22 standard deviation (SD) increase by Round 4. As a result, the share of beneficiaries with a Primary School Leaving Certificate (PSLC) increased by 5.8 and 8.1 percentage points in Rounds 3 and 4, respectively (Appendix Table S3, Panel A). However, earlier gains in test scores of English reading comprehension, mathematics, and cognitive skills (columns 4-7) did not translate into increased scores in tests of basic labor market skills, or "competencies," such as following instructions to apply fertilizer or calculating change in a market transaction (column 8).

The results for *baseline schoolgirls* suggest little, if any effects, on these outcomes for either treatment group (Table I, Panel B). Any significant effect in the CCT group at the end of the program was no longer detectable two years later. The reader should note that the mean number

[15] The reader should note that most of the one- and two-year impacts during and at the end of the program were reported in previous publications, which are clearly cited throughout the paper. What are new here are the findings from two years after the end of the program. Presenting program impacts over time within each domain allows the reader to examine the trajectory of program effects and assess whether earlier impacts were sustained.

of years completed in the control group is 10.4 in Round 4, at which point 88% of the control group had obtained a PSLC (Appendix Table S3, Panel B). Hence, close to 90% of the transfers to baseline schoolgirls were inframarginal with respect to primary school completion. Still, the cash transfer program did not cause any significant gains in secondary school completion either. Similarly, earlier gains in test scores in the CCT group did not translate into improved competencies in the longer-run, with the only significant improvement seen in the UCT group for the ability to send a simple text message using a mobile phone.

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

4.2 Marriage and Fertility

As with the education outcomes, CCTs had large effects on marriage and fertility for *baseline dropouts* that were sustained two years after cessation of transfers (Table II, Panel A). Baseline dropouts were 14.0, 15.7, and 10.7 percentage points less likely to have been ever married

during, at the end, and two years after the program, respectively (all significant at 99% confidence). The corresponding reductions were 5.7, 8.1, and 4.0 percentage points for being ever pregnant (all significant the 90% confidence or higher). Furthermore, there is a negative fertility gradient among CCT beneficiaries, leading to a reduction of 0.147 total live births at Round 4 (p-value < 0.001), which corresponds to a reduction of more than 10% and is consistent with the reduction in desired fertility. Age at first marriage and first birth were similarly higher by 0.43 and 0.27 years, respectively.

Among *baseline schoolgirls*, CCTs had no effects on marriage and fertility at any point during our study period (Table II, Panel B). On the other hand, UCTs were very effective in substantially reducing marriage and pregnancy rates among baseline schoolgirls during and at the end of the program (Baird, McIntosh and Özler 2011). However, we see an almost complete reversal of these outcomes at the four-year follow-up: there are no longer any differences in ever married, ever pregnant, total number of live births, or even age at first birth between the UCT group and either 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. Striking spikes in pregnancies and marriages in the UCT group immediately following the end of the transfers are shown in Figure II. The temporary nature of the fertility changes in this group is also reinforced by the fact that desired fertility remains unchanged (Table II, Panel B, column 12).

Cash transfers can have effects on marriage and fertility via two channels. The first pathway, apparent in the UCT arm, is through an income effect. The income effect is strong but it disappears immediately when the transfers stop – if the transfers have not led to some accumulation of physical or human capital. The other pathway, apparent in the CCT arm among baseline

dropouts, is through increased schooling. Increased schooling is strongly associated with delays in marriage and fertility and reductions in desired and total fertility, but the impacts of transfer programs on schooling have to be substantial to translate into meaningful and statistically significant knock-on effects on marriage and fertility.

4.3 Health

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

Among *baseline schoolgirls*, program impacts on HIV mirror those on marriage and fertility over time: there is no effect of CCTs on HIV at Rounds 3 or 4, but a 50% reduction in HIV prevalence in the UCT group at the end of the intervention dissipated two years later (Table III, Panel B). During the two-year post-intervention period, which saw a spike in pregnancies and marriage in the UCT group, the incidence of HIV was 3.5 percentage points (pp) – compared with 2.0 pp in the control group. Appendix Table S6 shows that effects of cash transfers were equally

transient on mental health and nutritional intake – strongly evident during the program and disappearing afterwards. There is weak evidence of a lower prevalence of anemia in the UCT group in Round 4, but the UCT effect on a continuous measure of hemoglobin levels does not support this finding. The evidence suggests that income effects among baseline schoolgirls were effective in causing temporary changes in wellbeing (lower risk of sexually transmitted infections, improved mental health, and increased food consumption), but were insufficient to put the beneficiaries on a different trajectory.

4.4 Labor Market Participation and Empowerment

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

For *baseline dropouts*, program impacts on empowerment echo those on competencies, health and labor market participation: despite significant gains in educational attainment, delays

in marriage and pregnancy, and reductions in total live births, there are no effects on the index of empowerment or on subjective welfare (Table IV, Panel A). This finding holds when we examine empowerment among those married and unmarried at Round 4 separately. Appendix Table S8 shows results by the components of the female empowerment index (self-esteem, social participation, preferences for child education, and aspirations).

For *baseline schoolgirls* in the CCT group we also see no significant impacts on empowerment or subjective wellbeing, although the coefficient estimates are generally positive. However, for UCTs, the empowerment index is significantly lower than both the control and the CCT groups (Table IV, Panel B). The -0.159 SD effect (p-value=0.05) on the super-index of overall empowerment among the UCT beneficiaries is reflected in the negative (but insignificant) effects in all sub-indices except aspirations (Appendix Table S8, Panel B), and is driven mainly by a large (-0.342 SD) and significant effect on empowerment among those who are married (Table IV, Panel B, column 7). The findings indicate a statistically significant divergence in female empowerment between CCT and UCT recipients among baseline schoolgirls two years after the end of the cash transfer program. We further explore these negative impacts on marital empowerment by directly studying husband characteristics below.

4.5 Husband Characteristics

The program impacts on empowerment presented above, particularly the negative effects apparent in the UCT group, motivate the examination of marriage market outcomes. As described earlier, two years after the end of the transfer program, CCT beneficiaries among baseline dropouts were less likely to be ever married or pregnant, had a smaller number of children, and were older at first marriage and pregnancy – knock-on effects of substantive increases in highest grade completed (0.62 years) and primary school completion (8.1 pp). While these gains did not translate

into increased empowerment or subjective wellbeing in this group, the program might have nonetheless caused changes in the characteristics of their spouses.

Table V presents the treatment-control comparison of husband characteristics. For *baseline dropouts*, the evidence is consistent with assortative matching (Panel A): husbands of CCT beneficiaries have completed 0.56 years more of schooling (p-value=0.11) and are 7.4 pp more likely to have successfully completed secondary school (p-value=0.05). By inducing large numbers of dropouts to return to school, CCTs drove them to marry husbands who are more educated than those they would have otherwise. This finding appears to be driven by the same types of women marrying more educated husbands, rather than differential selection into marriage.[16] These spouses, however, are not different in terms of labor market outcomes, cognitive ability, marital fidelity, mental health, HIV (Table V), or attitudes towards women's empowerment (Appendix Table S9).

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

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

control group (p-value=0.03). The differences between the CCT and UCT groups for the overall husband quality index, as well as MSCE and cognitive ability, are all statistically significant.[17]

The divergence in these marriage market outcomes between CCT and UCT recipients can be explained by program impacts on education and the timing of childbearing and marriage. Environments in which adolescent marriage is common may feature a preference for young brides (Foster and Khan 2000), and hence delaying marriage may worsen marriage prospects, resulting in either lower husband quality (or bride price) or higher dowry payments (Field and Ambrus 2008). However, potentially counteracting this effect of increased age at marriage is human capital accumulation: for example, (Ashraf et al. 2014) show that higher female education is associated with a higher bride price in Indonesia and Zambia. While bride price is uncommon in Zomba, Malawi (the setting for our study), it is likely that higher education is rewarded in the marriage market in other ways, such as husband quality. These factors lead to a tradeoff between increased age at marriage and higher education, which jointly determine husband quality in the absence of bride prices as a market clearing mechanism (Anderson and Bidner 2015).[18]

Among baseline dropouts, CCT recipients faced exactly this tradeoff and the evidence suggests that, by and large, they improved their marriage outcomes as a result of staying in school and delaying marriage. However, there was no such tradeoff for UCT beneficiaries: the temporary delays in marriage and pregnancy in this group were due to income effects and not accompanied by gains in educational attainment. As a result, they ended up with lower quality spouses (with respect to their education and cognitive ability) and have lower levels of empowerment than both

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

the control group and CCT beneficiaries. An examination of Figure II, which shows the relative timings of births and marriages in Panels A and B, respectively, suggests that a large share of these unions may have been shotgun marriages – forced by pregnancies: the large "baby boom" apparent in the UCT group 10-12 months after the end of the cash transfer program, indicating a spike in pregnancies immediately after the cessation of financial support, is preceded by a similarly-sized "marriage boom" only a few months earlier. Thus, consistent with the broader literature, it appears that the UCT group ended up with worse marriage market outcomes as a result of delaying childbearing and marriage without accumulating additional schooling.

4.6 Child Outcomes

Policies for child development often target the first 1,000 days – from conception to the second birthday (Barham, Macours and Maluccio 2013), a period during which improvements in family income may be particularly important for children's development.[19] In our experiment, more than 2,000 babies were born to study participants by Round 4 – with endogenous variation in duration of exposure to the cash transfer program. Give the fact that well-known channels for growth, such as maternal nutrition and stress (Black, Devereux and Salvanes 2016), improved during the two-year program and that there were significant changes in fertility patterns, we conclude this section with an examination of program effects on child outcomes.

As with the husband characteristics, we begin by presenting simple treatment-control comparisons for primary child outcomes. These comparisons, presented in Table VI, appear to

[19] (Aguero, Carter and Woolard 2006) study the effect of Child Support Grants in South Africa for children who were exposed to the program up to three years after birth and find sizeable effects of increased exposure to these unconditional cash transfers on child height. (Milligan and Stabile 2009), studying child benefits in Canada, find effects on cognitive and socio-emotional skills of children aged 4-6. (Dahl and Lochner 2012) using the variation in Earned Income Tax Credit in the U.S., find that increased income improves children's test scores. (Currie and Almond 2011) review the effects of "near cash" programs, such as food stamps, in the U.S. and find credible evidence of effects on birth weight. Finally, (Aizer et al. 2016) and (Hoynes, Schanzenbach and Almond 2016) find that children whose parents received cash transfers and food stamps in the U.S. had improved education, health, and income as adults.

show few significant differences; none among CCT children among *baseline dropouts* and only one (out of eight outcomes) among *baseline schoolgirls*. In the UCT group, we observe a significantly higher prevalence of exclusive breastfeeding and better parenting practices, with no significant differences between the UCT and CCT treatment arms.

These differences, however, are not interpretable as causal program impacts as childbearing itself is endogenous to the treatment. In addition, these simple differences ignore the timing of births in relation to the program. We would expect substantial heterogeneity of program impacts on child outcomes both by when the birth took place and whether the transfers to the mother were conditional on school attendance. As in other countries in the region, fertility and schooling are mutually exclusive in Malawi ((Baird, McIntosh and Özler 2011); (Ozier 2015)), meaning that the condition to attend school effectively screens out new mothers in the CCT arm: only in the UCT arm would mothers with newborn children continue receiving transfers. Secondly, even in the UCT arm, a child conceived after the end of the program would have had no direct exposure to the program and, as we have shown earlier, the average mother would have acquired no additional education that could provide subsequent human capital-driven benefits. On the other hand, increased mother's education can, for example, increase child height (Thomas, Strauss and Henriques 1991), so we might expect to see benefits among children born after the program in the CCT groups – particularly among baseline dropouts. These causal chains suggest that UCT benefits should be concentrated among children born or in utero during the program, while CCTs might be most beneficial to children born after the mother's additional human capital accumulation took place.[20]

[20] Increased age at first birth can also have positive effects on child height through improved gynecological maturity and decreased competition for nutrition between the mother and the child in utero, which could operate in both treatment groups that delayed pregnancies.

To investigate how differential exposure to CCTs and UCTs drives treatment effects, we consider the sample of children born during three epochs. The first epoch captures those *directly exposed* to the program, meaning those born during the program.[21] This cohort is exposed for a maximum of two years, with some combination of in utero and child exposure depending on the exact birth date of the child. The second epoch covers those born within nine months of the end of the program, who were *exposed in utero* for a maximum of nine months. Finally, the third epoch covers those born more than nine months after the end of the program, who were *not exposed* to cash transfers either as children or in utero and could only benefit from the program due to improved outcomes of their mothers**.** We concentrate this analysis on height-for-age z-scores (HAZ), which is an objectively measured indicator of stunting that affects almost 50% of children under the age of five in Malawi, and is a strong predictor of productivity as an adult in low income settings (LaFave and Thomas 2016).[22]

Figure III plots the "raw" differences between the HAZ of children under 60 months in the treatment and the control group.[23] The figures are consistent with the hypothesis that differences in children's heights between treatment and control are moderated by exposure to the program. Most strikingly, we see a very large difference in HAZ between the UCT and the control group during the program, which steadily declines, disappears at the end of the program, and even turns

[21] The percentage of baseline schoolgirls who reported having been ever pregnant was less than 2% at baseline. Hence, children directly exposed to the program in this stratum are almost exclusively born during the intervention. However, approximately 45% of baseline dropouts had already started childbearing at baseline. Therefore, our analysis includes children under two at the start of the program, who were at least partially exposed to cash transfers.

[22] Of the two anthropometric measures that we collected for children aged 0-59 months – height and weight – stunting (height-for-age z-score<-2) is the key indicator of malnutrition in Malawi: almost half of the children under the age of 5 were categorized as stunted in 2010, while wasting (weight-for-height z-score<-2) rates are low at 4% (Haddad et al. 2014). Child assessments (MDAT and SDQ) are also objectively measured outcomes of cognitive and socio-emotional development, but the target age group for these assessments (36-59 months) makes them unsuitable for analysis by epoch of exposure to the program because only children born during the first year of the program (less than 200 in the baseline schoolgirl stratum with less than 30 in the UCT arm) were eligible for assessment.

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

negative during the final epoch (Panel C). This pattern is consistent with the substantive but transient improvements in nutrition and mental health of the core respondents in the UCT arm. In contrast, no significant differences in child height are apparent between the CCT and the control groups during the program – also consistent with the fact that mothers of children born in this period would have likely dropped out of school as a result of their pregnancies, thus forgoing any cash transfers (Panels A and B). Column 1 in Tables VII and VIII reports the raw differences in HAZ by epoch, for baseline dropouts and baseline schoolgirls respectively, and confirms these patterns.

However, we need to interpret these differences between the treatment and control groups cautiously, because we know that the program caused changes in fertility patterns (Table II). There were significant declines in fertility in the CCT group among baseline dropouts, while UCTs caused temporary delays in childbearing among baseline schoolgirls. It is also possible that cash transfers may have caused differential selection into childbearing in both of these groups by epoch, causing different types of women to give birth during and after the program. For example, a temporary delay by UCT beneficiaries who were predicted to give birth to shorter children would generate precisely the observed HAZ pattern in Panel C of Figure III, but would not be indicative of any real benefits to children. This identification problem is endemic to all experimental studies of maternal interventions that begin prior to pregnancy and also influence fertility. The "raw" differences in child height between the treatment and control groups presented above contain a variety of potential selection effects, such as mothers' characteristics (including maternal age), father attributes, and trivial pathways such as the child's age in months. Hence, in the rest of this section, we try to describe and control for the observed extent of selection on observables to isolate direct program effects on child height. The goal is to move from a "naïve" treatment-control

comparison towards a suggestive causal estimate that answers a simple question: "do cash transfers confer a protective effect on the height of a given child?"

To use the terminology of the causal inference literature, selection of the sample of parents and children is a mediator for which we wish to control in order to isolate the direct effect of treatment on child height 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 that variable are no longer "apples-to-apples" ((Rosenbaum 1984); (Angrist and Pischke 2008)). Analysis that attempts to partial out selection effects 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); (Imai, Tingley and Yamamoto 2013)).

Here, to do this as flexibly as possible, we pursue two approaches simultaneously. First, we consider the selection into each epoch as a form of attrition, and implement standard inverse propensity weighting intended to make the observed sample of mothers be representative of the full sample of core respondents. This result is achieved by estimating a propensity to 'give birth to a child in this epoch' using a fully interacted regression with baseline covariates, treatment indicators, and (covariate * treatment) interactions. The analysis of child height is then weighted by the standard sampling weight divided by this predicted epoch- and treatment-specific fertility probability of the mother to give birth during that epoch. 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 happened to have given birth within any period.

Then, using the inverse propensity weights, we add a sequence of endogenous control variables that remove observable forms of heterogeneity via OLS regression adjustments on the intensive margin. This approach to mediation analysis is widely used in social science (Baron and Kenny 1986) and medicine (MacKinnon 1994).[24] An advantage of this approach is that it allows us to examine how treatment effect estimates change as we sequentially add controls to net out various pathways of impact in an attempt to isolate the direct treatment effect on a given sample of children.[25]

We begin the sequential set of adjustments by controlling for the same set of baseline characteristics of the mother used to estimate inverse propensity weights, except mother's age, which we separately control for later. These characteristics are predictive of child height and provide us with estimates that are "doubly robust" to these controls ((Robins and Rotnitzky 2001); (Van der Laan and Robins 2003); (Bang and Robins 2005)).[26] Next, we add controls for child age, using a flexible (cubic) polynomial in child age in months at the time of the anthropometric measurements. A program that reduces fertility (baseline dropout CCTs) or alters its timing (baseline schoolgirl UCTs) will produce a treatment distribution of child age that differs from the control, making child age a mediator for the treatment effect, but also a "bad control" in the

[24] While a more recent literature has suggested means to conduct mediation analysis using principle stratification (Jo 2008) or non-parametric estimators (Flores and Flores-Lagunes 2009), we present simple OLS regressions to make the control structure as transparent as possible.

[25] 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 practices arising from increased 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.

[26] The distinguishing feature of doubly robust estimates is that they remain asymptotically unbiased even if one of the two models (IPW and OLS with controls) is misspecified, i.e. they consistently estimate their targets if either model is true (Kang and Schafer 2007).

language of (Angrist and Pischke 2008). However, a difference in the mean age of children between the treatment and control groups can cause a spurious difference in mean HAZ.[27] Thus, child age is a "trivial" pathway and netting it out allows us to take another step towards a more meaningful comparison of children in treatment and control by estimating the average of the child age-specific treatment impacts.

A third pathway is maternal age. Even a temporary delay in fertility among adolescent mothers (i.e. one that is not accompanied by a reduction in the total number of births) can give rise to an improvement in child height. This is a meaningful, "non-trivial," effect of the program and we control for it using dummy variables for the mother's age in years at birth as well as interactions between age and all of the other baseline covariates. Finally, more "non-trivial" selection effects can arise because the treatment changes the composition of fathers; we control for this channel using characteristics of the father, including education, ethnicity, employment, and health status.[28]

Figure IV plots the cumulatively adjusted Fan regressions. The distribution of direct treatment effects in the UCT arm shown in Panel C with a bold line is remarkably consistent with what we would expect: a positive effect on height-for-age z-scores among children born during the program, which disappears immediately following the cessation of transfers.[29] Column 2 in Table VIII, Panel A shows that inverse propensity weights reduce the effect of UCTs during the program by roughly a third (or 0.3 SD), confirming significant (and positive) selection into

[27] The mean height-for-age z-score in our control group starts out very close to the mean of the reference group at birth, but declines steadily and rapidly as children get older, ending up almost two standard deviations below the global distribution by the time they are 36 months old (Appendix Figure S1). This seems to be a common feature of HAZ in poor countries (see, for example, Figure 1 in (Barham, Macours and Maluccio 2013)).

[28] While father type selection appears to be non-trivial, the program could generate a 'diversionary' treatment effect (as in (Crepon et al. 2013)) 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. 2015).

[29] Figure V shows the final direct treatment effects with confidence intervals, which are omitted from Figure IV for clarity of presentation. The UCT effects on HAZ during the program are statistically significant at the 95% level of confidence.

childbearing, while other pathways further lower the direct effect to 0.5 SD (column 6). The size of this remaining direct effect is consistent with (Barham, Macours and Maluccio 2013), who report that children in Nicaragua who received three years of cash transfers were 0.2-0.4 SD taller; and with (Aguero, Carter and Woolard 2006), who find that children in South Africa receiving child support grants for most of the period between 0-3 years of age gained as much as 0.45 SD in HAZ.

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

The findings here are consistent with the theory that underlies the tradeoff between CCTs and UCTs: UCTs primarily confer an income effect on children born during the program and no effects on children born later because they do not lead to an accumulation of capital (human, physical, or social).[30] On the other hand, CCTs deny such benefits to the children of non-compliers during the program, but may have effects on future children through increased human capital accumulation.

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

5. CONCLUSION

The most striking feature of the results presented in this paper is the transience of the impacts of cash transfers, particularly those given unconditionally. The short-term findings from this study provided hope that many problems that confront adolescent females in Southern and Eastern Africa could be combatted with modest monthly cash transfers ((Baird, McIntosh and Özler 2011); (Baird et al. 2012)), but here we show that when the money runs out, positive impacts quickly dissipate. Particularly glaring are the fleeting decreases in child marriage, teen pregnancy, psychological distress, and HIV in the UCT arm – the prevalence of all of which reverted to control group levels within just two years, implying significant but temporary income effects. Even when there were sustained program effects on school attainment and early marriage and pregnancy, which was the case for baseline dropouts receiving CCTs, these effects did not translate into reductions in HIV or gains in labor market outcomes or empowerment.[31]

Within months of the end of the program, a large number of UCT beneficiaries became pregnant, and were married soon thereafter. These delayed marriages, without any concomitant improvements in education, were to lower quality husbands and resulted in decreased empowerment in this group. This negative impact of waiting to marry in the absence of compensating gains is consistent with evidence from South Asia (Field and Ambrus 2008) and should give pause to interventions seeking to decrease teen marriage without making additional, complementary investments in human capital – at least in the context of sub-Saharan Africa. In this study, we are able to provide suggestive evidence that increased education among adolescent girls may in itself provide an important conduit to better marriage market outcomes: the increased

[31] These findings are consistent with (Duflo, Dupas and Kremer 2015), who find that education subsidies in Kenya reduce dropout, pregnancy, and marriage, but not sexually transmitted infections. They suggest a model in which choices between committed and casual relationships, rather than unprotected sex alone, affect pregnancy and HIV.

education levels of husbands almost exactly parallel the casual effects of the CCT program among baseline dropouts. For these women, delays in marriage were accompanied by large increases in education and there is no evidence that they suffered from delaying marriage and childbearing.

The improvements in test scores in both CCT arms at the end of the program and increased primary school completion among baseline dropouts, the latter of which actually accelerated during the post-program period, did not lead to any significant improvements in employment rates, wages, or basic labor market skills (competencies) two years after the cessation of transfers. This finding is discouraging, as it seems to suggest that increased formal schooling is ineffective at improving economic outcomes for women in this context. This is a fundamental challenge to CCT programs, for which the connection between improved schooling and higher adult incomes is the critical long-term link in the causal chain. As in many parts of sub-Saharan Africa, where the majority of the population is involved in agriculture and informal sector activities, a very small percentage of our sample reports working for wages (or in a household enterprise), but even our "competencies" tests failed to see any improvement under any treatment arm. Applied skills such as using a calculator, making correct change during a market transaction, calculating profits, or being able to follow instructions to correctly apply fertilizers are highly correlated with socio-economic status and overall years of education, but were not significantly improved by the program. Our findings suggest that the increased formal schooling induced by CCTs has not, at least not within two years, increased the earnings potential of program beneficiaries.

Our study has some limitations in terms of program duration and addressing the longer-term effects of cash transfers more generally. First, the data capture outcomes a little more than two years after the cash transfers stopped; therefore we cannot speak to long-term effects, like those analyzed in the U.S. context in recent studies ((Aizer et al. 2016); (Hoynes, Schanzenbach

and Almond 2016)). Particularly, the improved human capital obtained by the CCT dropout group may take longer than two years to bear fruit in terms of employment or wages. Second, the program itself only ran for two years, which is shorter in duration of eligibility than most safety net programs in developing and middle-income countries.[32] So, while we find the benefit of the transfers to be fleeting, this does not mean that a program of longer duration might not have created lasting benefits. However, the readers should also note that the Mothers' Pension program of the early 20th century U.S. had a median duration of three years and was of similar generosity to many cash transfer programs today, including ours (Aizer et al. 2016): that study showed long-term effects in health, education, and income among children of program beneficiaries. It is possible that children of cash transfer beneficiaries in our program might experience some benefits in the longer-run. Third, we track the sample of adolescent females as they get married and form their own households over time, which, while a strength of our study, also means that we cannot speak to medium-term benefits that may have accrued to their parental (original) households that received the majority of the transfers.

Our study provides some important guideposts for the design of effective adolescent-focused cash transfer programs. Given the powerful and multifaceted impacts of the program on the lives of the beneficiaries while it was in place and the near-total evaporation of these impacts when the money ran out, our results suggest a reconsideration of the current enthusiasm for cash transfers as a core anti-poverty tool. Their palliative benefits are uncontested and reinforced by

[32] Indeed, it is precisely the very long-term eligibility that Mexican households have now had under Oportunidades that partially led to the development of Prospera – a new productive inclusion program intended to segue beneficiaries off of cash transfers. However, other interventions for poverty reduction have limited durations like ours, such as the lump-sum transfers by GiveDirectly ((Haushofer and Shapiro 2016)) or the "ultra-poor graduation approach" originally developed by BRAC ((Banerjee et al. 2015); (Bandiera et al. 2016)), which provides support for 12-24 months.

our study, but the idea that they can contribute to a sustained transition out of poverty is unproven and not supported here.[33]

We shed further light on the tradeoffs between the benefits of conditional and unconditional transfers. The lack of knock-on effects from schooling in this context implies that the imperative to use conditions to generate increased investments in human capital may be weak when few formal job options exist. Moreover, the costs of imposing schooling conditions to cash transfers are visible throughout our study. By denying them cash transfers at precisely the moment when they are most likely to start childbearing, a myriad of potential benefits for adolescent girls and young women are missed under CCT programs. A potentially promising way of resolving this tradeoff is to view CCT and UCT programs as complements to each other rather than alternatives: policymakers could provide a basic unconditional cash transfer to adolescent girls topped up by conditional cash transfers for human capital accumulation and desired health behaviors – providing both an incentive to invest in education and health while still guaranteeing a basic level of protection to those who are unable or unwilling to comply with the conditions.

The promising (if only suggestive) evidence of the positive effect of UCTs on children's height provides an important extension to this argument. In previous work, we argued that UCTs are effective in protecting adolescents in this context precisely because it is those out of school who are most vulnerable (Baird, McIntosh and Özler 2011). The most promising evidence presented in this paper of a mechanism for durable impacts of UCTs arises again from a form of vulnerability (this time to malnutrition and stress during a child's first 1,000 days), and the causal

[33] We do not mean to downplay or underestimate the effects of redistributive policies on current poverty and inequality reduction, even if they do not lead to substantive increases in human capital accumulation. Welfare gains from such effects can be as large as, if not larger than, those from human capital investments (Alderman, Behrman and Tasneem 2015).

pathway for benefits requires that transfers not be conditioned on school attendance.[34] In the African context, a schooling CCT is *de facto* conditioned on not being pregnant, meaning that expectant mothers are cut off from transfers at precisely the moment when the social benefits of providing them may be highest. Indeed, (Currie and Almond 2011) have suggested that targeting transfers towards women of childbearing age may be beneficial in the U.S. context, so as to maximize benefits to children *in utero*. This form of targeting would suffer from remarkably little 'leakage' in the Malawian context; two thirds of women aged 20-24 gave birth by age 20 and virtually all females have started childbearing by age 25 (NSO 2005). Our results suggest that targeting unconditional transfers towards low-income adolescents and young women can generate substantial human capital benefits for the next generation in Sub-Saharan Africa.

[34] The importance of resources and shocks during early childhood are documented, among others, for high-income countries like the U.S. ((Aizer et al. 2016); (Hoynes, Schanzenbach and Almond 2016)), middle-income countries like Uruguay (Amarante et al. 2016), and low-income countries like Burkina Faso, India, and Nicaragua ((Akresh et al. 2016); (Shah and Steinberg 2012); (Barham, Macours and Maluccio 2013)). Food prices spiked in Malawi during the program in 2008-09, causing significant decreases in per capita food consumption in Round 2, followed by recovery in Round 3. It is possible that UCTs insulated mothers and their children from the deleterious effects of this large negative covariate shock.

REFERENCES

Aguero, Jorge, Michael Carter, and Ingrid Woolard. 2006. The impact of unconditional cash transfers on nutrition: The South African child support grant. SALDRU Working Paper Series No. 06/08.

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. The long-run impact of cash transfers to poor families. *Am Econ Rev* 106 (4): 935-971.

Akresh, Richard, Damien De Walque, and Harounan Kazianga. 2013. Cash transfers and child schooling: Evidence from a randomized evaluation of the role of conditionality. *World Bank Policy Research Working Paper*, No. 6340.

Akresh, Richard, Emilie Bagby, Damien De Walque, and Harounan Kazianga. 2016. Child labor, schooling, and child ability. *World Bank Policy Research Working Paper*, No. 5965.

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

Amarante, Verónica, Marco Manacorda, Edward Miguel, and Andrea Vigorito. 2016. Do cash transfers improve birth outcomes? Evidence from matched vital statistics, program, and social security data. *American Economic Journal: Economic Policy* 8 (2): 1-43.

Anderson, Siwan and Chris Bidner. 2015. Property rights over marital transfers. *Q J Econ* 130 (3): 1421-1484.

Angrist, Joshua D and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton university press.

Ashraf, Nava, Natalie Bau, Nathan Nunn, and Alessandra Voena. 2014. Bride price and female education. http://scholar.harvard.edu/files/nbau/files/paper_draft_indonesia_zambia_20140821.pdf.

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

Baird, S., C. McIntosh, and B. Özler. 2011. Cash or condition? Evidence from a cash transfer experiment. *Q J Econ* 126 (4): 1709-1753.

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

Baird, Sarah, J Aislinn Bohren, Craig McIntosh, and Berk Özler. 2015. Designing experiments to measure spillover effects, second version. *PIER Working Paper 15-021*.

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

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

Bandiera, Oriana, Robin Burgess, Narayan Das, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2016. Labor markets and poverty in village economies. https://sites.google.com/site/orianabandiera/home/research.

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

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

Barham, Tania, Karen Macours, and John A Maluccio. 2013. Boys' cognitive skill formation and physical growth: Long-Term experimental evidence on critical ages for early childhood interventions. *Am Econ Rev* 103 (3): 467-471.

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

Barrera-Osorio, Felipe, Leigh L Linden, and Juan E Saavedra. 2015. Medium term educational consequences of alternative conditional cash transfer designs: Experimental evidence from Colombia. *CESR-Schaeffer Working Paper*, no. 2015-026.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2015. Self-selection into credit markets: Evidence from agriculture in Mali. http://faculty.wcas.northwestern.edu/~lab823/bktu_mali.pdf.

Behrman, Jere R, Susan W Parker, and Petra E Todd. 2011. Do conditional cash transfers for schooling generate lasting benefits? *Journal of Human Resources* 46 (1): 93-122.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2015. Turning a shove into a nudge? A "labeled cash transfer" for education. *American Economic Journal: Economic Policy* 7 (3): 86-125.

Bianchi, Milo and Matteo Bobba. 2013. Liquidity, risk, and occupational choices. *The Review of Economic Studies* 80 (2): 491-511.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2016. Does grief transfer across generations? Bereavements during pregnancy and child outcomes. *American Economic Journal: Applied Economics* 8 (1): 193-223.

Blattman, Christopher, Julian C Jamison, and Margaret Sheridan. 2015. Reducing crime and violence: Experimental evidence on adult noncognitive investments in Liberia. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2594868.

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2014. Generating skilled self-employment in developing countries: Experimental evidence from Uganda. *Quarterly Journal of Economics, Forthcoming* 129 (2): 697-752.

Bruhn, Miriam and David McKenzie. 2009. In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics* 1 (4): 200-232.

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

Crepon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora. 2013. Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Q J Econ* 128 (2): 531-580.

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

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

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

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

Fernald, Lia CH, Paul J Gertler, and Lynnette M Neufeld. 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, Erica and Attila Ambrus. 2008. Early marriage, age of menarche, and female schooling attainment in Bangladesh. *Journal of Political Economy* 116 (5): 881-930.

Filmer, Deon and Norbert Schady. 2014. The medium-term effects of scholarships in a low-income country. *Journal of Human Resources* 49 (3): 663-694.

Fiszbein, Ariel, Norbert Rüdiger Schady, and Francisco H G Ferreira. 2009. *Conditional Cash Transfers : Reducing Present and Future Poverty*. http://public.eblib.com/choice/publicfullrecord.aspx?p=459451.

Flores, Carlos A and Alfonso Flores-Lagunes. 2009. Nonparametric partial and point identification of net or direct causal effects. In *American Economic Association, Annual Meeting Paper 2009*.

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

Gertler, Paul. 2004. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *Am Econ Rev* 94 (2): 336-341.

Gertler, Paul J, Sebastian W Martinez, and Marta Rubio-Codina. 2012. Investing cash transfers to raise long-term living standards. *American Economic Journal: Applied Economics* 4 (1): 164-192.

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

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

Haddad, Lawrence James, Endang Achadi, Mohamed Ag Bendech, Arti Ahuja, Komal Bhatia,

Zulfiqar Bhutta, Monika Blössner, Elaine Borghi, Esi Colecraft, and Mercedes de Onis. 2014. *Global Nutrition Report 2014: Actions and Accountability to Accelerate the World S Progress on Nutrition*. IFPRI.

Haushofer, Johannes and Jeremy Shapiro. 2016. The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *Quarterly Journal of Economics* forhcoming.

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

Heckman, James J and Stefano Mosso. 2014. The economics of human development and social mobility. NBER *Working Paper*, No. 19656.

Heckman, James J and Tim Kautz. 2013. Fostering and measuring skills: Interventions that improve character and cognition. IZA Discussion Paper, No. 800.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. Long-Run impacts of childhood access to the safety net. *Am Econ Rev* 106 (4): 903-934.

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

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

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

Kang, Joseph D. Y. and Joseph L. Schafer. 2007. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* 22 (4): 523-539.

LaFave, Daniel and Duncan Thomas. 2016. Height and cognition at work: Labor market productivity in a low income setting. *NBER Working Paper 22290*.

MacKinnon, David P. 1994. Analysis of mediating variables in prevention and intervention research. *NIDA Research Monograph* 139:127-127.

Macours, Karen, Norbert Schady, and Renos Vakis. 2012. Cash transfers, behavioral changes, and cognitive development in early childhood: Evidence from a randomized experiment. *American Economic Journal: Applied Economics* 4 (2): 247-273.

McQueston, Kate, Rachel Silverman, and Amanda Glassman. 2013. The efficacy of interventions to reduce adolescent childbearing in low- and middle-income countries: A systematic review. *Stud Fam Plann* 44 (4): 369-88.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. Returns to capital in microenterprises: Evidence from a field experiment. *Q J Econ*1329-1372.

———. 2012. One-time transfers of cash or capital have long-lasting effects on microenterprises in Sri Lanka. *Science* 335 (6071): 962-966.

Milligan, Kevin and Mark Stabile. 2009. Do child tax benefits affect the wellbeing of children? Evidence from canadian child benefit expansions. *Am Econ Rev* 99 (2): 128-132.

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

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

National Statistical Office, and ICF Macro. 2005. Malawi demographic and health survey 2010. *National Statistical Office of Malawi (NSO) and ORC Macro, Zomba, Malawi/Calverton, MD, USA*.

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

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

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

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

Shah, Manisha and Bryce Millett Steinberg. 2012. Could droughts improve human capital? Evidence from India. IZA Discussion Paper No. 9440.

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

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

Van der Laan, Mark J and James M Robins. 2003. *Unified Methods for Censored Longitudinal Data and Causality*. Springer Science & Business Media.

Woerner, W, B Fleitlich-Bilyk, R Martinussen, J Fletcher, G Cucchiaro, P Dalgalarrondo, M Lui, and R Tannock. 2004a. The strengths and difficulties questionnaire overseas: Evaluations and applications of the SDQ beyond europe. *Eur Child Adolesc Psychiatry* 13 Suppl 2:47-54.

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

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

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

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

Table II: Program impacts on marriage and fertility (beneficiaries)

Panel A: Baseline Dropouts

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

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

Table III: Program impacts on HIV and Anemia (beneficiaries)

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

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

Table IV: Program impacts on labor market outcomes and empowerment (beneficiaries)

Panel A: Baseline Dropouts

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Opportunity cost of time is calculated by taking the minimum daily wage the respondent would take for one year of work in her village. Detail on the construction of the super-indices can be found in the pre-analysis plan and in appendix A. The change in subjective wellbeing asks the respondent where she sees herself on a 10-step ladder comparing five years ago to today, where zero represents the worst possible life she could have and 10 represents the best possible life she could have. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

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

	Husband Quality Super Index (Standardized)	Highest Grade Completed	=1 if Passed Primary School (PSLC)	=1 if Passed Junior Secondary School (JCE)	=1 if Passed Secondary School (MSCE)	Cognitive Test Score (Standardized)	Typical Wage in Past Three Months	=1 if Currently Employed	Sexual Activity and Marital Fidelity (Standardized)	=1 if Does Not Suffers from Psychological Distress	=1 if HIV Positive
	Two Years After Program										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Baseline Dropouts											
=1 if Conditional Schoolgirl	0.084	0.561	0.032	0.029	0.074**	-0.049	-0.081	-0.024	0.032	0.007	-0.005
	(0.106)	(0.348)	(0.054)	(0.046)	(0.037)	(0.110)	(0.225)	(0.040)	(0.106)	(0.061)	(0.035)
Mean in Control Group	0.000	7.806	0.526	0.314	0.097	0.000	1.194	0.246	0.000	0.634	0.055
Sample Size	326	326	326	326	326	323	325	326	325	326	265
Panel B: Baseline Schoolgirls											
=1 if Conditional Schoolgirl	0.141	0.046	0.024	0.012	0.059	0.014	0.014	0.045	0.284***	0.074	0.001
	(0.096)	(0.271)	(0.043)	(0.049)	(0.053)	(0.109)	(0.262)	(0.051)	(0.091)	(0.060)	(0.033)
=1 if Unconditional Schoolgirl	-0.186	-0.454	0.005	0.017	-0.088	-0.357**	-0.406	-0.091	0.013	0.008	0.010
	(0.180)	(0.425)	(0.068)	(0.086)	(0.054)	(0.163)	(0.344)	(0.093)	(0.219)	(0.093)	(0.041)
p-value UCT vs. CCT	0.084	0.240	0.776	0.954	0.042	0.044	0.225	0.17	0.196	0.508	0.845
p-value Treatment	0.145	0.490	0.845	0.964	0.118	0.087	0.432	0.358	0.006	0.471	0.971
Mean in Control Group	0.000	9.743	0.699	0.541	0.258	0.000	1.42	0.352	0.000	0.647	0.052
Sample Size	543	543	543	543	543	539	540	543	542	541	457

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

Table VI: Program impacts on child outcomes (children of beneficiaries)

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

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

Table VII: Program impacts on height-for-age z-scores (children of beneficiaries: **baseline dropouts**)

Panel A: Born During Program	Child Gender	+ Selection weights	+ Mother Attributes, no age	+ Child Age	+ Mother Age	+ Father Attributes
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	-0.015	-0.200	-0.193	-0.178	-0.066	-0.051
	(0.128)	(0.174)	(0.155)	(0.143)	(0.136)	(0.136)
Sample Size	367	367	367	367	367	367
Panel B: Born Within 9 Months of Program Ended						
=1 if Conditional Schoolgirl	0.353	0.405	0.385	0.498*	0.611**	0.577**
	(0.296)	(0.316)	(0.286)	(0.262)	(0.297)	(0.260)
Sample Size	88	88	88	88	88	88
Panel C: Born More than 9 Months After Program Ended						
=1 if Conditional Schoolgirl	-0.269	-0.175	-0.186	-0.171	-0.214	-0.183
	(0.168)	(0.173)	(0.164)	(0.158)	(0.154)	(0.152)
Number of observations	287	287	287	287	287	287
Control Structure:						
Uses attrition propensity weight		X	X	X	X	X
Mother baseline controls, except age			X	X	X	X
Cubic in child age in months				X	X	X
Maternal age in years, age interactions					X	X
Father characteristics						X

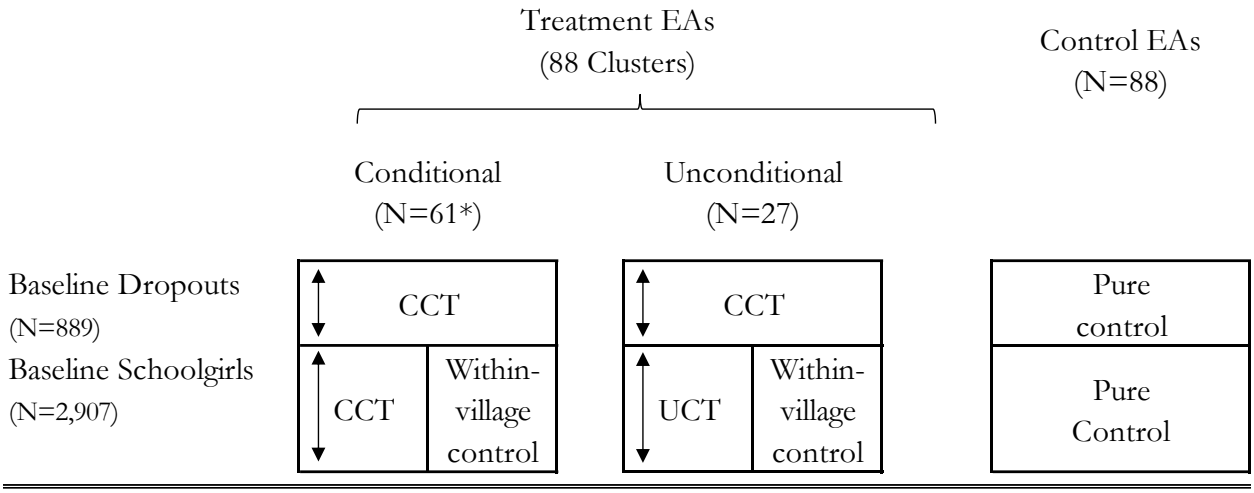
Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights. Specification (3) adds the following maternal baseline controls: stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. Specification (4) adds a cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. Finally, specification (6) adds the following characteristics of the father: highest education level, religion, ethnicity, main activity, and likely HIV status.. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table VIII: Program impacts on height-for-age z-scores (children of beneficiaries: **baseline schoolgirls**)

Panel A: Born During Program	Child Gender	+ Selection weights	+ Mother Attributes, no age	+ Child Age	+ Mother Age	+ Father Attributes
	(1)	(2)	(3)	(4)	(5)	(6)
=1 if Conditional Schoolgirl	0.155	-0.038	-0.050	0.024	0.128	0.124
	(0.162)	(0.190)	(0.192)	(0.184)	(0.150)	(0.155)
=1 if Unconditional Schoolgirl	0.953**	0.658***	0.525**	0.609**	0.523**	0.523*
	(0.476)	(0.219)	(0.221)	(0.259)	(0.250)	(0.299)
p-value UCT vs. CCT	0.091	0.005	0.022	0.026	0.115	0.156
p-value Treatment	0.123	0.007	0.040	0.051	0.114	0.218
Sample Size	315	315	315	315	315	315
Panel B: Born Within 9 Months of Program Ended						
=1 if Conditional Schoolgirl	0.251	0.156	0.127	0.086	0.040	0.086
	(0.279)	(0.280)	(0.292)	(0.205)	(0.207)	(0.194)
=1 if Unconditional Schoolgirl	0.177	0.130	-0.346	-0.445	-0.430**	-0.434**
	(0.514)	(0.348)	(0.399)	(0.271)	(0.216)	(0.193)
p-value UCT vs. CCT	0.887	0.945	0.242	0.078	0.087	0.028
p-value Treatment	0.663	0.835	0.502	0.189	0.124	0.047
Sample Size	211	211	211	211	211	211
Panel C: Born More than 9 Months After Program Ended						
=1 if Conditional Schoolgirl	-0.011	0.186	0.497	0.275	0.259	0.257
	(0.187)	(0.245)	(0.445)	(0.212)	(0.187)	(0.179)
=1 if Unconditional Schoolgirl	-0.351**	-0.365*	-0.651***	-0.093	-0.121	-0.123
	(0.174)	(0.200)	(0.242)	(0.155)	(0.164)	(0.183)
p-value UCT vs. CCT	0.115	0.028	0.006	0.098	0.083	0.078
p-value Treatment	0.114	0.055	0.002	0.252	0.214	0.186
Sample Size	507	506	506	506	506	506
Control Structure:						
Uses attrition propensity weight		X	X	X	X	X
Mother baseline controls, except age			X	X	X	X
Cubic in child age in months				X	X	X
Maternal age in years, age interactions					X	X
Father characteristics						X

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. The height-for-age z-score is calculated using the 2006 WHO child growth standards. Specification (1) controls for the gender of the child. Specification (2) adds selection weights. Specification (3) adds the following maternal baseline controls: stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. Specification (4) adds a cubic in child age. Specification (5) adds maternal age and maternal age interacted with the other baseline covariates. Finally, specification (6) adds the following characteristics of the father: highest education level, religion, ethnicity, main activity, and likely HIV status.. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

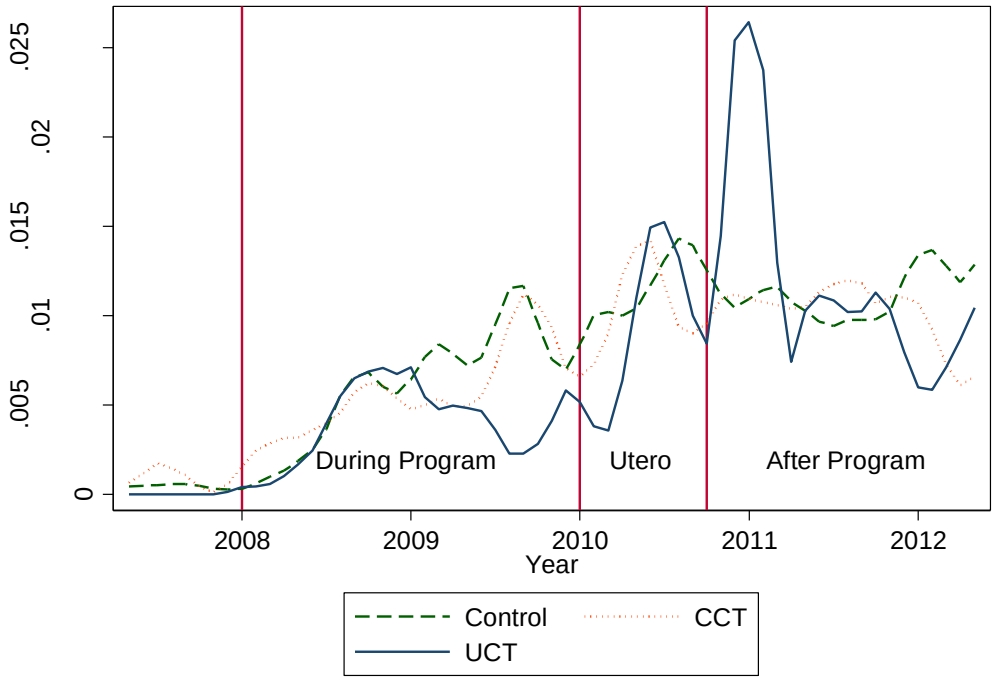
Figure I: Research Design



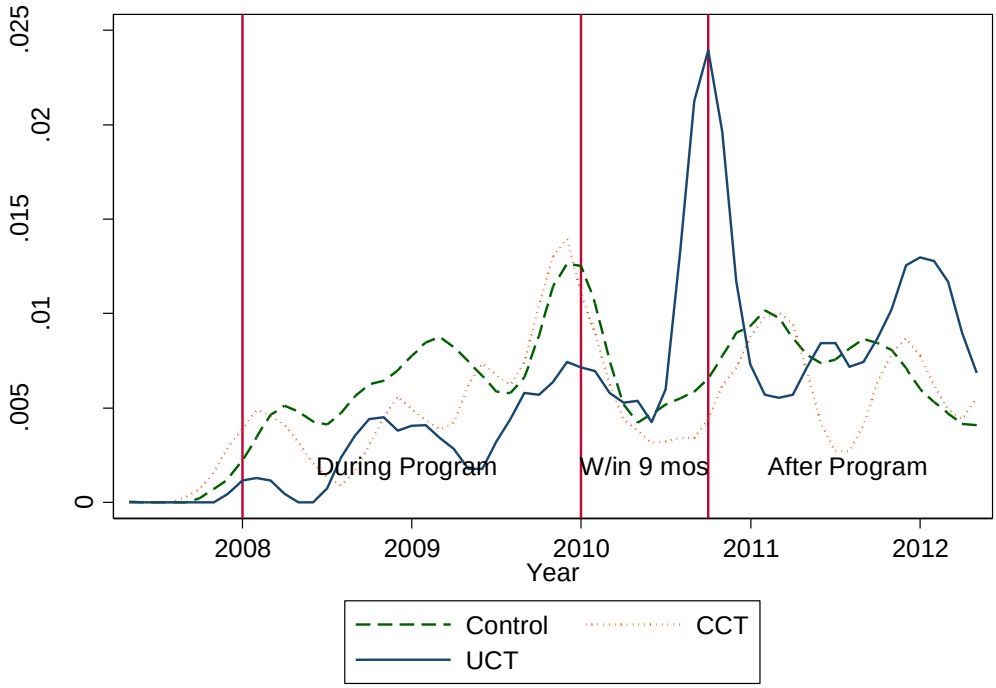
*In 15 of the 61 conditional treatment clusters only baseline dropouts were treated.

Figure II: Monthly marriage and fertility rates for baseline schoolgirls

Panel A: Monthly Fertility Rates



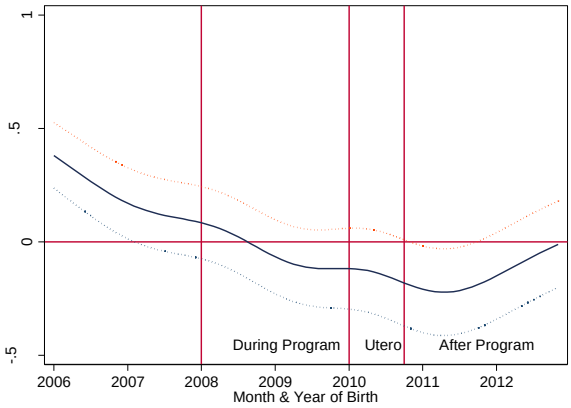
Panel B: Monthly Marriage Rates



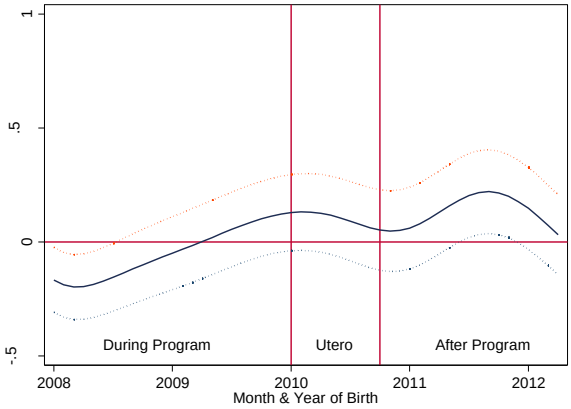
Notes: Figures illustrate the smoothed fraction of core respondents who give birth (Panel A) or get married (Panel B) in each month using retrospective information on the month of birth and marriage, respectively.

Figure III: Fan regressions of height-for-age z-scores by month of birth, raw treatment effects with 95% confidence intervals

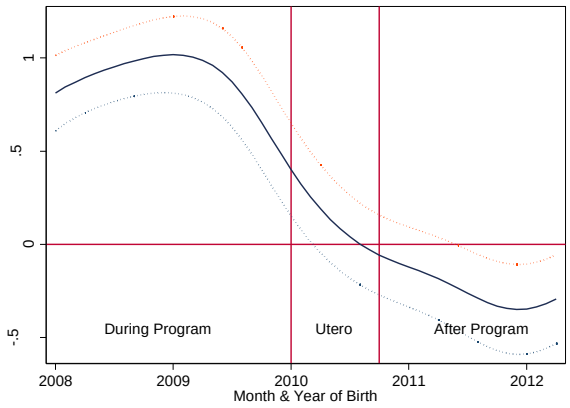
Panel A: Baseline Dropouts



Panel B: Baseline Schoolgirls, CCT



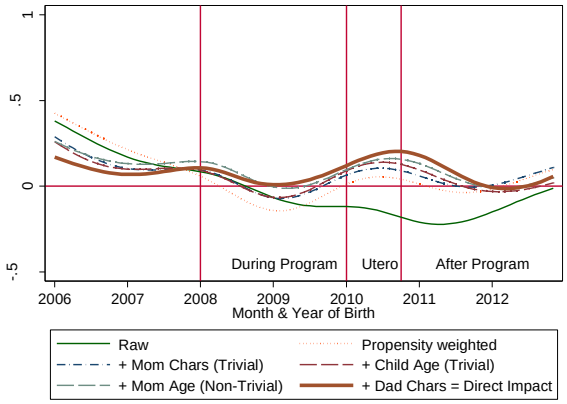
Panel C: Baseline Schoolgirls, UCT



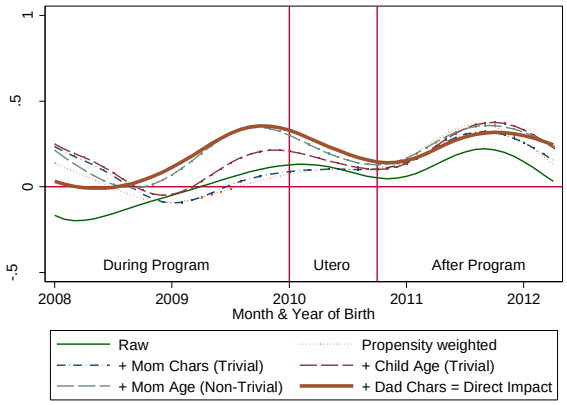
Notes: Figures utilize a Fan regression that estimates the raw OLS regressions (as in Table VI but without control variables) locally at each month and year of birth for every observation, using the product of a standard sampling weight and a weight decreasing in time differences from that month of birth.

Figure IV: Fan regressions of height-for-age z-scores by month of birth, treatment effects including successive mediators

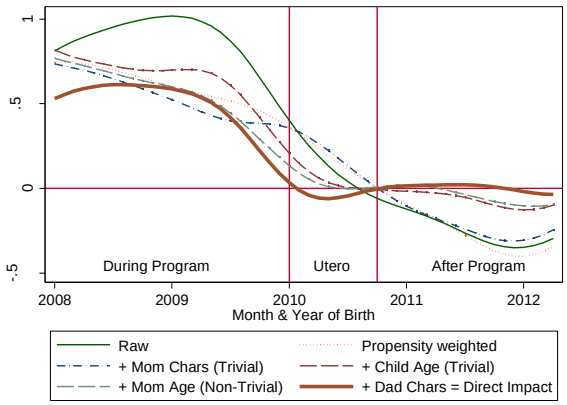
Panel A: Baseline Dropouts



Panel B: Baseline Schoolgirls, CCT



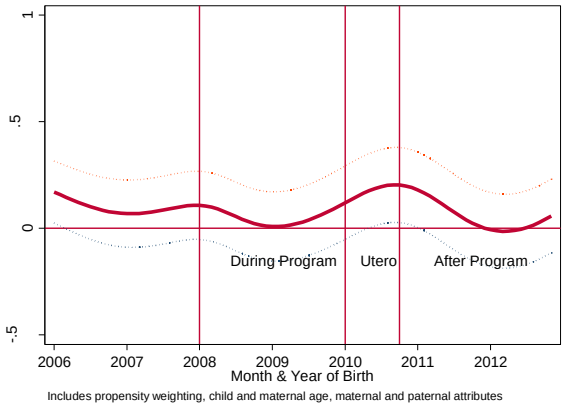
Panel C: Baseline Schoolgirls, UCT



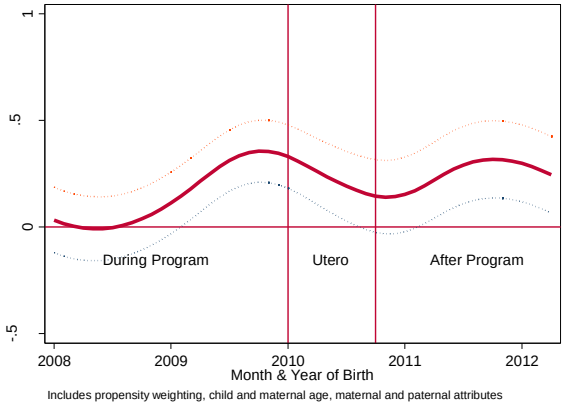
Notes: Figures utilize six different Fan regressions for each treatment group that estimate impacts at each month and year of birth, beginning with no controls and then including successive controls for endogenous attributes of the child and parents that remove an increasingly stringent set of mediating variables from the estimated impact.

Figure V: Fan regressions of height-for-age z-scores by month of birth, final adjusted treatment effects with 95% confidence intervals

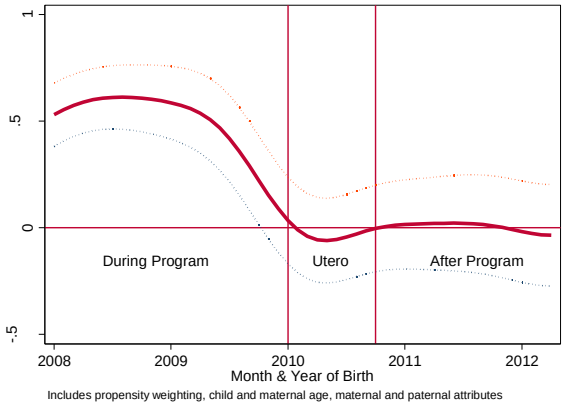
Panel A: Baseline Dropouts



Panel B: Baseline Schoolgirls, CCT



Panel C: Baseline Schoolgirls, UCT



Notes: Figures based on the fully controlled mediation Fan regressions shown in Figure V, and present 95% confidence intervals for the direct treatment effect on child HAZ with all the selection mediators removed.

Supplemental Online Appendix for:

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

Sarah Baird, Craig McIntosh and Berk Özler*

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

June 13, 2016

This PDF file includes:

Appendix A: Detailed description of construction of outcome variables

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

Outcome	Table	Population	Primary?	Rounds	Description
Highest Grade Completed	I	Core Respondent	Yes	2,3,4	Highest grade completed is the self-reported highest grade attended/completed by the core respondent at the time of the household survey. In rounds 2 and round 3 this is the highest grade attended and in round 4 in is highest grade completed.
Passed Primary School (PSLC)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Primary School Leaving Certificate (PSLC) at the end of 8th grade.
Passed Junior Secondary School (JCE)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Junior Certificate of Education at the end of 10th grade
Passed Secondary School (MSCE)	S1	Core Respondent	No	2,3,4	=1 if core respondent passed the Malawi School Certificate of Education at the end of 12th grade
English Test Score (Standardized)	I	Core Respondent	N/A	3	Total score from a professionally developed test of English reading comprehension based on the Malawian school curricula for the grades the target population was attending.

Outcome	Table	Population	Primary?	Rounds	Description
TIMMS Math Score (Standardized)	I	Core Respondent	N/A	3	Total score from a professionally developed test of mathematics based on the Malawian school curricula for the grades the target population was attending. Five questions (four from the Fourth Grade test and one from the eighth Grade test) from Trends in Mathematics and Science Study (TIMMS) 2007, which is a cycle of internationally comparative assessments in mathematics and science carried out at the fourth and eighth grades every 4 years, were added to the Math test.
Non-TIMMS Math Score (Standardized)	I	Core Respondent	N/A	3	
Cognitive Test Score (Standardized)	I	Core Respondent	N/A	3	The cognitive test score is based on Raven's Colored Progressive Matrices.
Competencies Score (Standardized)	I	Core Respondent	Yes	4	Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The overall competencies score is a standardized index of the five competencies listed immediately below.
Fertilizer Application (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to follow instructions for applying fertilizer to maize.
Change Given (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to make change following hypothetical scenarios of market transactions.
Sending a Text Message (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to send a text message saying "hello" to a specified phone number.
Using a Calculator (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to use a cell-phone calculator to calculate 873*17.
Calculating Profits (Standardized)	S3	Core Respondent	No	4	Standardized index of ability to calculate profits from hypothetical scenarios.

Outcome	Table	Population	Primary?	Rounds	Description
Total Time Spent on Competencies (Standardized)	S3	Core Respondent	No	4	Standardized index of average time taken across the five competencies
Ever Married	II	Core Respondent	Yes	2,3,4	=1 if core respondent ever married at the time of the household survey. Variable is corrected for inconsistences across rounds.
Age at First Marriage	II	Core Respondent	Yes	4	Age of core respondent at first marriage.
Ever Pregnant	II	Core Respondent	Yes	2,3,4	=1 if core respondent ever pregnant at time of the household survey. Variable is corrected for inconsistences across rounds.
Number of Live Births	II	Core Respondent	Yes	2,3,4	Number of live births reported by core respondent at time of the household survey.
Age of First Birth	II	Core Respondent	Yes	4	Age of core respondent at first live birth.
Desired Fertility	II	Core Respondent	No	4	Reported number of total children desired by the core respondent (including any they already have).
HIV Positive	III	Core Respondent	Yes	2,3,4	=1 if core respondent is HIV positive. Biomarker data for HIV were collected through home based voluntary counseling and testing (HCT).
Anemic	III	Core Respondent	Yes	4	=1 if core respondent is anemic (biomarker data). An individual is considered anemic if her hemoglobin count is less than or equal to 11g/dL if pregnant and less than or equal to 12d/dL if non-pregnant based on WHO guidelines to define mild anemia.

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

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

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

Outcome	Table	Population	Primary?	Rounds	Description
Change in Subjective Wellbeing from Five Years Ago to Today	IV	Core Respondent	Yes	4	The change in subjective wellbeing asks the respondent where she sees herself on a 10-step ladder comparing five years ago to today, where zero represents the worst possible life she could have and 10 represents the best possible life she could have.
Super-Index of Unmarried Empowerment (Standardized)	IV	Core Respondent	Yes	4	Standardized super index of unmarried empowerment, measured for core respondents who are not married at the time of the household survey. Includes two components described below: index of non-abuse and index of autonomy.
Index of Unmarried non-Abuse (Standardized)	-	Core Respondent	No	4	Standardized index of unmarried non-abuse, measured for core respondents who are not married at the time of the household survey. Includes question on whether the core respondent has been beaten or mistreated physically, whether the core respondent has been forced to have sex against her will, and a series of five questions on threats or physical violence. The variable is constructed so that higher values are an indication of not experiencing physical abuse.

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

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

Outcome	Table	Population	Primary?	Rounds	Description
Microenterprise Participation	-	Core Respondent	No	4	=1 if the core respondent controls the use of profits from any household microenterprise.
Livestock Control	-	Core Respondent	No	4	=1 if the core respondent is involved in decision-making around any of the household's livestock.
Ratio of Wife to Husband Consumption (Standardized)	-	Core Respondent	No	4	Standardized index of the ratio of wife's to husband's consumption focusing on gender specific goods.
Husband Quality Super Index (Standardized)	V	Husband	Yes	4	Standardized quality super index for husbands of core respondents. Includes nine components listed below: highest grade completed, passed PSLC, passed JCE, passed MSCE, cognitive test score, typical wage in the past three months, employment status, sexual activity and marital fidelity, and physiological distress.
Highest Grade Completed	V	Husband	No	4	Highest grade completed is the self-reported highest completed by the husband at the time of the household survey.
=1 if Passed Primary School (PSLC)	V	Husband	No	4	=1 if husband passed the Primary School Leaving Certificate (PSLC) at the end of 8th grade.
=1 if Passed Junior Secondary School (JCE)	V	Husband	No	4	=1 if husband passed the Junior Certificate of Education at the end of 10th grade
=1 if Passed Secondary School (MSCE)	V	Husband	No	4	=1 if husband passed the Malawi School Certificate of Education at the end of 12th grade
Cognitive Test Score (Standardized)	V	Husband	No	4	The cognitive test score is based on Raven's Colored Progressive Matrices.
Typical Wage in Past Three Months	V	Husband	No	4	Typical wage the husband reports earning in the past three months. It takes on a value of zero if the husband earned nothing.

Outcome	Table	Population	Primary?	Rounds	Description
=1 if Currently Employed	V	Husband	No	4	=1 if husband currently employed.
Sexual Activity and Marital Fidelity (Standardized)	V	Husband	No	4	The husband's sexual activity and marital fidelity standardized index is constructed from three variables: number of sexual partners ever, number of sexual partners in the past 12 months and an indicator for concurrent multiple partners. Constructed so that higher values are better.
=1 if Does Not Suffer from Psychological Distress	V	Husband	No	4	=1 if husband does not suffer from psychological distress. Psychological distress is equal to one if the summed General Health Questionnaire- 12 score is equal to three or higher, and is zero otherwise.
=1 if HIV Positive	V	Husband	No	4	=1 if husband is HIV positive. Biomarker data for HIV was collected through home based voluntary counselling and testing (HCT).
Attitudes Towards Women's Empowerment Super Index (Standardized)	S9	Husband	Yes	4	Standardized super index of five indicators listed below: preferences for children's education, attitudes towards wife's autonomy, attitudes towards abuse, divorce prospects, and desired fertility.
Index of Preferences for Children's Education (Standardized)	S9	Husband	No	4	Standardized index of preferences for child education. Includes nine questions regarding attitudes towards the importance of schooling for girls.

Outcome	Table	Population	Primary?	Rounds	Description
Index of Attitudes Towards Wife's Autonomy (Standardized)	S9	Husband	No	4	Standardized index of the husband's attitude towards his wife's autonomy. Series of questions asking whether the wife needs permission to engage in day to day activities. This variable takes on a higher value if the wife does not need permission.
Index of Attitudes Towards (non)-Abuse (Standardized)	S9	Husband	No	4	Standardized index of husband's attitudes towards beating his wife. This variable uses a series of three questions asking when wife beating is justified. This variable takes on a higher value if the husband does not think beating is justified.
Index of Wife's Divorce Prospects (Standardized)	S9	Husband	No	4	Standardized index of wife's divorce prospects. This looks at a series of variables related to a man's ability to leave the marriage and keep things from the household. This variable takes on a higher value the less ability a man has to divorce his wife.
Desired Fertility	S9	Husband	No	4	Husband's desired fertility. Note that this variable is standardized and made negative when added to the super index.
Height-for-Age Z-Score	VI-VIII	Child	Yes	4	The height-for-age (length-for-age) z-score is calculated using the 2006 WHO child growth standards. See the following for more details: Leroy, Jef L (2011). zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards. http://www.ifpri.org/staffprofile/jef-leroy.
Neonatal Mortality	VI	Child	Yes	3,4	=1 if child died at or before 31 days.

Outcome	Table	Population	Primary?	Rounds	Description
Post-Neonatal Mortality	VI	Child	Yes	3,4	=1 if child died between the ages of one month and one year. Defined for those that survived the first month of life.
Parenting Practices Percentage Score	VI	Child	Yes	4	Percentage score across a series of 16 questions on parenting practices related to addressing behavior problems to interacting with the child (the total number of questions asked various by the age of the child). These questions are only asked of living children.
Exclusively Breastfed for First Six Months	VI	Child	Yes	3,4	=1 if mother exclusively breastfed for the first 6 months of life, is still exclusively breastfeeding for those under 6 months, or who breastfed until death.
Standardized Malawi Development Assessment Tool	VI	Child	Yes	4	The standardized Malawi Developmental Assessment Tool (MDAT) is a test of fine motor skills, language, and hearing administered directly to the child.
Reported Child Difficulties	VI	Child	Yes	4	The standardized reported child difficulties and reported pro-social behaviors are created using the Strengths and Difficulties Questionnaire (SDQ) and administered to the parent. Details on construction of these variables can be found at http://www.sdqinfo.com/c3.html.
Reported Pro-Social Behaviors	VI	Child	Yes	4	

Table S1: Attrition

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. All regressions include baseline centered values of the following variables: age indicators, stratum indicators, household asset index, highest grade attended, an indicator for never had sex. Columns (2) and (4) interact the centered baseline controls with treatment. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S2: Baseline means and balance

	Baseline Schoolgirl				Baseline Dropout	
	Mean (s.d.)			p-value	Mean (s.d.)	
	Control group	Conditional group	Unconditional Group	(CCT-UCT)	Control group	Conditional group
	(1)	(2)	(3)	(4)	(5)	(6)
Urban Household	0.346	0.478	0.418	0.726	0.181	0.129
	(0.476)	(0.500)	(0.494)		(0.385)	(0.335)
Mother Alive	0.839	0.800	0.828	0.431	0.783	0.749
	(0.368)	(0.401)	(0.378)		(0.413)	(0.434)
Father Alive	0.709	0.718	0.76	0.341	0.656	0.649
	(0.454)	(0.451)	(0.428)		(0.476)	(0.478)
Household Size	6.375	6.341	6.659	0.156	6.120	6.104
	(2.262)	(2.134)	(2.063)		(2.388)	(2.617)
Asset Index	0.632	1.100	1.373*	0.572	-0.831	-0.743
	(2.575)	(2.721)	(2.444)		(2.233)	(2.484)
Age	15.228	14.919	15.466	0.002	17.579	17.162
	(1.904)	(1.828)	(1.926)		(2.397)	(2.478)
Highest Grade Attended	7.506	7.262	7.928**	0.004	6.105	5.940
	(1.651)	(1.601)	(1.587)		(2.856)	(2.864)
Never Had Sex	0.800	0.807	0.790	0.682	0.315	0.294
	(0.400)	(0.395)	(0.408)		(0.465)	(0.456)
Ever Pregnant	0.021	0.029	0.029	0.964	0.445	0.420
	(0.144)	(0.169)	(0.168)		(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 S3: Program impacts on educational qualifications and competencies (beneficiaries)

	Educational Qualifications									Competencies (Standardized)					
	=1 if Passed Primary School (PSLC)			=1 if Passed Junior Secondary School (JCE)			=1 if Passed Secondary School (MSCE)			Fertilizer Application	Change Given	Sending a Text Message	Using a Calculator	Calculating Profits	Total Time Spent
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	Two Years After Program					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Baseline Dropouts															
=1 if Conditional Schoolgirl	0.030	0.058**	0.081***	0.012	0.049**	0.034	0.004	0.003	0.016	-0.044	-0.014	0.101	0.065	0.094	-0.007
	(0.025)	(0.025)	(0.026)	(0.019)	(0.021)	(0.022)	(0.008)	(0.010)	(0.011)	(0.069)	(0.062)	(0.072)	(0.071)	(0.076)	(0.091)
Mean in Control Group	0.328	0.351	0.366	0.085	0.123	0.136	0.008	0.025	0.026	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	697	718	744	697	718	744	697	718	744	742	741	741	741	742	742
Panel B: Baseline Schoolgirls															
=1 if Conditional Schoolgirl	0.030	0.013	-0.014	-0.013	0.055*	0.033	-0.004*	0.005	0.006	0.015	0.048	0.077	0.060	-0.006	-0.113
	(0.039)	(0.024)	(0.019)	(0.022)	(0.028)	(0.028)	(0.002)	(0.011)	(0.021)	(0.071)	(0.071)	(0.070)	(0.054)	(0.076)	(0.085)
=1 if Unconditional Schoolgirl	0.046	0.030	0.017	0.002	0.016	0.010	-0.006*	-0.009	-0.065**	0.096	-0.017	0.161**	0.098	-0.045	-0.118
	(0.038)	(0.026)	(0.016)	(0.022)	(0.045)	(0.035)	(0.003)	(0.015)	(0.027)	(0.092)	(0.057)	(0.079)	(0.064)	(0.090)	(0.085)
p-value UCT vs. CCT	0.755	0.600	0.166	0.546	0.439	0.565	0.325	0.385	0.022	0.378	0.389	0.364	0.584	0.636	0.963
p-value Treatment	0.386	0.488	0.359	0.797	0.148	0.486	0.150	0.683	0.045	0.570	0.685	0.105	0.249	0.862	0.258
Mean in Control Group	0.496	0.776	0.879	0.144	0.337	0.537	0.004	0.054	0.170	0.000	0.000	0.000	0.000	0.000	0.000
Sample Size	1,967	2,019	2,047	1,967	2,019	2,047	1,967	2,019	2,047	2,048	2,046	2,047	2,047	2,048	2,048

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Competencies represent a set of skills that were anticipated to be sensitive to education and relevant for non-formal employment. The skills tested included reading and following instructions to apply fertilizer; making correct change during hypothetical market transactions; sending text messages and using the calculator on a mobile phone, and calculating profits under hypothetical business scenarios. All competency components are standardized to have a mean of zero and a standard deviation of one in the control group. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

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

	=1 if Ever Had Sex			Number of Sexual Partners (lifetime)			=1 if Sexually Active During the Past 12 Months		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Baseline Dropouts									
=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, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S5: Program impacts on sexual behavior (beneficiaries: **intensive** margin)

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). We correct for inconsistencies in 'ever had sex' across rounds. 'Age at First Sex' is defined for those that had ever had sex. 'Older Partner' is defined as having a partner who is 5 years older or more in the past 12 months. 'Condom Use' is defined as using a condom at last sex with most recent sexual partner. Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

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

	=1 if Suffers from Psychological Distress			Number of Times Respondent Ate Protein Rich Foods During the Past 7 Days (out of 21)		
	During Program	End of Program	Two Years After Program	During Program	End of Program	Two Years After Program
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline Dropouts						
=1 if Conditional Schoolgirl	-0.002	0.010	0.038	0.326	0.224	0.228
	(0.039)	(0.036)	(0.042)	(0.202)	(0.192)	(0.181)
Mean in Control Group	0.463	0.314	0.424	3.678	3.989	3.741
Sample Size	698	715	743	698	718	744
Panel B: Baseline Schoolgirls						
=1 if Conditional Schoolgirl	-0.068**	-0.037	-0.030	0.385**	0.596***	0.072
	(0.032)	(0.047)	(0.032)	(0.195)	(0.174)	(0.141)
=1 if Unconditional Schoolgirl	-0.139***	-0.026	-0.002	0.445**	0.338**	-0.043
	(0.035)	(0.054)	(0.046)	(0.199)	(0.153)	(0.240)
p-value UCT vs. CCT	0.068	0.860	0.552	0.814	0.215	0.672
p-value Treatment	0.000	0.677	0.627	0.023	0.001	0.858
Mean in Control Group	0.372	0.313	0.369	3.967	4.052	4.134
Sample Size	1,963	2,013	2,045	1,967	2,018	2,047

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Psychological distress is equal to one if the summed GHQ- 12 score is equal to three or higher, and is zero otherwise. Protein rich foods are defined as those containing animal proteins, i.e. meat, fish, and eggs. The number of days each item was consumed over the past week are summed to create the outcome variable. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S7: Program impacts on labor market outcomes and consumption (beneficiaries: **secondary** outcomes)

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Effective daily wage is calculated using earnings and activities in the past seven days. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S8: Program impacts on empowerment (beneficiaries: **secondary** outcomes)

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. Detail on the construction of the indices can be found in the pre-analysis plan and in appendix A. Aspirations asks the respondent where she sees herself on a 10-step ladder comparing today to five years from now, where zero represents the worst possible life she could have and 10 represents the best possible life she could have. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Table S9: Program impacts on marriage market outcomes (husband attitudes towards women's empowerment)

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

Notes: Regressions are OLS models with robust standard errors clustered at the EA level. All regressions are weighted to make them representative of the target population in the study EAs. All variables are constructed so that higher values are better. The husband gender empowerment super index is a standardized index of the other variables in this table. Additional detail on the construction of the indices can be found in the pre-analysis plan and in appendix A. Baseline values of the following variables are included as controls in the regression analyses: age indicators, stratum indicators, household asset index, highest grade attended, and an indicator for never had sex. We restrict the sample to husbands of respondents who were surveyed during the latest household survey conducted two years after the program (Round 4). Parameter estimates statistically different than zero at 99% (***), 95% (**), and 90% (*) confidence.

Figure S1. Height-for-age z-score by age in months (control group)