# Family Planning and Children's Human Capital: Experimental Evidence from Urban Malawi*

Daniel Maggio† Mahesh Karra‡ David Canning§

February 3, 2023

**Abstract**

We conduct a randomized controlled trial that provides pregnant and immediate postpartum women with improved access to family planning through counseling, free transport to a clinic, and financial reimbursement for family planning services over two years. We study the effects of our intervention on child growth and development outcomes among 1,034 children born to participating women directly before intervention rollout. We find that children born to mothers assigned to the family planning intervention arm were 0.28-0.34 standard deviations taller for their age and 10.7-12.0 percentage points less likely to be stunted within a year of exposure to the intervention. Children born to mothers assigned to the intervention arm also scored 0.19-0.23 standard deviations higher on a caregiver-reported measure of cognitive development after two years of intervention exposure. Non-measurement of children is a challenge in our study; however, we show that our estimates are robust to multiple methods of correcting for potential attrition bias. Our results are consistent with models of fertility that link couples' fertility decisions to child health and human capital. Our results also suggest that improved access to family planning may have positive downstream effects on child health beyond contraceptive use and fertility outcomes.

*This trial was registered at the American Economics Association Registry for randomized controlled trials on May 7, 2015 (AEARCTR-0000697) and at the Registry for International Development Impact Evaluations (RIDIE) on May 28, 2015 (RIDIE-STUDY-ID-556784ed86956). This research uses original data collected by Canning and Karra with support from Innovations for Poverty Action (IPA) in Malawi. The authors thank John Hoddinott, Chris Barrett, John Cawley, Günther Fink, Jessica Leight, Jack Cavanaugh, and seminar participants at Cornell University, NEUDC 2021, ICDE 2021, IPC 2021, and CSAE 2022 for their valuable feedback and comments. The authors would also like to acknowledge the dedication and support of Carly Farver, Patrick Baxter, Bagrey Ngwira, Reginald Chunda, Viola Nyirongo, Violet Chitsulo, Macdonald Salamu, and the entire Malawi Family Planning Study team. This project was supported by two grants from the Hewlett Foundation and the Program for Women's Empowerment Research (POWER) at the Boston University Global Development Policy Center. Ethical approval to conduct the study was received from the Harvard University IRB (protocol number IRB16-0421) and the Malawi National Health Sciences Research Committee (protocol number 16/7/1628). The findings, interpretations, and conclusions expressed in this study are entirely those of the authors. They do not represent the views of Innovation for Poverty Action and its affiliated organizations, or those of the Executive Directors of the Innovation for Poverty Action they represent.

**Keywords**: Family Planning, Child Health, Stunting, Cognitive Development, Randomized Control Trial

†**Corresponding Author.** Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY USA. dwm272@cornell.edu.

‡Frederick S. Pardee School of Global Studies, Boston University, Boston, MA USA. mvkarra@bu.edu.

§Harvard T.H. Chan School of Public Health, Boston, MA USA. dcanning@hsph.harvard.edu.

# 1 Introduction

Roughly 14 million unwanted pregnancies occur in Sub-Saharan Africa each year, and an estimated 37 percent of these pregnancies are terminated (Bankole et al., 2020). For those pregnancies that result in live births, women and couples are left to care for a child that was unplanned and that they may have been unprepared or unwilling to raise (Bongaarts, 2016). While unplanned children confer many benefits, they are also costly to the household (e.g. refer to Robinson (1987); Becker and Tomes (1976); Barro and Becker (1989); Blundell et al. (1994)). To this end, a mistimed birth may affect how limited household resources and parental investments are allocated among children in the household, which may have significant implications for child health and human capital attainment, the effects of which, in turn, may persist into adulthood (Almond and Mazumder, 2013; Almond and Currie, 2011; Becker and Tomes, 1976; Adhvaryu and Nyshadham, 2016).

While access to family planning and reproductive health (FP/RH) services has improved in recent decades, the average contraceptive prevalence in Sub-Saharan Africa (22 percent) is less than half that of South Asia (51 percent) and less than a third that of East Asia (78 percent) (United Nations, 2018). Previous studies have shown that women from low-income and disadvantaged backgrounds form one of the largest groups that have an unmet need for family planning - that is, they are sexually active and report wanting to delay or stop childbearing but are not using a contraceptive method (Bradley et al., 2012; Casterline and Sinding, 2004; Westoff and Ochoa, 1991). This is partly because women may face financial and social barriers to accessing contraceptives and other FP/RH services (Haider and Sharma, 2013). To this end, improved access to FP/RH services may contribute to averting unintended births by allowing women and couples to more effectively meet their desired family size. Moreover, the use of contraception also enables women to time and space future pregnancies with more certainty (Casterline and Agyei-Mensah, 2017; Agesa and Agesa, 2020), which, in turn, may enable them to more effectively invest in their children's health, education, and well-being (Gipson et al., 2008; Singh et al., 2013).

In this study, we assess the causal impact of improved access to family planning on child health and human capital outcomes using experimental evidence from a randomized controlled trial in urban Malawi. Our investigation is motivated by classical models of the quantity-quality trade-off that were first developed by Becker (1960) and later extended by Schultz (1969, 2007). Most recently, work by Cavalcanti et al. (2021) adapts the Becker (1960) model by accounting for the uncertainty in fertility faced by households and assuming that families use contraception to reduce this uncertainty; the reduction in the cost to access family planning would, in turn, lead to increased investments in child health and human

capital by households. In addition, we are motivated by a limited evidence base that has documented the extent to which FP/RH services may contribute to improved child health by lengthening interpregnancy intervals and promoting healthy birth spacing (Cleland et al., 2012; Conde-Agudelo et al., 2006; Rutstein, 2006; Fink et al., 2014; Miller and Karra, 2020). This literature has identified potential physiological channels linking family planning and interpartum spacing to birth outcomes and early-life growth. In particular, the role of maternal nutritional depletion, which is linked to the close succession of pregnancies and periods of lactation, has been proposed as a key risk factor (King, 2003; Dewey and Cohen, 2007; Kozuki et al., 2013; Mayo et al., 2017; Molitoris, 2017; Miller, 1991). This evidence has served as the basis for the development of the World Health Organization's recommendation to women to wait at least 24 months after a live birth before becoming pregnant again (World Health Organization, 2005).

As part of our field experiment, 2,143 women who were either pregnant or up to six months postpartum at baseline were recruited in 2016. Following a baseline survey, women were randomized into either intervention or control arms, and women who were assigned to the intervention arm received the following package of services: 1) up to six free family planning counseling sessions; 2) free transport to a high-quality family planning clinic; and 3) financial reimbursement for family planning services received at the clinic, including for the treatment of contraceptive-related side effects. Two follow-up surveys were conducted in 2017 and 2018, respectively. For our analysis, we use anthropometric data collected for all children under six at baseline and the 2017 survey in combination with caregiver-reported survey data on children that was added to the 2018 survey, which measured cognitive functioning for all children under 3.

We focus our analysis on "index" children[1] who were alive to be measured at baseline, resulting in a potential sample of 1,034 children. We show that children born to women who received improved access to family planning through our package of services have better health and cognitive development outcomes. In particular, children born to women assigned to the intervention group were 0.34 standard deviations (SD) taller for their age relative to children born to women assigned to the control group after one year of exposure to the intervention. Children born to women assigned to the intervention group also performed 0.23 SD better on a caregiver-reported measure of cognitive development after two years of exposure to the intervention. While these results are promising, we must note that we do not find evidence of similar improvements in children's weights or blood hemoglobin levels.

Throughout the study, anthropometric measurement rates among eligible children are

[1]We define the child that resulted from the pregnancy or recent birth that made the woman eligible for the study as the index child.

low. Two factors drive these low rates. First, this study was conducted in a densely urban environment, Lilongwe, where children (particularly those of school age) have relatively high freedom of mobility and were often not at home at the time of the interview. Given that women were the primary study respondent for the main experimental intervention, the survey commenced as long as the primary respondent was present. Although attempts were made to revisit homes and measure eligible children, these efforts were not always successful. The second reason for low measurement is due to mothers' low rate of consent, which may, in part, have been driven by religious beliefs and mistrust. Among mothers in Lilongwe, skepticism around anthropometric measurement, even after having been informed of its minimal risks and potential benefits, significantly hampered data collection, particularly at endline when all anthropometric surveys were forced to pause amidst a national panic surrounding blood sampling and measurement. For this reason, we present findings from our midline follow-up survey where we have more complete data.

We cope with low levels of measurement in three ways. First, we present an analysis of attrition by levels and characteristics and show that, while some differences exist, attrition is not widely differential across treatment groups. Second, we employ propensity score weighting and Heckman selection models (Heckman, 1979) to correct for potential bias created by non-measurement. Finally, we present bounds in the tradition of Kling et al. (2007) to explore the severity with which non-measurement may impact our estimates. Using these three methods, we present evidence to indicate that the high rates of non-measurement do not bias our estimates. Nevertheless, our incomplete sample weakens our ability to make stronger causal claims throughout this study. We therefore view our estimates as a call for additional research into the relationship between improved access to family planning and child health outcomes.

Our study contributes to the limited and mixed evidence base on the relationship between family planning and child health. A Lancet series on Maternal and Child Malnutrition by Black et al. (2013), which put forward a framework of international action to alleviate global malnutrition and childhood stunting, denoted family planning as a "nutrition-sensitive" intervention. Although an extensive literature has examined the contributions of fertility decline on child welfare,[2] few studies focus directly on the impacts of family planning programs. To date, the best evidence of the effect of family planning on child health outcomes beyond child survival has predominantly been obtained from quasi-experimental evaluations of multi-component health interventions (Miller and Singer Babiarz, 2016). Analyses of the Matlab maternal and child health/family planning program found positive impacts of the intervention on child height, cognitive function, and educational attainment (Barham,

[2]For example, refer to DaVanzo et al. (2004); Li et al. (2008); Gertler and Molyneaux (1994).

2012; Joshi and Schultz, 2007), while an evaluation of health program placement from the Philippines observed a 7 percent increase in child height and a 12 percent increase in child weight from exposure to a family planning facility (Rosenzweig and Wolpin, 1986). In both programs, however, family planning was introduced as part of a larger bundle of maternal and child health and nutrition services, making it difficult to isolate the contribution of family planning exclusively. While a few observational studies have identified associations to suggest impact,[3] there is little empirical evidence of the causal link between increased access to family planning and early-life growth and development, particularly in low- and middle-income settings. Our experimental design allows us to more rigorously estimate these relationships and disentangle key behavioral and physiological pathways through which our family planning intervention impacts child health and human capital.

The rest of this study is presented as follows. Section 2 presents our theoretical motivation and conceptual framework. Section 3 describes the design of the field experiment. Section 4 presents our empirical strategy. Section 5 presents the results. Section 6 concludes. In the Appendixes, we include heterogeneous treatment effects of the intervention, an intervention component analysis, and a set of robustness checks to support the main findings that are presented.

# 2 Theoretical Framework

The study of the relationships between contraception, fertility, and child outcomes has received considerable attention in recent decades, much of which has been motivated by Becker (1960)'s groundbreaking "quality-quantity" model of fertility. In the Becker model, the impact of family planning and contraception on children's health and well-being is acknowledged directly through the trade-off that couples make when determining their family size (child quantity) and the level of investments that can be made in each child (child quality) (Becker, 1960). Interestingly, in his following adaptations of the model, Becker does not directly mention the role of contraception and family planning; instead, a couple's contraceptive behavior is assumed to be inseparable from their fertility outcome, which is perfectly deterministic (Becker and Tomes, 1976; Becker and Lewis, 1973). To this end, the role of family planning and contraception on child quantity and quality within the classic Becker model is entirely predicated on the fact that women and couples can perfectly target, and achieve, their realized fertility - families are either perfectly using contraceptives or have no demand for contraceptives. This assumption, in which the demand for children can be perfectly realized, is hard to reconcile with the fact that an estimated 74 million unwanted

[3]For example, Portner et al. (2011); Miller (2010).

pregnancies occur in developing countries each year (Bongaarts, 2016).

Several models, most recently by Cavalcanti et al. (2021), have extended the classic Becker model of fertility to allow for stochastic realizations of a couple's underlying fertility preferences and demand for children. The defining feature and key contribution of the Cavalcanti et al. (2021) model is that couples cannot perfectly choose the number of children they have; instead, they may face uncertainty in achieving their desired fertility, and this uncertainty can be mitigated through contraceptive use. As contraceptives are costly to a couple, the model predicts that when either contraceptive costs decrease or contraceptive effectiveness increases, couples will increase their level of household consumption and investments that they make in their children, thereby improving child health and development.

Following this prediction, we expect that our intervention may encourage investments in child health by: 1) decreasing the realized price of contraceptives through financial reimbursement and free transport; and 2) improving the perceived effectiveness of contraceptives through counseling. As couples increase their contraceptive use in response to improved access to services, they may be more effectively able to control their fertility. Couples also may be more certain of the time horizon until their next birth event, which may allow for greater human capital accumulation for their existing children if human capital and birth parity are substitutes in couples' preferences. The improved certainty over the timing of birth events allows couples to commit more credibly to human capital investments in children. Additionally, if in line with the predictions above, households consume more as a result of either reduced precautionary savings or reduced birth parity, this would be likely to indirectly improve child health outcomes through improved maternal nutrition or an improved home environment.

In adopting this theory of change, we rule out two other potential causal channels, namely those related to improved interpartum intervals and "marginal children". In Appendix A, we discuss how our choice of sample and observations from our data make these alternative channels unlikely. Further, by following Cavalcanti et al. (2021), we implicitly assume that couples use contraceptives primarily to control fertility and reduce uncertainty surrounding pregnancy. However, couples may use contraceptives for many reasons that are not directly related to the prevention or delay of pregnancy, including the prevention of sexually transmitted infections, menstrual regulation, or sexual well-being. In Appendix B, we utilize data on contraceptive and fertility intentions from our intervention and show that, in line with this assumption, the largest proportion of women report efficacy at preventing pregnancy as the feature that was most important to them when choosing a contraceptive method. We also show that women who were using contraceptives at baseline report a greater desire for birth spacing and lower fertility intentions than women who were not using contraceptives.

# 3 Study Design

Our empirical analysis is based on data from a randomized controlled trial conducted in Lilongwe, Malawi between November 2016 and November 2018. We provide an abbreviated description of the trial; a more detailed protocol describing the study design and intervention can be found in Karra and Canning (2020).

Women who were either pregnant or immediately postpartum (had a live birth within six months of the baseline screening) were recruited to participate in the study. Following a baseline survey implemented from September 2016 to January 2017, women were randomly assigned to either an intervention or control arm. Women assigned to the intervention arm received a comprehensive family planning package of services over a two-year period, which included: 1) an information brochure and up to six counseling visits from trained family planning counselors; 2) free transport to a high-quality private family planning clinic in Lilongwe; 3) financial reimbursement for family planning services, including for the treatment of contraceptive-related side effects. Annual follow-up surveys were conducted with women and children under age six who were available in the household at the time of the interview. Data collection for the first follow-up survey began in August 2017 and was completed in February 2018. Data collection for the second follow-up survey began in August 2018 and was completed in February 2019.

In this study, we use baseline and follow-up data on children to present findings related to the intervention's impact on the subsample of children who were alive at baseline and resulted from the pregnancy or recent birth that made their mothers eligible to be enrolled in the study (the "index" children). This choice of sample allows us to circumvent sample selection issues created by women who reported a pregnancy at baseline but for whom a live birth was not recorded, either due to a termination of pregnancy or initial misreporting by women.[4] Among these children, we test the extent to which the family planning intervention may have impacted these early-life linear growth patterns due to changes in parental investment behavior.

[4] As a robustness check, we expand our sample to include index children born during the study in Appendix C. The findings in this expanded sample are congruent with our main results. By focusing on children who were alive at baseline, we can identify baseline characteristics for those children whom we do not observe later in the study, which in turn allows us to estimate the extent to which attrition may bias our results.

## 3.1 Study Sample

### 3.1.1 *Eligibility*

Women were recruited to participate in the baseline survey if they were:[5]

1. Married;
2. Currently pregnant or had given birth within the previous six months;
3. Between the ages of 18 and 35;
4. A permanent resident of Lilongwe, Malawi;
5. Not sterilized nor had undergone a hysterectomy.

For women who were enrolled in the study, anthropometric data[6] was collected from children who were:

1. Under age six at baseline;
2. Identified as the biological or adopted child of the woman who was enrolled in the study;
3. Present in the household at the time of the interview.

Parental consent for collecting height, weight, and hemoglobin measurements was obtained from the woman for each of her eligible children.

### 3.1.2 *Sample Size and Randomization*

A total of 2,143 women were enrolled in the study in 2016. Following the baseline survey, women were randomly assigned to intervention and control groups by stratified covariate balanced randomization as proposed by Bruhn and McKenzie (2009). Women were allocated to strata based on the following baseline characteristics: number of living children, contraceptive use, age of sexual debut, level of educational attainment, work status, and their neighborhood of residence at baseline. In total, 1,026 women were assigned to the intervention arm, while 1,113 women were assigned to the control group. Additional details of the randomization protocol are presented in Karra et al. (2022).

For this study, we restrict the sample of children to those "index" children who were already born at baseline. Among this sample, there are 1,034 children,[7] 538 were born

[5] In addition to these inclusion criteria, no two women were enrolled from the same household. If more than one woman in the same household were eligible to participate in the study, the youngest eligible woman was recruited to participate. To minimize potential spillover effects between intervention and control groups, women chosen to be enrolled in the study were selected to be sufficiently distant from each other (at least five households apart).

[6] Heights, weights, and hemoglobin measures, as an indicator for anemia, were collected in all three survey waves. The study team collected hemoglobin measures using a rapid on-site blood diagnostic test (HemoCue). In the second follow-up survey, a set of questions were added to measure the cognitive development of the child.

[7] Among the women sampled for the study, 1,037 report being immediately postpartum at the time of

Table 1: Baseline Balance

| | (1) Full Sample | (2) Treatment | (3) Control | (4) Difference (3) - (2) |
|---|---|---|---|---|
| ***Sample Women*** | | | | |
| Current Use of FP (1 = Yes) | 0.489 | 0.492 | 0.486 | -0.006 |
| Long Acting Method Use (1=Yes) | 0.069 | 0.070 | 0.069 | -0.002 |
| Injectable Use (1=Yes) | 0.386 | 0.390 | 0.382 | -0.007 |
| Implant Use (1=Yes) | 0.065 | 0.064 | 0.065 | 0.001 |
| Ever Use of FP (1 = Yes) | 0.825 | 0.857 | 0.796 | -0.062*** |
| Woman's Age (Years) | 24.954 | 25.046 | 24.868 | -0.178 |
| Total Number of Children | 2.416 | 2.435 | 2.400 | -0.035 |
| Primary Education (1 = Yes) | 0.571 | 0.574 | 0.568 | -0.007 |
| Secondary Education (1 = Yes) | 0.429 | 0.426 | 0.432 | 0.007 |
| Tertiary Education (1 = Yes) | 0.029 | 0.024 | 0.033 | 0.009 |
| Religion (1 = Christian) | 0.836 | 0.855 | 0.818 | -0.037 |
| Ethnicity (1 = Chewa) | 0.415 | 0.423 | 0.409 | -0.014 |
| Woman Works (1 = Yes) | 0.083 | 0.080 | 0.085 | 0.005 |
| Age of First Cohabitation (Years) | 18.981 | 19.006 | 18.957 | -0.049 |
| **Observations** | **1,037** | **498** | **539** | |
| **P-Value (Joint Significance)** | | | | **0.462** |
| ***Sample Children*** | | | | |
| Child's Age (Months) | 2.909 | 2.901 | 2.915 | 0.014 |
| Child's Biological Sex (1=Male) | 0.492 | 0.504 | 0.481 | -0.023 |
| Child's Birth Order | 2.366 | 2.393 | 2.340 | -0.053 |
| Preceding Birth Interval (Months) | 51.731 | 52.401 | 51.046 | -1.355 |
| Height-for-Age Z-Score | -0.199 | -0.186 | -0.211 | -0.026 |
| Weight-for-Age Z-Score | 0.126 | 0.115 | 0.136 | 0.022 |
| **Observations** | **1,034** | **496** | **538** | |
| **P-Value (Joint Significance)** | | | | **0.276** |

$**p < 0.05$, $*p < 0.1$

Notes: For the top panel, the unit of analysis is a woman. For the bottom panel, it is a child. Our sample includes three women who report being immediately postpartum but whose children we do not record. These women are dropped from further analysis but included here for completeness. Stars are based on the critical value from individual t-tests.

to women assigned to the control group, while the remaining 496 children were born to women assigned to receive the family planning intervention. Table 1 presents key descriptive statistics for children enrolled in our study at baseline, as well as women who had already given birth to their "index" child. We see from Table 1 that apart from prior experience with contraception, where treatment women were slightly more likely to have ever used a

the baseline survey. However, for 3 of these women, their child was not available at the time of the baseline survey, and they are therefore excluded from the study.

contraceptive method relative to control women, women are generally balanced across key characteristics at baseline. We also note that there are no significant differences among children in the treatment group relative to the control group at baseline. Through joint significance tests, we provide additional evidence that children and women are not likely to be systematically different across intervention arms at baseline.

## 3.2 The Intervention

Women assigned to the treatment group were offered a comprehensive, multi-component postpartum family planning package over a two-year intervention period. The intervention was designed to overcome multiple barriers to access to care, including knowledge barriers as well as geographic and financial accessibility constraints.

The intervention included three main components. First, women were offered up to six free family planning counseling sessions over the intervention period. During these hour-long sessions, conducted by trained counselors, women received information on a full range of contraceptive methods, their potential side effects, and the health benefits of birth spacing. Second, Women were offered free transportation to the Good Health Kauma Clinic, a high-quality local service provider that offered clients a comprehensive list of family planning methods and related services.[8] The transport service was provided by a private taxi driver hired exclusively for the project and was accompanied at all times by a female field manager to help mitigate any social stigma. Finally, women assigned to the intervention arm received up to 17,500 MKW (∼$25.00 USD) in financial reimbursement for any costs incurred while receiving family planning services at the Kauma Clinic. Costs eligible to be reimbursed included those related to the procurement of family planning methods, family planning consultations, lab test fees, and exam fees.

Women assigned to the control arm received publicly available information on family planning methods and information about their nearest family planning clinic. The women were only contacted again at follow-up.

## 3.3 Findings From Previous Analyses

This study is the second in a series of analyses that present findings from the randomized controlled trial. As prescribed in our pre-analysis plan and the trial protocol (Karra and

[8]Services offered by the Kauma Clinic include insertion and removal of long-acting methods by trained clinicians, capacity for the treatment of contraceptive-related side effects and contraindications, referrals for sterilization, and additional counseling on family planning and methods. A full range of contraceptive methods was available at the Kauma Clinic, there were no reported stockouts of methods, and waiting times at the clinic were reported to be low.

Canning, 2020), child growth and development are identified as secondary outcomes that provide evidence for the potential long-run causal impact of improved access to family planning. As such, the trial was powered to detect effects in contraceptive use, contraceptive method mix, and short birth intervals. While these outcomes are beyond the current scope of this study, the theory that we present is predicated on the condition that FP/RH services provide women with the opportunity to exercise greater reproductive control. Indeed, Karra et al. (2022) shows that women assigned to the intervention arm were 5.9 percentage points more likely to be using contraceptives after two years of exposure to the intervention. The effect seems particularly strong among women who were immediate postpartum at the time of recruitment; these women were 7.2 percentage points more likely to be using contraception. Moreover, the increase in contraceptive use seem to be driven by an increase in demand for long-acting reversible methods, as we observe a 4.6 percentage point increase in contraceptive implants.

While the impacts that we observe on contraceptive use are in line with our prior predictions, the strongest results that we find in our previous analysis, which are most relevant for this study, are those that examine intervention impact on inter-birth intervals. A survival analysis finds that women assigned to the intervention arm were 43 percent less likely to have a second pregnancy within 24 months of their index birth at baseline. Together, these results show that FP/RH services allowed women to exercise greater control over their contraceptive use, which in turn impacted their likelihood of experiencing a subsequent pregnancy. These findings suggest that the impact of the intervention is likely to extend to downstream child growth and development.

# 4 Empirical Strategy

## 4.1 Key Outcomes

In this paper, we study the impact of our intervention on four main outcomes. Below, we describe how each is calculated.

### 4.1.1 Height

The main outcome of this study is the impact of family planning services on children's height-for-age Z-score (HAZ). A child's HAZ is calculated as the number of standard deviations their height is from a healthy reference or standard child of the same age and sex. The reference height distribution is derived from the WHO Multicentre Growth Reference Study (MGRS), which measured the heights of healthy children around the world to determine

standard growth patterns (de Onis et al., 2004). This measure serves as a standard metric of health and nutrition and, as a reflection of the path dependency of height, are sensitive to nutrition and health shocks that may affect growth (Victora et al., 2010; Hoddinott et al., 2013).

In addition to HAZ, we report treatment effects on moderate and extreme stunting in children. A child is considered to be moderately stunted if their height is more than two standard deviations below the height of an average reference child of the same age and sex. A child is considered extremely stunted if her height is more than three standard deviations less than the height of an average reference child of the same age and sex.

#### 4.1.2 Weight

While linear growth patterns serve as a meaningful proxy for a child's cumulative health, they are slow to change. For short-term shocks to a child's health, Weight-for-Height z-score (WHZ) is often used as a proxy. Given that weights change more readily than heights, children's weight, relative to their height, is more responsive to shocks to the child's health or nutrition, such as a prolonged period of fasting or diarrheal disease. As such, we accompany our results on HAZ with results on WHZ, which is similarly measured as the number of standard deviations a child's weight is from a healthy reference or standard child of the same age and height.

#### 4.1.3 Hemoglobin

Meanwhile, we may observe changes in children's growth patterns if we observe changes in their anemia status. Iron deficient anemia, defined by insufficient hemoglobin levels, has been shown to affect children's cognitive and physical development, leading to decreased productivity in the long run (Grantham-McGregor and Ani, 2001). As such, we also present results on children's hemoglobin levels, measured, in grams per milliliter, of hemoglobin in a child's blood.

#### 4.1.4 Child Cognitive Measures

Finally, as early life growth patterns are linked to cognitive development in children, we report the effect of the family planning intervention on a composite measure of cognition created by the Caregiver Reported Early Development Instruments (CREDI) (Fink et al., 2018). The CREDI instrument was designed to produce a population-level measurement of early childhood development for children under age three and consists of sets of questions

that are asked to caregivers of children of different ages up to age 3. This instrument focuses exclusively on key child development milestones and behaviors reported by the caregiver.[9]

## 4.2 Intent-to-Treat Analysis

Our main results present findings from an Intent-to-Treat (ITT) analysis, which identifies the average treatment effect (ATE) on children born to mothers in the treatment group relative to the control group. The ITT is preferred, in our case, to a local average treatment effect (LATE) specification, which represents the ATE for children born to mothers who utilized the program and would likely overestimate the program's effect due to selection into program uptake. Given that not every woman will decide to participate in the program even if such a program were ever taken to scale, an ITT specification would better reflect the policy relevance of integrating family planning services into a larger campaign to ameliorate stunting. On the other hand, a LATE analysis would allow us to infer which intervention component drives our result. In Appendix D, we discuss some of the causal challenges we face when disaggregating component effects and present correlational evidence on this question.

For our ITT analysis, we present results from the following specification:

$$Y_{im} = \alpha + \beta_T T_m + \delta MOB_{im} + \gamma f(age_{im}) + \mathbf{X}_{im}\lambda + \mathbf{Z}_m\zeta + \chi_m + \varepsilon_{im} \tag{1}$$

In Equation 1, $Y_{im}$ is the outcome of interest for child $i$ born to mother $m$. This study presents findings on children's height-for-age z-score, stunting (a binary outcome), weight-for-height z-scores, hemoglobin levels, and cognition scores. $T_m$ is an indicator variable representing the child's mother's assignment to treatment, and $\beta_T$, the impact of the intervention, is our coefficient of interest. $MOB_{im}$ is an indicator variable for the child's birth month. $\mathbf{X}_{im}$ is a vector of child characteristics measured at baseline, including biological sex and birth order, and $\mathbf{Z}_m$ is a vector of maternal characteristics measured at baseline, including age, work status, education, religion, ethnicity, total number of children, ever use of family planning, and age of sexual debut. $\chi_i$ is a neighborhood fixed effect defined by the sampled enumeration areas within Lilongwe. $f(age_i)$ captures a flexible specification over the child's age, and the results we present are for a spline with knots at 6, 12, 18, 24, and 30 months. This specification allows our model to account for the biological changes in linear growth patterns within the early years of life (Cummins, 2013). In the presentation of our results, we present $\beta_T$ estimated by a naïve model containing only treatment assignment, one in which

[9]Since the survey is exclusively administered to the caregiver and involves no direct interaction with the child, it was not subject to the suspension of anthropometric measurement in 2017; as a result, we have sufficient data to report findings from second-year follow-up survey.

we add the age-specific controls $MOB_{im}$, and finally, the fully adjusted model presented in Equation 1. In each of these specifications, we cluster standard errors by mother to account for within-mother correlations among non-singleton births.

## 4.3 Multiple Hypothesis Testing

We run Equation 1 on several child health and cognitive outcomes. To correct for the testing of multiple hypotheses, we present frequentist q-values that are adjusted for the False Discovery Rate (FDR) associated with that estimation (Storey and Tibshirani, 2003; Newson, 2010). We calculate these frequentist q-values using the method described in Simes (1986). Q-values are calculated by estimation and sample (unadjusted vs. unadjusted and index sample vs. expanded sample) and are reported alongside the more traditionally calculated p-values.

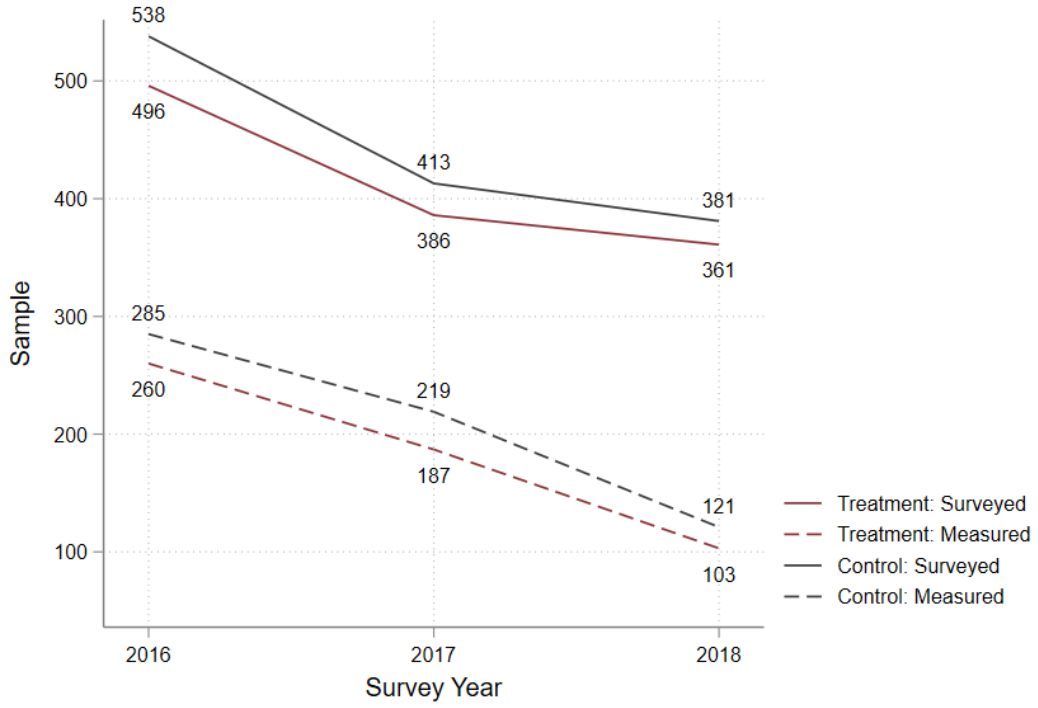
## 4.4 Attrition Adjustment

From the onset of our study, we observe high rates of non-measurement for anthropometric outcomes among children in our sample. At baseline we can capture anthropometric data for only 52.7 percent of children, providing us with a sample of 545 children in total. These measurement issues persist over time, and we can only obtain anthropometric data for 406 children during the first-year follow-up for our study. Finally, during our endline survey, we were forced to pause our anthropometric data collection, particularly the administration of on-site blood testing for anemia, out of concern for the safety of our enumerators.[10] For this reason, although we were able to collect some anthropometric data at endline, we report treatment effects using data from the more complete first-year follow-up. Figure 1 breaks down these sample metrics by intervention arm.

Of the children who were unmeasured at first-year follow-up, seven children had died before the survey, 113 children were not at home at the time of the interview, 146 children did

[10]In September and October 2017, at least five people were killed by lynch mobs who accused them of vampirism. News sources reported that mobs searching for accused vampires in communities had been mounting roadblocks, which raised safety and security concerns throughout the country (Reuters, 2017). In response to these rumors, the United Nations, international NGOs, and other institutions in Malawi withdrew many staffers from southern districts and suspended any research-related collection of blood samples from respondents. For these reasons, children interviewed after the suspension of anthropometric data collection were not measured.

Figure 1: Sample Description



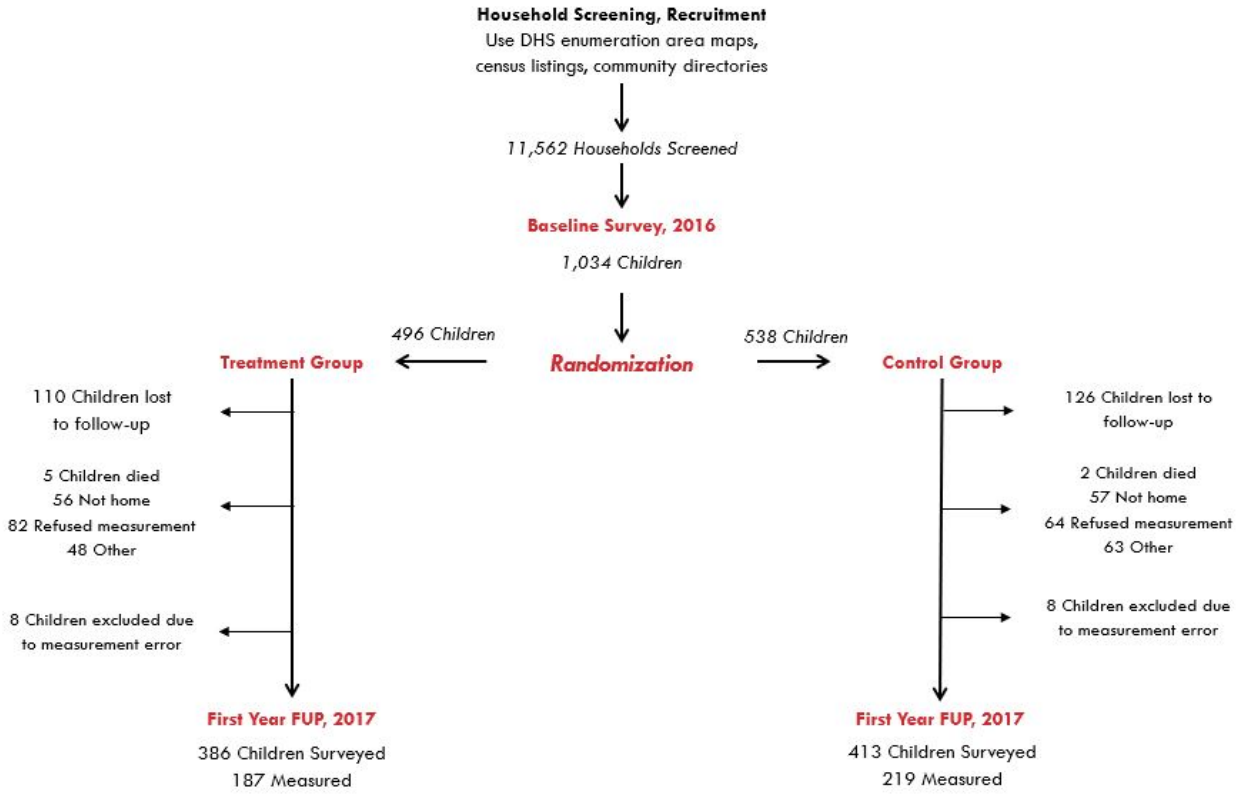

not have consent granted by their parents, and 111 were not measured for other reasons.[11,12] Figure 2 presents a flowchart combining the discussion throughout this section thus far, demonstrating the randomization protocol and identifying the reasons for attrition and non-measurement by treatment groups during the first-year follow-up survey.

With attrition rates this high, we cannot simply present the intent-to-treat (ITT) results and consider them to be causal estimates of the intervention on outcomes. High attrition rates may bias our results in two ways. First, if measured children are systematically different across treatment groups, the differential non-measurement would invalidate our randomization and introduce the possibility of omitted variable bias. This type of systematic attrition is the more severe of the two possible biases we explore because it would threaten *internal validity* since we may no longer have valid estimates of the causal impact of family planning.

[11]Based on our observations from data collection, we note that a majority of the "other" instances of non-measurement resulted when the child refused to cooperate with the field enumerator after their parent had granted consent for measurement. Surveyors on our field team were informed of the importance of measurement and were instructed to make every attempt to work with the child to be measured; however, no child was forced into measurement, and the study team was requested to cease measurement after three attempts.

[12]In addition to those not measured, a small group of 16 children were excluded from the sample because their Z-scores fell outside of the WHO recommended exclusion criteria due to measurement error in age.

Figure 2: Experimental Design and Attrition Tracking




In addition, we may observe that the characteristics of the mother or the child systematically determine the probability of non-measurement. That is to say that the attrition is not necessarily different across treatment groups but that those children who are measured are systematically different than those who are not. While this form of potential bias is less severe, it would still change the interpretation of our results. In the case where attrition is systematically determined by characteristics but not systematically different across groups, our *external validity* would be threatened, whereby we would only have valid causal estimates for the subsample of children for whom we have data.

In section 5.4, we explore the threat posed by attrtion to the validity of our estimates by presenting an analysis of this attrition in both levels and characteristics. We show that while our initial baseline balance is largely unchanged, suggesting that the validity of our randomization remains intact, there is evidence that children who are unmeasured were born to mothers who are younger, less educated, and less likely to be using contraception. In following a large literature on attrition adjustment in randomized trials, we cope with these findings by adjusting our ITT estimates using a number of attrition adjustment methods

described below.

### 4.4.1 Inverse Probability Weighting

In following Macours and Molina Millán (2017), we first correct for sample attrition using a propensity score weighting approach. This approach allows us to correct for imbalances due to sample attrition of study women, which may be correlated with observable woman-level characteristics such as education and age. We accompany our main ITT specification with estimates from an inverse probability weighting adjustment using the estimated propensity scores as weights. By creating the propensity scores that we use to adjust our estimates, we also create the opportunity to further characterize the attrition that we observe in the sample. As such, in addition to presenting a balance table of characteristics among attritiors and non-attritiors, section 5.4 presents the results from the selection models we use to create our propensity scores.

### 4.4.2 Heckman Selection Model

As much of our attrition is driven by low participation, we may be concerned that it is correlated with some non-observable characteristic that determines measurement across waves. To account for this possibility, we accompany our main ITT estimates with estimates from a Heckman selection model (Heckman, 1979; Koné et al., 2019). We model missingness as a function of the enumerator assigned to interview the household under the assumption that measurement is, at least in part, driven by enumerator ability; this approach to use variation in enumerator assignment has been implemented in previous studies[13] (Hogan et al., 2012; Bärnighausen et al., 2011).

### 4.4.3 Bounding

Finally, a large literature has proposed to account for attrition using bounding methods, which make assumptions about the potential outcomes of missing data and estimate a range of coefficients using this assumed data (Macours and Molina Millán, 2017). We present bounds of the Kling-Liebman type (Kling et al., 2007), which assume that missing outcomes are within a given number of standard deviations from the within-intervention arm mean. We calculate bounds under two sets of assumptions; 1) we assume missing data takes the outcome within 0.1 standard deviation away from the group mean; and 2) in a more conservative case, we assume that the missing data are within 0.2 standard deviations of the group mean. Both

[13]During training, our enumerators reported difficulty in to agree to sit still for measurement. Our model's identification assumes that some enumerators are better at administering consent to participate.

of these assumptions follow the example of Özler et al. (2021) to correct for sample bias due to repatriation and third-country relocation in Turkish refugee camps.

# 5 Results

## 5.1 Height-for-Age Z-Scores

To measure the effects of improved access to family planning on linear growth in children, we report the treatment effect on HAZ, as described in Section 4.1.1.

Table 2: Results on Children's Heights, Year 1

| | Unadjusted Estimates | | Bounds and Adjusted Estimates | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) +/- 0.2 SD | (4) +/- 0.1 SD | (5) Heckman | (6) Adjusted Estimate | (7) IPW | (8) +/- 0.1 SD | (9) +/- 0.2 SD |
| Panel A: Height-for-Age Z-Score | | | | | | | | | |
| Treatment | 0.284* | 0.308* | -0.0624 | 0.123** | 0.340** | 0.339** | 0.316* | 0.492*** | 0.677*** |
| | (0.155) | (0.157) | (0.0628) | (0.0621) | (0.151) | (0.162) | (0.162) | (0.0617) | (0.0620) |
| | [0.119] | [0.090] | | | | [0.064] | | | |
| Control Mean | -1.67 | -1.67 | | | | -1.67 | | | |
| Panel B: Moderate Stunting (HAZ $\leq$ -2) | | | | | | | | | |
| Treatment | -0.107** | -0.114** | 0.00617 | -0.0536*** | -0.119** | -0.120** | -0.118** | -0.173*** | -0.233*** |
| | ‘(0.0491) | ‘(0.0499) | (0.0199) | (0.0197) | (0.0489) | (0.0519) | (0.0548) | (0.0196) | (0.0198) |
| | [0.071] | [0.052] | | | | [0.051] | | | |
| Control Mean | 0.42 | 0.42 | | | | 0.42 | | | |
| Panel C: Extreme Stunting (HAZ$\leq$ -3) | | | | | | | | | |
| Treatment | -0.0276 | -0.0294 | 0.0610*** | 0.0154 | -0.0411 | -0.041 | -0.0336 | -0.0757*** | -0.121*** |
| | (0.0377) | (0.0383) | (0.0151) | (0.0149) | (0.0372) | (0.0400) | (0.0445) | (0.0149) | (0.0150) |
| | [0.478] | [0.516] | | | | [0.430] | | | |
| Control Mean | 0.17 | 0.17 | | | | 0.17 | | | |
| Age and Birth Month Controls | | X | X | X | X | X | X | X | X |
| Mother and Child Characteristics | | | X | X | X | X | X | X | X |
| Observations | 406 | 406 | 1,037 | 1,037 | 404 | 404 | 399 | 1,037 | 1,037 |

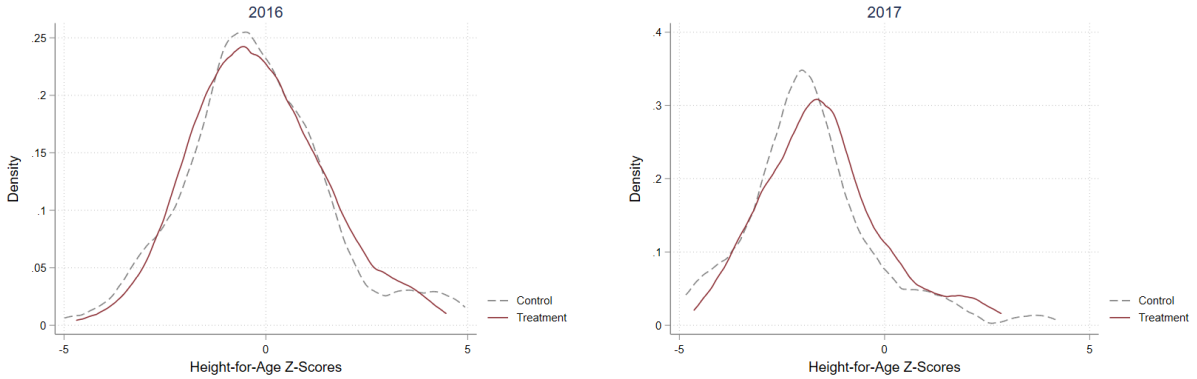
$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. OLS estimates are accompanied by estimates from an inverse probability weighting model (Column 7), Heckman-type selection model (Column 5), and Kling-Leibman-type bounds (Columns 3, 4, 8, & 9). The adjusted regressions (Column 6) include woman-level controls such as use of family planning methods, ever use of family planning methods, the total number of children who are alive (included as a set of dummy variables), educational attainment of the woman (primary or less versus secondary and higher), age of the woman (in three age groups), age of sexual debut, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut and child-level controls such as biological sex, birth order (included as a set of dummy variables), birth month fixed effects and a flexible form over the child's age. All controls are included at their baseline value. Neighborhood fixed effects are included and standard errors, clustered at the mother level, are reported in parentheses. Sharpened q-values are presented in brackets.

Table 2 presents the unadjusted (Columns 1 & 2) and adjusted (Column 6) ITT estimates for the treatment effect on child height-for-age Z-scores, moderate stunting, and extreme stunting during the first year. The ITT effects are presented alongside estimates from the

inverse probability weighting model (Column 7), Heckman-type selection model (Column 5), and Kling-Leibman-type bounds (Columns 3, 4, 8, & 9). In addition to Table 2, Figure 3 presents non-parametric density distributions for children's height-for-age.

Figure 3: Non-Parametric Density Distributions of Children's Height-for-Age Z-Scores




Notes: Density estimations are produced using a Gaussian Kernel and the rule-of-thumb optimal bandwidth. The left panel displays the distributions of children at baseline by treatment group. The right panel displays children at first-year follow-up

The results from Figure 3 and Table 2 suggest that improved access to family planning led to a 0.28-0.34 standard deviation increase in children's height for age and a 10.7-12.0 percentage point decrease in moderate stunting. To explain this result, we refer back to Figure 3, which shows that the modal child in the control group is only slightly above the moderate stunting threshold at the first year follow-up.[14] This means that the increase in observed height-for-age pushes a potentially large number of children over the stunting threshold. While it is unlikely that there are significant differences in the health of a child who is just above the stunting threshold compared to a child who is just below the threshold, our findings suggest that family planning may play a more significant role in the larger effort to end childhood stunting than what has previously been speculated (Fink and Rockers, 2014).

Based on our models of attrition adjustment, we find the potential impact of attrition bias on our estimates to likely be minor. Both the inverse probability and Heckman-type models report coefficients that are very similar to our adjusted estimate.[15] Meanwhile, our results are robust to the moderate version of our assumptions on the Kling-Leiban-type

[14]Consistent with findings from previous studies (Roche and Himes, 1980), we observe a leftward shift in the distribution of z-scores between waves. Other studies have shown that children's HAZ tends to decrease over the first two years of life before evening out and allowing for possible catch-up.

[15]While the inverse probability coefficient for HAZ is only statistically significant to the 10 percent level, we note that the corresponding t-statistic associated with this estimate is 1.95

bounds. Under the more conservative assumptions, our estimates change in sign but lose statistical significance.

In Appendix E, we study potential heterogeneous effects of our intervention. We show that the treatment effects may be larger among male children, children whose parents were interested in having another child, and children with older siblings. In Appendix F, we reconcile our results with the motivating theory by conducting a causal mediation analysis to determine the channels driving our results. Due to our modest sample size, we lack the statistical power to make concrete statements about the underlying causal mechanisms driving our results. With this said, we find evidence to suggest that changes in health investments and birth spacing may partially mediate the results.

## 5.2 Weight-for-Height Z-Scores & Hemoglobin Levels

We present estimates of the treatment effect of our intervention on children's WHZ and hemoglobin levels, as described in Sections 4.1.2 and 4.1.3 respectively, in Table 3. Our main estimates do not find evidence of a treatment effect on children's weights or hemoglobin levels. Given the imprecision of the estimates, it is difficult to draw conclusions from these estimates. However, studying the point estimates, we observe a negative point estimate on children's WHZ. It is perhaps unsurprising that we do not observe a treatment effect of our intervention on children's WHZ as we expect family planning services to act, in this context, through increased investments in children. It is unlikely, however, that family planning services increase household resiliency to shocks or the sanitary environment in which children live, which we would expect to reduce incidences of underweight.

Conversely, we observe a positive coefficient on hemoglobin level. It is certainly possible that there is a positive impact of family planning services here that we are underpowered to detect. However, to say this definitively, we would need a substantially larger sample size, given the imprecision of the estimate.

## 5.3 Cognitive Development

Table 4 presents unadjusted and adjusted estimates of the impact of our intervention on child CREDI scores and on internally standardized scores of cognition.[16] Table 4 is accompanied by Figure 4, which presents non-parametric density distributions of children's CREDI Z-scores by treatment group.

[16]These Z-scores are calculated by standardizing the distribution of CREDI scores in the control group. The control distribution is preferred to the distribution of all children because it does not include any possible treatment effects induced by our intervention.

Table 3: Results on Weight-for-Height Z-Scores & Hemoglobin Levels, Year 1

| | Unadjusted Estimates | | Bounds and Adjusted Estimates | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) +/- 0.2 SD | (4) +/- 0.1 SD | (5) Heckman | (6) Adjusted Estimate | (7) IPW | (8) +/- 0.1 SD | (9) +/- 0.2 SD |
| Panel A: Weight-for-Height Z-score | | | | | | | | | |
| Treatment | -0.152 | -0.181 | -0.523*** | -0.338*** | -0.205 | -0.145 | -0.224 | 0.0314 | 0.216*** |
| | (0.153) | (0.155) | (0.0622) | (0.0617) | (0.184) | (0.162) | (0.195) | (0.0618) | (0.0624) |
| | [0.450] | [0.338] | | | | [0.431] | | | |
| Control Mean | 0.503 | 0.503 | | | | 0.503 | | | |
| Observation | 409 | 409 | 1037 | 1037 | 408 | 408 | 413 | 1037 | 1037 |
| Panel B: Hemoglobin (g/L) | | | | | | | | | |
| Treatment | 0.11 | 0.0785 | -0.280*** | -0.0890 | 0.0757 | 0.117 | 0.0631 | 0.294*** | 0.485*** |
| | (0.155) | (0.156) | (0.0574) | (0.0569) | (0.154) | (0.165) | (0.167) | (0.0570) | (0.0575) |
| | [0.478] | [0.614] | | | | [0.478] | | | |
| Control Mean | 9.83 | 9.83 | | | | 9.83 | | | |
| Observations | 374 | 374 | 1037 | 1037 | 374 | 374 | 368 | 1037 | 1037 |
| Age and Birth Month Controls | | X | X | X | X | X | X | X | X |
| Mother and Child Characteristics | | | X | X | X | X | X | X | X |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. OLS estimates are accompanied by estimates from an inverse probability weighting model (Column 7), Heckman-type selection model (Column 5), and Kling-Leibman-type bounds (Columns 3, 4, 8, & 9). The adjusted regressions (Column 6) include woman-level controls such as use of family planning methods, ever use of family planning methods, the total number of children who are alive (included as a set of dummy variables), educational attainment of the woman (primary or less versus secondary and higher), age of the woman (in three age groups), age of sexual debut, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut and child-level controls such as biological sex, birth order (included as a set of dummy variables), birth month fixed effects and a flexible form over the child's age. All controls are included at their baseline value. Neighborhood fixed effects are included and standard errors, clustered at the mother level, are reported in parentheses. Sharpened q-values are presented in brackets.

We find that exposure to the family planning intervention increased CREDI scores in children by 0.19-0.23 standard deviations. Figure 4 suggests that these mean treatment effects may be driven by a reduction of the number of children with lower scores and a lower kurtosis of the distribution, implying a more centralized distribution in the treatment group. As opposed to an increase in highly positive scores, these results suggest a lower rate of developmental delays, although the CREDI instrument is not specifically designed to detect these delays. Again, our attrition-adjusted results indicate that any selection bias created by our high attrition rate is minimal, and our estimates do not change substantively in response to adjustment.

Once again, we further explore our results in Appendix E and Appendix F. In Table E1, we show that the treatment effects on CREDI scores may be larger among female children, children whose parents were not interested in having another child, and children with older siblings. Interestingly, these findings contrast with those on children's heights. Further, in Table F2, we extend our mediation analysis to CREDI scores. However, we do not detect any mediating effects, possibly due to our limited sample size.

Table 4: Results on CREDI Scores, Year 2

| | Unadjusted Estimates | | Bounds and Adjusted Estimates | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) +/- 0.2 SD | (4) +/- 0.1 SD | (5) Heckman | (6) Adjusted Estimate | (7) IPW | (8) +/- 0.1 SD | (9) +/- 0.2 SD |
| Panel A: CREDI Scores | | | | | | | | | |
| Treatment | 0.568*** (0.212) [0.027] | 0.590*** (0.195) [0.009] | 0.0841 (0.0760) | 0.342*** (0.0751) | 0.694*** (0.190) | 0.693*** (0.218) [0.006] | 0.682*** (0.252) | 0.859*** (0.0747) | 1.118*** (0.0753) |
| Control Mean | 54.6 | 54.6 | | | | 54.6 | | | |
| Panel B: CREDI Z-Score | | | | | | | | | |
| Treatment | 0.191*** (0.0716) [0.027] | 0.199*** (0.0656) [0.009] | 0.0283 (0.0256) | 0.115*** (0.0253) | 0.234*** (0.0640) | 0.233*** (0.0736) [0.006] | 0.230*** (0.0851) | 0.290*** (0.0252) | 0.377*** (0.0254) |
| Control Mean | 0.54 | 0.54 | | | | 0.54 | | | |
| Age and Birth Month Controls | | X | X | X | X | X | X | X | X |
| Mother and Child Characteristics | | | X | X | X | X | X | X | X |
| Observations | 362 | 362 | 1,037 | 1,037 | 362 | 362 | 357 | 1,037 | 1,037 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. OLS estimates are accompanied by estimates from an inverse probability weighting model (Column 7), Heckman-type selection model (Column 5), and Kling-Leibman-type bounds (Columns 3, 4, 8, & 9). The adjusted regressions (Column 6) include woman-level controls such as use of family planning methods, ever use of family planning methods, the total number of children who are alive (included as a set of dummy variables), educational attainment of the woman (primary or less versus secondary and higher), age of the woman (in three age groups), age of sexual debut, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut and child-level controls such as biological sex, birth order (included as a set of dummy variables), month of birth fixed effects and a flexible form over the child's age. All controls are included at their baseline value. Neighborhood fixed effects are included and standard errors, clustered at the mother level, are reported in parentheses. Sharpened q-values are presented in brackets.

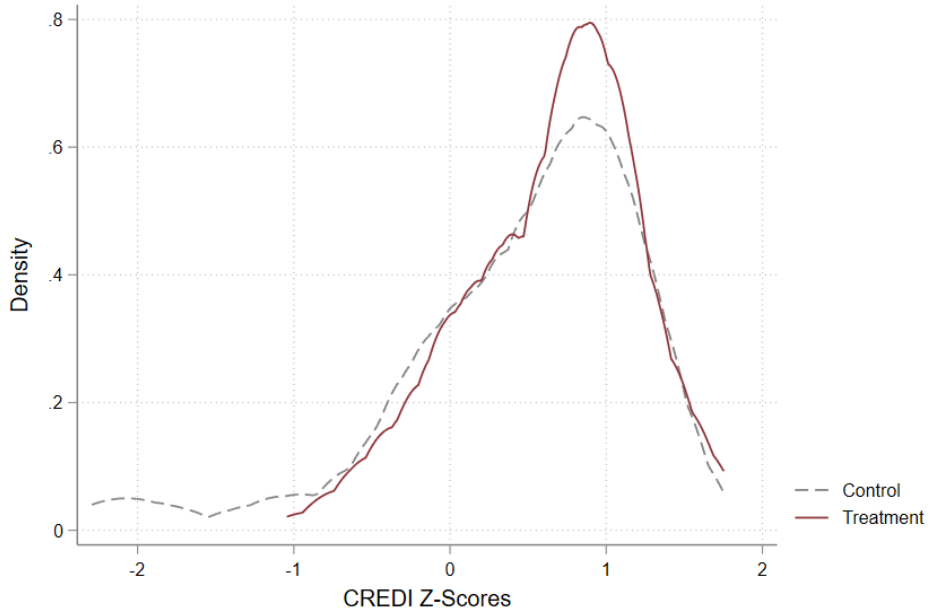
## 5.4 Attrition Analysis

Throughout this section, we have adjusted our results to account for threats posed to our analysis by a high rate of attrition that is a key feature of our sample. This attrition comes in two forms. First, a non-negligible 22.8 percent of children are not observed at any time at first-year follow-up. While this amount of attrition may be considered standard in an urban setting, among these children, we only have outcome data for 406 children of the 799 children who we observe,[17] creating a measurement rate of only 50.8 percent.

To explore the threats discussed in section 4.3, Table 5 compares the level of attrition by baseline characteristics of mothers and children by measurement status and by treatment groups at the first-year follow-up. We observe that among measured children, the pattern of balance across treatment groups remains largely unchanged, with mothers of children

[17] In this study, we present results for three different anthropometric outcomes: heights, weights, and hemoglobin. There are slightly different levels of missingness for each of these variables; we observe heights for 406 children, weights for 409 children, and hemoglobin levels for 374 children. The rate of hemoglobin measurement is slightly lower, likely because the process is more demanding on the child, and they were less likely to consent to measurement. In this section, we conduct our analysis of the missingness in heights.

Figure 4: Non-Parametric Density Distributions of Children's CREDI Z-Scores




Notes: Density estimations are produced using a Gaussian Kernel and the rule-of-thumb optimal bandwidth. The left panel displays the distributions of children at baseline, by treatment group. The right panel displays children at second-year follow-up

reporting greater contraceptive experience. However, we do not observe jointly significant differences in the baseline characteristics of children measured at first-year follow-up. This is also the case for children who are not measured. We observe that children who were not measured in the treatment group seem to have been born to slightly older mothers than those not measured in the control group. However, much like for those children who were measured, we do not observe jointly significant differences across characteristics. The inability to detect differences across intervention arms among measured or non-measured children means we cannot reject the hypothesis that the attrition we observe is balanced across groups. This fact is suggestive evidence that the estimates we present are *internally valid*, in lieu of attrition adjustment.

While the differences that we observe *across* treatment groups are promising for our aims of causal estimation, the differences that we observe *within* treatment groups do give a reason for pause. Within the treatment arm, we observe no statistically significant differences among baseline characteristics for those children who were measured at first-year follow-up versus those who were not. However, within the control group, we observe moderately-sized differences in mothers' contraceptive use and age. Further, our results suggest that the baseline characteristics of children who were measured versus those who were not at

Table 5: Attrition Balance Table

| | Remaining Participants | | Lost to Follow-up | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Treatment (N=185) | Control (N=216) | Treatment (N=311) | Control (N=322) | Difference (1) - (2) | Difference (1) - (3) | Difference (2) - (4) | Difference (3) - (4) |
| | (1) | (2) | (3) | (4) | | | | |
| ***Maternal Characteristics*** | | | | | | | | |
| Current Use of FP (1 = Yes) | 0.497 | 0.417 | 0.489 | 0.525 | 0.081 | 0.009 | -0.108** | -0.036 |
| Long Acting Method Use (1=Yes) | 0.059 | 0.088 | 0.074 | 0.059 | -0.029 | -0.014 | 0.029 | 0.015 |
| Injectable Use (1=Yes) | 0.395 | 0.310 | 0.386 | 0.419 | 0.084* | 0.009 | -0.109** | -0.033 |
| Implant Use (1=Yes) | 0.054 | 0.074 | 0.068 | 0.059 | -0.020 | -0.013 | 0.015 | 0.009 |
| Ever Use of FP (1 = Yes) | 0.886 | 0.796 | 0.839 | 0.795 | 0.090** | 0.047 | 0.001 | 0.044 |
| Woman's Age (Years) | 25.119 | 25.606 | 25.161 | 24.335 | -0.488 | -0.042 | 1.271*** | 0.825** |
| Total Number of Children | 2.443 | 2.537 | 2.479 | 2.329 | -0.094 | -0.036 | 0.208 | 0.150 |
| Primary Education (1 = Yes) | 0.535 | 0.537 | 0.588 | 0.590 | -0.002 | -0.053 | -0.053 | -0.002 |
| Secondary Education (1 = Yes) | 0.465 | 0.463 | 0.412 | 0.410 | 0.002 | 0.053 | 0.053 | 0.002 |
| Tertiary Education (1 = Yes) | 0.038 | 0.046 | 0.019 | 0.025 | -0.008 | 0.019 | 0.021 | -0.006 |
| Religion (1 = Christian) | 0.870 | 0.838 | 0.849 | 0.804 | 0.032 | 0.021 | 0.034 | 0.045 |
| Ethnicity (1 = Chewa) | 0.432 | 0.417 | 0.419 | 0.408 | 0.016 | 0.013 | 0.009 | 0.011 |
| Woman Works (1 = Yes) | 0.076 | 0.088 | 0.084 | 0.084 | -0.012 | -0.008 | 0.004 | -0.000 |
| Age of First Cohabitation (Years) | 19.120 | 19.223 | 19.016 | 18.800 | -0.104 | 0.103 | 0.423* | 0.216 |
| ***Child Characteristics*** | | | | | | | | |
| Child Age (Months) | 2.751 | 2.791 | 2.991 | 2.999 | -0.040 | -0.240 | -0.208 | -0.008 |
| Biological Sex (1=Male) | 0.470 | 0.500 | 0.524 | 0.469 | -0.030 | -0.054 | 0.031 | 0.055 |
| Child's Birth Order | 2.384 | 2.435 | 2.399 | 2.276 | -0.051 | -0.015 | 0.159 | 0.122 |
| **P-Value (Joint Significance)** | | | | | **0.435** | **0.738** | **0.025**** | **0.315** |
| **Observations (Joint Significance)** | | | | | **399** | **493** | **534** | **628** |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. Stars are based on the critical value from individual t-tests. Joint tests are computed using only observations for which we possess full data on all characteristics.

first-year follow-up are jointly different. These differences may suggest two things. First, they may suggest that those children who were measured were systematically different than those who were not, regardless of their assigned intervention arm. Given the context and hesitancy with measurement, it is certainly possible that measurement refusals correlate with a mother's characteristics.

As discussed in section 4.3, differences of the kind that we observe may threaten the *external validity* of our estimates, meaning that any unadjusted causal effects that we may find cannot be considered an unbiased estimate of the treatment effect on the study sample. Above, we have confronted this possibility by utilizing propensity score weighting to create a sample-representative estimate of the treatment effect. Table 6 present the results of the logit model that we use to estimate these propensity scores. In addition to creating weights for our estimates, this methodology allows us to expand the investigation of our sample attrition beyond the balance table presented in Table 5.

We see from Table 6 that children in the treatment group were less likely (have a lower odds ratio) to be measured than those in the treatment group. We cannot, however, reject the hypothesis that the probability of measurement is equal across intervention arms. Extending

Table 6: Selection into Attrition

| | (1) Height-for-Age Z-Scores | (2) Weight-for-Age Z-Scores | (3) Hemoglobin | (4) Cognition Z-Scores |
|---|---|---|---|---|
| Treatment | 0.847 | 0.853 | 0.854 | 0.907 |
| | (0.115) | (0.115) | (0.117) | (0.128) |
| Measured in 2016 | 1.764*** | 1.685*** | 2.359* | |
| | (0.243) | (0.233) | (1.046) | |
| Woman Works | 0.928 | 0.878 | 0.764 | 0.970 |
| | (0.235) | (0.222) | (0.200) | (0.251) |
| Ever Use of Contraceptives | 1.329 | 1.360 | 1.179 | 0.921 |
| | (0.308) | (0.312) | (0.270) | (0.221) |
| Current Use of Contraceptives | 0.799 | 0.848 | 0.877 | 1.302 |
| | (0.147) | (0.155) | (0.163) | (0.252) |
| Age of First Cohabitation | 1.023 | 1.026 | 1.014 | 1.015 |
| | (0.0314) | (0.0314) | (0.0314) | (0.0329) |
| Primary Education | 0.847 | 0.850 | 0.761* | 0.938 |
| | (0.127) | (0.127) | (0.113) | (0.144) |
| Religion (1=Christian) | 1.252 | 1.165 | 1.011 | 1.380* |
| | (0.233) | (0.213) | (0.188) | (0.266) |
| Ethnicity (1=Chewa) | 0.995 | 0.970 | 0.867 | 1.327* |
| | (0.138) | (0.134) | (0.122) | (0.193) |
| Child Sex (1=Male) | 0.938 | 1.062 | 0.902 | 1.021 |
| | (0.129) | (0.145) | (0.124) | (0.144) |
| Observations | 1026 | 1026 | 1023 | 1023 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. Odds rations from a logit model are reported. Standard errors are reported in parentheses. In the interest of space, Coefficients are not reported for the spline over age or fixed effects on birth month, birth order, maternal birth parity, maternal age group, or neighborhood.

this analysis, there seems to be little correlation between maternal and child characteristics.[18] Instead, across outcomes, we see that the best predictor of measurement at follow-up is having been measured at baseline.[19] Taking the findings from Table 5 and Table 6 together, we note that while attrition is high, we find little evidence that it is differential across groups.

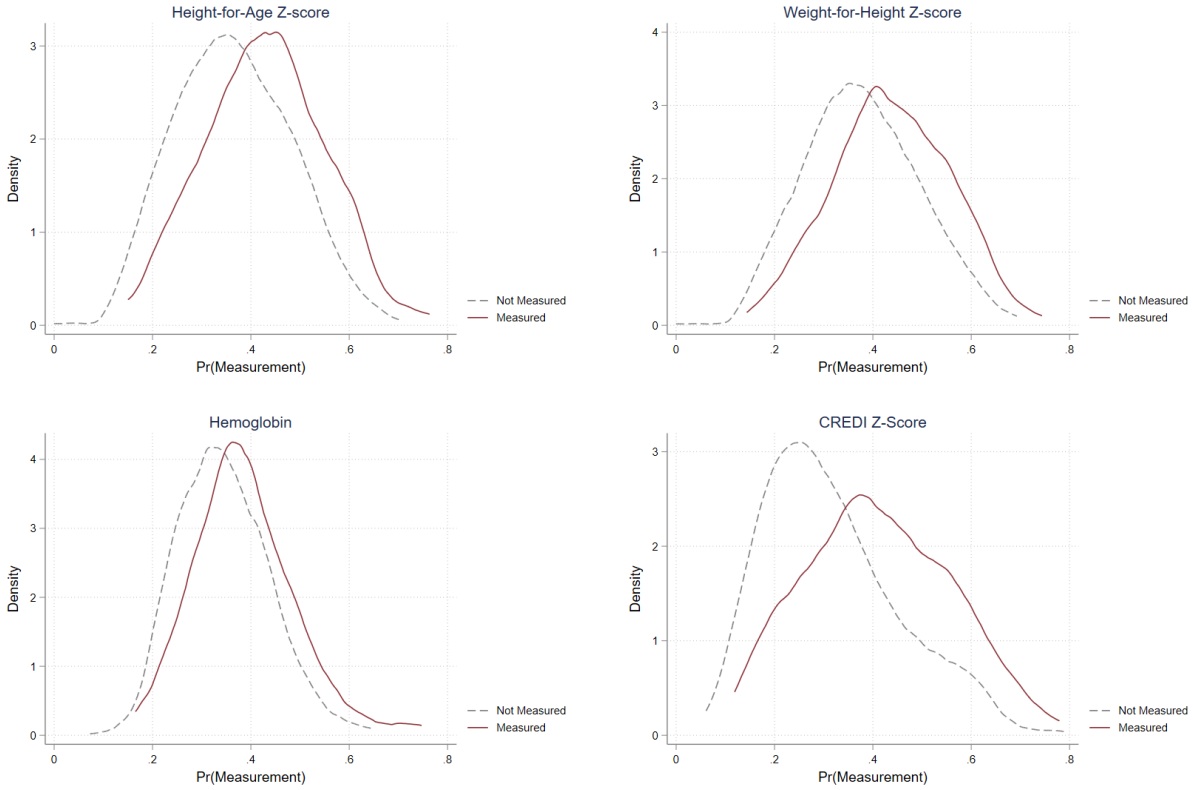
Estimating the propensity scores of measurement at follow-up also allows us to visually inspect for differences in the probability of measurement. By plotting propensity scores at baseline of children who would be measured versus those who were lost from the sample, we can get a sense of the similarity between these children across observable characteristics. Figure 5 displays the distribution of propensity scores at baseline. From this figure, we see that there is significant overlap across groups, implying that our weighting is valid and that the characteristics we observe in this study have little effect on the probability of

[18]Propensity scores are estimated using the same specification as the ITT model, which include fixed effects for birth order, month of birth, maternal age group, and neighborhood, and a spline over child's age. We do not report these coefficients, but we note that our estimates imply that younger children at baseline were less likely to be measured across all outcomes.

[19]The caregiver-reported measures of cognition were added to our survey instrument only at endline, so no children were measured at baseline.

measurement at follow-up.

Figure 5: Non-Parametric Density Distributions of Attrition Propensity Scores




Notes: Density estimations are produced using a Gaussian Kernel and the rule-of-thumb optimal bandwidth. Distributions are presented at study baseline.

# 6 Discussion

Recent extensions to the traditional economic models of fertility have accounted for the uncertainty in achieving desired family size, which many households in low- and middle-income countries face due to a lack of access to family planning and reproductive health services. These models predict that when fertility is uncertain, human capital outcomes for children are likely to be worse than in a counterfactual state of the world where fertility preferences are more certain. A decrease in the price of contraceptives or an increase in their effectiveness should, in turn, improve human capital attainment in children.

To test these hypotheses, we use data from a randomized controlled trial in Malawi that improved pregnant and postpartum women's access to family planning and reproductive health services, which improved women's knowledge of contraceptives and decreased their

effective price. Using data on children, we find that improved access to family planning positively affects children's heights, a standard marker for health and nutritional status, and children's cognitive development. These effects are relatively large for an intervention that did not explicitly target children. Our results broadly support the theoretical conclusions of frameworks that highlight the links between family planning, fertility, and children's outcomes. Further, they suggest that we may observe effects of our family planning intervention on downstream outcomes that are impacted by child health and cognitive development, such as schooling, employment, and labor market productivity, among others. Finding a causal impact on educational attainment or income would substantially further the case for family planning as an effective development strategy.

Our results contribute to an extensive literature on models of fertility and the quantity-quality trade-off in fertility preferences. We contribute to this literature by acknowledging the uncertainty in fertility that many families face and studying how changes to this uncertainty affect the decisions parents make as they relate to their children. The work that relates the most closely to ours, Cavalcanti et al. (2021), develops a model of fertility under uncertainty and uses it to discuss educational investment. In their analysis, the authors go as far as to say that family planning services may induce larger improvements in educational attainment than most educational programs. While we cannot say whether family planning may out-compete other child health programs, our results show that they may be an overlooked intervention in child health programming more generally. To this end, our study also contributes to a smaller and more limited evidence base on the link between family planning and child health by providing experimental evidence of this relationship.

Our study has several limitations. First, we observe high attrition and non-measurement among children in our sample. We use multiple methods to assess and overcome selection bias from this attrition, including inverse probability weighting, Heckman-style selection models, and bounding techniques. Throughout our study, our results seem to be comprehensively robust to these adjustments except for under severe assumptions on our bounding estimates. However, this does not allow us to definitively state that selection does not exist in our study as we cannot rule out potential bias that arises from unobservable characteristics that are associated with both attrition and non-measurement.

Additionally, our results from a causal mediation analysis highlight the potential causal mechanisms driving the results. However, in restricting our sample to children who were alive at baseline (which alleviated the attrition issue to an extent), we lose the statistical power needed to make more robust inferences. To this end, our results highlight the need for future work to more rigorously study the mediating effects of family planning as a primary outcome.

The case for expanding family planning programs and improving access to contraceptives is strong. Our results support this case by showing there are likely large and positive externalities associated with contraceptive use that have not been incorporated into the cost-benefit calculus of these services. Our results also validate the inclusion of family planning services in frameworks to reduce childhood stunting and improve early life outcomes (Black et al., 2013). Taken together, it is likely that improved access to and utilization of family planning and reproductive health services would lead to significant short- and longer-term welfare gains for women, their children, and their families over the life course.

# References


Acharya, A., M. Blackwell, and M. Sen (2016, August). Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects. *American Political Science Review 110*(3), 512–529. Publisher: Cambridge University Press. 41

Adhvaryu, A. and A. Nyshadham (2016, June). Endowments at Birth and Parents' Investments in Children. *Economic journal (London, England) 126*(593), 781–820. 1

Agesa, J. and R. U. Agesa (2020). Child Spacing in Sub-Saharan Africa: Evidence from Kenya. *Journal of African Development 21*(2), 213–226. Publisher: Penn State University Press. 1

Alderman, H. and E. M. King (1998, December). Gender differences in parental investment in education. *Structural Change and Economic Dynamics 9*(4), 453–468. 40

Almond, D. and J. Currie (2011, August). Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perspectives 25*(3), 153–172. 1

Almond, D. and B. Mazumder (2013). Fetal Origins and Parental Responses. *Annual Review of Economics 5*(1), 37–56. _eprint: https://doi.org/10.1146/annurev-economics-082912-110145 tex.ids= almond_fetal_2013. 1

Bankole, A., L. Remez, O. Owolabi, J. Philbin, and P. Williams (2020, December). From Unsafe to Safe Abortion in Sub-Saharan Africa: Slow but Steady Progress. Technical report, Guttmacher Institute, New York, NY. 1

Barcellos, S. H., L. S. Carvalho, and A. Lleras-Muney (2014, January). Child Gender and Parental Investments in India: Are Boys and Girls Treated Differently? *American Economic Journal: Applied Economics 6*(1), 157–189. 40

Barham, T. (2012, January). Enhancing Cognitive Functioning: Medium-Term Effects of a Health and Family Planning Program in Matlab. *American Economic Journal: Applied Economics 4*(1), 245–273. 3

Baron, R. M. and D. 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), 1173–1182. Place: US Publisher: American Psychological Association. 42

Barro, R. J. and G. S. Becker (1989). Fertility Choice in a Model of Economic Growth. *Econometrica 57*(2), 481–501. 1

Becker, G. (1960). An Economic Analysis of Fertility. NBER Chapters, National Bureau of Economic Research, Inc. 1, 4

Becker, G. S. and H. G. Lewis (1973, March). On the Interaction between the Quantity and Quality of Children. *Journal of Political Economy 81*(2, Part 2), S279–S288. Publisher: The University of Chicago Press. 4

Becker, G. S. and N. Tomes (1976, August). Child Endowments and the Quantity and Quality of Children. *Journal of Political Economy 84*(4, Part 2), S143–S162. tex.ids= becker_child_1976-4 publisher: The University of Chicago Press. 1, 4

Bhatia, S., W. H. Mosley, A. S. G. Faruque, and J. Chakraborty (1980). The Matlab Family Planning-Health Services Project. *Studies in Family Planning 11*(6), 202–212. 32

Black, R. E., C. G. Victora, S. P. Walker, Z. A. Bhutta, P. Christian, M. de Onis, M. Ezzati, S. Grantham-McGregor, J. Katz, R. Martorell, and R. Uauy (2013, August). Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet 382*(9890), 427–451. tex.ids= black_maternal_2013-2. 3, 27

Blundell, R., I. Preston, and I. Walker (1994, September). *The Measurement of Household Welfare*. Cambridge, UK: Cambridge University Press. Google-Books-ID: 3MAcwOSEOCcC. 1

Bongaarts, J. (2016, February). Development: Slow down population growth. *Nature News 530*(7591), 409. Section: Comment. 1, 5

Bradley, S. E. K., T. Croft, J. Fishel, and C. Westoff (2012). Revising Unmet Need for Family Planning. DHS Analytical Studies No. 25, ICF International, Calverton, MD. 1

Bruhn, M. and D. McKenzie (2009, September). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics 1*(4), 200–232. tex.ids= bruhn_pursuit_2009-4. 7

Bärnighausen, T., J. Bor, S. Wandira-Kazibwe, and D. Canning (2011, January). Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. *Epidemiology (Cambridge, Mass.) 22*(1), 27–35. 16

Casterline, J. B. and S. Agyei-Mensah (2017). Fertility Desires and the Course of Fertility Decline in sub-Saharan Africa. *Population and Development Review 43*, 84–111. 1

Casterline, J. B. and S. W. Sinding (2004, January). Unmet Need for Family Planning in Developing Countries and Implications for Population Policy. *Population and Development Review 26*(4), 691–723. 1

Cavalcanti, T., G. Kocharkov, and C. Santos (2021, February). Family Planning and Development: Aggregate Effects of Contraceptive Use. *The Economic Journal 131*(634), 624–657. 1, 5, 26, 34, 41, 45

Cleland, J., A. Conde-Agudelo, H. Peterson, J. Ross, and A. Tsui (2012, July). Contraception and health. *The Lancet 380*(9837), 149–156. 2, 32

Conde-Agudelo, A., A. Rosas-Bermudez, F. Castano, and M. H. Norton (2012). Effects of Birth Spacing on Maternal, Perinatal, Infant, and Child Health: A Systematic Review of Causal Mechanisms. *Studies in Family Planning 43*(2), 93–114. 32

Conde-Agudelo, A., A. Rosas-Bermúdez, and A. C. Kafury-Goeta (2006, April). Birth Spacing and Risk of Adverse Perinatal Outcomes: A Meta-analysis. *JAMA 295*(15), 1809–1823. 2, 32

Cummins, J. (2013, August). On the Use and Misuse of Child Height-for-Age Z-score in the Demographic and Health Surveys. Technical report. Publication Title: Working Papers.

12

DaVanzo, J., A. Razzaque, M. Rahman, L. Hale, K. Ahmed, M. A. Khan, G. Mustafa, and K. Gausia (2004). The effects of birth spacing on infant and child mortality, pregnancy outcomes, and maternal morbidity and mortality in Matlab, Bangladesh. *Technical Consultation and Review of the Scientific Evidence for Birth Spacing*. 3, 32

de Onis, M., C. Garza, C. G. Victora, A. W. Onyango, E. A. Frongillo, and J. Martines (2004, January). The who Multicentre Growth Reference Study: Planning, Study Design, and Methodology. *Food and Nutrition Bulletin 25*(1_suppl_1), S15–S26. Publisher: SAGE Publications Inc. 11

Debpuur, C., J. F. Phillips, E. F. Jackson, A. Nazzar, P. Ngom, and F. N. Binka (2002, June). The Impact of the Navrongo Project on Contraceptive Knowledge and Use, Reproductive Preferences, and Fertility. *Studies in Family Planning 33*(2), 141–164. 32

Dewey, K. G. and R. J. Cohen (2007). Does birth spacing affect maternal or child nutritional status? A systematic literature review. *Maternal and Child Nutrition 3*(3), 151–173. 2

Fink, G., M. Pierre-Louis, and D. C. McCoy (2018, January). CREDI User Guide. pp. 16. 11

Fink, G. and P. C. Rockers (2014, July). Childhood growth, schooling, and cognitive development: further evidence from the Young Lives study. *The American Journal of Clinical Nutrition 100*(1), 182–188. 18

Fink, G., C. R. Sudfeld, G. Danaei, M. Ezzati, and W. W. Fawzi (2014, July). Scaling-Up Access to Family Planning May Improve Linear Growth and Child Development in Low and Middle Income Countries. *PLOS ONE 9*(7), e102391. 2, 32

Gertler, P. J. and J. W. Molyneaux (1994, February). How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility. *Demography 31*(1), 33–63. 3

Gipson, J. D., M. A. Koenig, and M. J. Hindin (2008). The Effects of Unintended Pregnancy on Infant, Child, and Parental Health: A Review of the Literature. *Studies in Family Planning 39*(1), 18–38. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4465.2008.00148.x. 1

Grantham-McGregor, S. and C. Ani (2001, February). A Review of Studies on the Effect of Iron Deficiency on Cognitive Development in Children. *The Journal of Nutrition 131*(2), 649S–668S. Publisher: Oxford Academic. 11

Gruber, J., P. Levine, and D. Staiger (1997, May). Abortion Legalization and Child Living Circumstances: Who is the "Marginal Child?". Working Paper 6034, National Bureau of Economic Research. Series: Working Paper Series. 33

Haider, T. L. and M. Sharma (2013, October). Barriers to Family Planning and Contraception Uptake in Sub-Saharan Africa: A Systematic Review. *International Quarterly of Community Health Education 33*(4), 403–413. Publisher: SAGE Publications Inc. 1

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica 47*(1), 153–161. Publisher: [Wiley, Econometric Society]. 3, 16

Hoddinott, J., J. R. Behrman, J. A. Maluccio, P. Melgar, A. R. Quisumbing, M. Ramirez-Zea, A. D. Stein, K. M. Yount, and R. Martorell (2013, November). Adult consequences of growth failure in early childhood. *The American Journal of Clinical Nutrition 98*(5), 1170–1178. 00150. 11

Hogan, D. R., J. A. Salomon, D. Canning, J. K. Hammitt, A. M. Zaslavsky, and

T. Bärnighausen (2012, December). National HIV prevalence estimates for sub-Saharan Africa: controlling selection bias with Heckman-type selection models. *Sexually Transmitted Infections 88 Suppl 2*(Suppl_2), i17–23. 16

Huttly, S., C. Victora, F. Barros, and J. P. Vaughan (1992). Birth Spacing and Child Health in Urban Brazil. *Pediatrics 89*(6), 1049–1053. 32

Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods 15*(4), 309–334. Place: US Publisher: American Psychological Association. 43, 44

Imai, K., L. Keele, and T. Yamamoto (2010, February). Identification, Inference and Sensitivity Analysis for Causal Mediation Effects. *Statistical Science 25*(1), 51–71. Publisher: Institute of Mathematical Statistics. 41, 42, 43, 44

Joshi, S. and T. P. Schultz (2007, February). Family Planning as an Investment in Development: Evaluation of a Program's Consequences in Matlab, Bangladesh. SSRN Scholarly Paper ID 962938, Social Science Research Network, Rochester, NY. 4

Joshi, S. and T. P. Schultz (2013, February). Family Planning and Women's and Children's Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh. *Demography 50*(1), 149–180. 32

Karra, M. and D. Canning (2020). The Effect of Improved Access to Family Planning on Postpartum Women: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols 9*(8), e16697. Company: JMIR Research Protocols Distributor: JMIR Research Protocols Institution: JMIR Research Protocols Label: JMIR Research Protocols Publisher: JMIR Publications Inc., Toronto, Canada. 6, 9

Karra, M., D. Maggio, M. Guo, B. Ngwire, and D. Canning (2022). The Causal Effect of Improved Access to Postpartum Family Planning: Evidence from a Single-Blind, Randomized Controlled Trial in Urban Malawi. *Proceedings of the National Academy of Sciences*. 7, 10, 32

King, J. C. (2003, May). The risk of maternal nutritional depletion and poor outcomes increases in early or closely spaced pregnancies. *The Journal of Nutrition 133*(5 Suppl 2), 1732S–1736S. 2

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2007.00733.x. 3, 16

Koné, S., B. Bonfoh, D. Dao, I. Koné, and G. Fink (2019, December). Heckman-type selection models to obtain unbiased estimates with missing measures outcome: theoretical considerations and an application to missing birth weight data. *BMC Medical Research Methodology 19*(1), 231. 16

Kozuki, N., A. C. Lee, M. F. Silveira, C. G. Victora, L. Adair, J. Humphrey, R. Ntozini, R. E. Black, J. Katz, and Child Health Epidemiology Reference Group Small-for-Gestational-Age-Preterm Birth Working Group (2013, September). The associations of birth intervals with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. *BMC Public Health 13*(3), S3. 2

Li, H., J. Zhang, and Y. Zhu (2008, February). The quantity-Quality trade-Off of children In a developing country: Identification using chinese twins. *Demography 45*(1), 223–243. 3

Macours, K. and T. Molina Millán (2017, April). Attrition in Randomized Control Trials:

Using Tracking Information to Correct Bias. 16

Malawi, N. S. O. and ICF (2017). Malawi Demographic and Health Survey 2015-16. Technical report, NSO and ICF, Zomba, Malawi:. 43

Mayo, J., B. Shachar, D. Stevenson, and G. Shaw (2017, August). Interpregnancy Interval and Adverse Pregnancy Outcomes. *Obstetrics & Gynecology 130*(2). 2

Miller, G. (2010). Contraception as Development? New Evidence from Family Planning in Colombia*. *The Economic Journal 120*(545), 709–736. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0297.2009.02306.x. 4

Miller, G. and K. Singer Babiarz (2016). Family Planning Program Effects: Evidence from Microdata. *Population and Development Review 42*, 7–26. Series: Working Paper Series. 3

Miller, J. E. (1991). Birth Intervals and Perinatal Health: An Investigation of Three Hypotheses. *Family Planning Perspectives 23*(2), 62–70. 2

Miller, R. and M. Karra (2020). Birth Spacing and Child Health Trajectories. *Population and Development Review 46*(2), 347–371. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/padr.12335. 2, 32

Molitoris, J. (2017). The Effect of Birth Spacing on Child Mortality in Sweden, 1878–1926. *Population and Development Review 43*(1), 61–82. 2

Molitoris, J., K. Barclay, and M. Kolk (2019, August). When and Where Birth Spacing Matters for Child Survival: An International Comparison Using the DHS. *Demography 56*(4), 1349–1370. 32

Newson, R. B. (2010, December). Frequentist Q-values for Multiple-test Procedures. *The Stata Journal 10*(4), 568–584. Publisher: SAGE Publications. 13

Pearl, J. (2014, December). Interpretation and identification of causal mediation. *Psychological Methods 19*(4), 459–481. 41

Portner, C. C., K. Beegle, and L. Christiaensen (2011, September). Family Planning and Fertility : Estimating Program Effects Using Cross-Sectional Data. Policy Research Working Paper WPS 5812, World Bank, Washington, DC. Accepted: 2012-03-19T18:04:51Z. 4

Reuters (2017, October). Vampire scare prompts U.N. pullout from southern Malawi. *Reuters*. 13

Robinson, W. C. (1987, July). The Time Cost of Children and Other Household Production. *Population Studies 41*(2), 313–323. Publisher: Routledge _eprint: https://doi.org/10.1080/0032472031000142816. 1

Roche, A. F. and J. H. Himes (1980, September). Incremental growth charts. *The American Journal of Clinical Nutrition 33*(9), 2041–2052. 18

Rosenzweig, M. R. and K. I. Wolpin (1986). Evaluating the Effects of Optimally Distributed Public Programs: Child Health and Family Planning Interventions. *The American Economic Review 76*(3), 470–482. Publisher: American Economic Association. 4

Rutstein, S. O. (2006, January). Effects of preceding birth intervals on neonatal, infant and under-five years mortality and nutritional status in developing countries: Evidence from the Demographic and Health Surveys. *International Journal of Gynecology & Obstetrics 89*(S1), S7–S24. 2, 32

Schultz, T. P. (1969, March). An Economic Model of Family Planning and Fertility. *Journal of Political Economy 77*(2), 153–180. 1

Schultz, T. P. (2007, January). Chapter 52 Population Policies, Fertility, Women's Human Capital, and Child Quality. In T. P. Schultz and J. A. Strauss (Eds.), *Handbook of Development Economics*, Volume 4, pp. 3249–3303. Elsevier. 1

Simes, R. J. (1986). An Improved Bonferroni Procedure for Multiple Tests of Significance. *Biometrika 73*(3), 751–754. Publisher: [Oxford University Press, Biometrika Trust]. 13

Singh, A., A. Singh, and B. Mahapatra (2013, April). The Consequences of Unintended Pregnancy for Maternal and Child Health in Rural India: Evidence from Prospective Data. *Maternal and Child Health Journal 17*(3), 493–500. 1

Smits, L. J. and G. G. Essed (2001, December). Short interpregnancy intervals and unfavourable pregnancy outcome: role of folate depletion. *Lancet (London, England) 358*(9298), 2074–2077. 32

Storey, J. D. and R. Tibshirani (2003, August). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences 100*(16), 9440–9445. Publisher: National Academy of Sciences Section: Physical Sciences. 13

United Nations (2018). World Contraceptive Use 2018. 1

Victora, C. G., M. de Onis, P. C. Hallal, M. Blössner, and R. Shrimpton (2010, March). Worldwide timing of growth faltering: revisiting implications for interventions. *Pediatrics 125*(3), e473–480. 11

Westoff, C. F. and L. H. Ochoa (1991). *Unmet need and the demand for family planning*, Volume 5. Columbia, MD: Institute for Resource Development/Macro International. 1

World Health Organization (2005). Report of a WHO Technical Consultation of Birth Spacing. Technical report, World Health Organization, Geneva, Switzerland. 2, 32

Özler, B., Çelik, S. Cunningham, P. F. Cuevas, and L. Parisotto (2021, November). Children on the move: Progressive redistribution of humanitarian cash transfers among refugees. *Journal of Development Economics 153*, 102733. 17

# A Appendix A: Alternative Causal Mechanisms

While our theory of change focuses on the mediating role of human capital investments, we also acknowledge two alternative mechanisms that, while plausible at the macro level, are unlikely to drive our study results. First, since healthy interpartum intervals are a central objective of family planning programs, it is possible that our intervention improved child health outcomes through a reduction in maternal depletion syndrome, the insufficient repletion of maternal folate levels between pregnancies, which poses risks to child and maternal health (Smits and Essed, 2001). While evidence is limited, studies have documented the role of family planning in allowing women and couples to more effectively time and space pregnancies (Bhatia et al., 1980; Cleland et al., 2012; Debpuur et al., 2002; Joshi and Schultz, 2013). This is promising, as short birth intervals have been associated with an increased risk of neonatal, infant, and child mortality (Rutstein, 2006; Conde-Agudelo et al., 2006, 2012; DaVanzo et al., 2004; Molitoris et al., 2019) and have also been linked to higher rates of stunting and wasting in children (Rutstein, 2006; World Health Organization, 2005; Huttly et al., 1992; Miller and Karra, 2020; Fink et al., 2014). Indeed, in Karra et al. (2022), we show that pregnant and postpartum women who received our family planning intervention

had a significantly lower risk of having a subsequent pregnancy within 24 months and giving birth again within three years of their last birth. These results imply that changes in inter-partum intervals due to our intervention may be driving our results. However, in this study, we restrict our sample to children born at baseline; as such, these children's preceding birth intervals were determined before intervention initiation and cannot have been influenced by our intervention.

Table A1: Baseline Characteristics Disaggregated by Fertility at Year 2

| | Woman had Second Birth? | | | |
|---|---|---|---|---|
| | No N=1487 | Yes N=119 | Total N=1606 | Difference (1)-(2) |
| ***Baseline Characteristic*** | (1) | (2) | (3) | |
| Current Use of FP (1 = Yes) | 0.252 | 0.303 | 0.255 | -0.051 |
| Long Acting Method Use (1=Yes) | 0.038 | 0.025 | 0.037 | 0.013 |
| Injectable Use (1=Yes) | 0.195 | 0.252 | 0.199 | -0.057 |
| Implant Use (1=Yes) | 0.036 | 0.008 | 0.034 | 0.028 |
| Ever Use of FP (1 = Yes) | 0.782 | 0.723 | 0.778 | 0.059 |
| Woman's Age (Years) | 24.876 | 24.706 | 24.863 | 0.170 |
| Total Number of Children | 2.439 | 2.370 | 2.434 | 0.069 |
| Primary Education (1 = Yes) | 0.552 | 0.563 | 0.553 | -0.011 |
| Secondary Education (1 = Yes) | 0.463 | 0.479 | 0.464 | -0.016 |
| Tertiary Education (1 = Yes) | 0.027 | 0.034 | 0.027 | -0.007 |
| Religion (1 = Christian) | 0.838 | 0.815 | 0.836 | 0.023 |
| Ethnicity (1 = Chewa) | 0.424 | 0.429 | 0.425 | -0.004 |
| Woman Works (1 = Yes) | 0.102 | 0.076 | 0.100 | 0.026 |
| Age of First Cohabitation (Years) | 18.945 | 18.933 | 18.944 | 0.013 |
| **P-Value (Joint Significance)** | | | | **0.153** |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a woman. Means are presented for the value of observable characteristics at baseline. The column a woman is assigned to is based on if she did or did not have a second birth before the second-year follow-up survey.

Second, for any intervention that induces a decline in fertility, there is a set of children who are not observed but would have been born in the absence of the intervention. It is often the case that these "marginal children" would have been born to mothers from more disadvantaged settings.[20] It is, therefore, possible that changes in child health outcomes due to fertility programs may reflect the fact that fewer children were born in disadvantaged settings rather than reflecting actual benefits conferred to children. This may also be the case with the results presented in this study; however, as we restrict our analysis to children already alive at baseline, these "marginal" children would not directly enter our estimation

[20]In the United States, studies have shown that children who are not born as a result of legalized abortion would have been 70 percent more likely to be born to a single mother and 40 percent more likely than the average child to grow up in poverty (Gruber et al., 1997).

and, as such, cannot drive our results. Nevertheless, children in our sample may be less likely to have a younger sibling due to our intervention, implying that the compositional effects we discuss above may enter our estimation indirectly. To explore this possibility, we compare baseline characteristics of mothers who did and did not have a second birth before our endline survey in Appendix Table A1. We see from this table that there are no significant differences in observable characteristics across these birth groups. These results lead us to believe that the compositional effects are unlikely to be driving any intervention impacts that we find and that other physiological changes likely cause any effects that are driven by changes in birth spacing.

# B Appendix B: Contraceptive Choice

One key feature of the model developed by Cavalcanti et al. (2021) is that the derived demand for contraceptives is predicated on women's fertility preferences (both in terms of spacing and timing births) and the uncertainty surrounding pregnancy. However, there are many reasons why women and couples choose to use contraceptives that are not driven by fertility intentions or pregnancy prevention. Common reasons include preventing HIV and other sexually transmitted infections, facilitating control over menstruation, and improving sexual satisfaction and well-being. With these in mind, the predictions of the Cavalcanti et al. (2021) model may not be as relevant for our setting if the predominant drivers for contraceptive use are not fertility-related.

In Table B1, we present attributes and features that women reported to be most important to them when considering their choice of contraceptive method during our pre-intervention baseline survey.[21] This table highlights effectiveness at preventing pregnancy as the most important attribute women consider when choosing a contraceptive method. In contrast, fewer women choose their method by considering more effective methods for preventing HIV and STIs. Rather, alongside pregnancy prevention, women in our sample seek contraceptives that they perceive as not threatening to their health, that do not have side effects, that are easy to use, and that do not require frequent visits to a clinic. Almost half of women report prioritizing pregnancy prevention in their contraceptive choice, which implies that fertility preferences play a significant role in shaping their derived demand for contraception.

This study is motivated by a causal framework in which contraceptives reduce uncertainty in pregnancy prevention and subsequent fertility outcomes, which may allow women to invest more effectively in their and their children's human capital. While this pathway has been

[21] We note that this table is not conditional on the woman using contraceptives. Instead, we asked women about the features that were important to them when making contraceptive choices, regardless of which choices were made.

Table B1: Most Important Features in Contraceptive Choice at Baseline

| Contraceptive Feature | Proportion of Women |
|---|---|
| Prevents Pregnancy | 0.48 |
| Concealable | 0.06 |
| Prevents STI/HIV | 0.02 |
| No Health Risks | 0.29 |
| Maintains Regular Menstruation | 0.18 |
| No Side Effects | 0.20 |
| Easy to Use | 0.37 |
| Easy to Obtain | 0.19 |
| Long Acting | 0.27 |

Notes: Statistics represent the percentage of postpartum women at baseline who name each feature in response to the question, "In choosing a contraceptive method, what feature(s) would be most important to you?". The options were not mutually exclusive, and women could name multiple attributes. Thus, the sum of the statistics is greater than 100 percent.

identified in recent work, we note it is not the only channel through which family planning may contribute to children's human capital. While many women use family planning to delay births, many also use contraceptives to avert pregnancy and limit future births. If women use family planning to limit and space births, then a reduction in parity may also contribute to increased human capital investment in children.

We identify the motivations behind women's contraceptive use among the sub-sample of women who were immediately postpartum when they were recruited at baseline; we exclude women who were pregnant at baseline because they would not have a stated preference for family planning at the time of the interview. Columns (1), (2), and (3) of Table B2 present baseline statistics of postpartum women's fertility preferences and stated intentions around future childbearing. We present these statistics for all postpartum women, those who were using contraceptives, and those who were not using contraceptives. In addition, we also present significance tests of the difference between both means and standard deviations.

We find that 57 percent of postpartum women in our sample stated an intention to have a subsequent pregnancy. Interestingly, women who use contraceptives are more likely to report wanting to have a subsequent birth, suggesting a preference for spacing. Moreover, we find the gap between ideal and desired family size to be non-differential by assignment to either the intervention or control arm; on average, women in our sample have had 0.7 fewer births than their desired parity. The evidence implies that the principal reason for women using contraceptives within our sample is not to limit births but rather to space births. When comparing women's perceptions around their probability of getting pregnant within the next year, we see that women who use contraception report a significantly lower perceived

Table B2: Preferences and Perceptions among Postpartum Women at Baseline

| | Using Contraceptives? | | | | |
|---|---|---|---|---|---|
| | No | Yes | Total | Difference | P-Value |
| | N=528 | N=507 | N=1035 | (1)-(2) | $\frac{sd(1)}{sd(2)} > 1$ |
| ***Variable*** | (1) | (2) | (3) | | |
| Want's Another Child (1=Yes) | 0.552 | 0.599 | 0.575 | -0.047 | 0.372 |
| Desired Wait Time Until Next Child (Months) | 63.944 | 68.429 | 66.251 | -4.485** | 0.099* |
| Desired Number of Children | 3.261 | 3.150 | 3.206 | 0.111* | 0.483 |
| Number of Births | 2.511 | 2.318 | 2.416 | 0.194** | 0.291 |
| Perceived Probability of Pregnancy (1 = Not likely, 5=Very Likely) | 4.058 | 1.616 | 2.852 | 2.442*** | 0.00*** |

Notes: The unit of analysis is a woman. We exclude pregnant women from this analysis as they lack the need for contraceptives. Perceived pregnancy is a composite variable created using the questions "If you were to not use any family planning method, how likely do you think it is that you will become pregnant during the next year?" and "If you were to continue to use your family planning method, how likely do you think it is that you would become pregnant during the next year?. The variable is then created using the answer to the question that matches the woman's contraceptive status.

probability of getting pregnant than those who do not. Moreover, there is significantly less variation in the perceived risk of pregnancy among women who are using contraceptives than those who are not using contraceptives. These results imply that women within our sample predominantly use contraceptives to more effectively space and time future pregnancies and reduce the uncertainty over their future fertility.

# C Appendix C: Expanded Sample Robustness Check

Our analysis uses data from all index children alive at baseline. However, considering our study enrolled 2,143 women into our study, and half of the women enrolled were pregnant at the time of baseline, there should be 1,109 index children not observed at baseline but born during the first year of the intervention.[22] Of these potential children, we observe 739 of them and record anthropometric data from 314 of them. We chose not to use these children in our main sample because of the 370 potential children we never observed; it is unclear how many were truly lost to follow-up and how many were the result of misreported pregnancy status. As we did not conduct a urine test to verify pregnancy status, it is unclear how many women misreported their pregnancy status. The potential of misreporting creates uncertainty about how we should treat these children in our attrition analysis. Further, in Section 4.4.1, we discuss our use of inverse probability weighting to correct for attrition bias. Within this methodology, if we were to consider the full set of index children as our primary sample, we would be forced to restrict the creation of our propensity scores to only maternal characteristics, as we never observed 370 potential children.

[22]The number of potential children could be higher, allowing for non-singleton births.

Table C1: ITT Effects on Main Outcomes for All Index Children

| | Height-for-Age Z-Scores | | CREDI Z-Scores | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment | 0.231* | 0.257** | 0.101** | 0.103** |
| | (0.121) | (0.122) | (0.0488) | (0.0501) |
| Control Mean | -1.5 | -1.5 | 0.1047 | 0.1047 |
| Survey Wave | First Year Follow-Up | | Second Year Follow-Up | |
| Age and Birth Month Controls | X | X | X | X |
| Mother and Child Characteristics | | X | | X |
| Observations | 720 | 717 | 727 | 726 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. Both adjusted and unadjusted regressions include birth month fixed effects and a flexible form over the child's age. The adjusted regressions include woman-level controls such as use of family planning methods, ever use of family planning methods, the total number of children who are alive (included as a set of dummy variables), educational attainment of the woman (primary or less versus secondary and higher), age of the woman (in three age groups), age of sexual debut, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut and child-level controls such as biological sex, and birth order (included as a set of dummy variables). All controls are included at their baseline value. Neighborhood fixed effects are included, and heteroskedasticity-robust standard errors are reported in parentheses. Sharpened

In restricting our sample of children, we recognize that the interpretation of our treatment effects may be different relative to an analysis where the entire sample of index children is used. To account for potential differences due to the sample selection, we present our main ITT estimates for all index children on height-for-age z-scores and CREDI scores in Table C1. We see that, for both outcomes, our estimates are attenuated but are not statistically different than the estimate we present using our main sample. Given the congruence of our estimates across specifications, we find that the choice to exclude children of women who were pregnant at baseline from our sample does not have a meaningful effect on the interpretation of our results.

# D Appendix D: Intervention Components Analysis

The multi-component structure of our intervention naturally begs the question to identify the intervention component(s) that are the most effective in driving outcomes. However, several causal identification challenges prevent us from effectively responding to this question. First, there are likely to be several latent and unobservable factors that determine women's use

of each intervention component. While we can navigate this challenge using exogenous variation from treatment assignment, we would have, in an ideal world, multiple sources of said variation. As women assigned to the treatment arm received the full combined intervention, it is impossible to compare, for example, women who *only* received counseling to women who *only* used transport. In this appendix, we will present instrumental variable estimates for the counseling and transport component that would identify the LATE of treatment assignment in a single-component intervention. However, as we are analyzing individual components of a multi-component intervention, this estimation is likely to be under-identified, and we, therefore, treat this evidence as suggestive.

## D.1 *Specification*

Having acknowledged that such a specification should not be considered causal, we estimate the relationship between each intervention component and our outcomes of interest using the following specification:

$$Y_{im} = \alpha + \beta_C C_{im} + \delta MOB_{im} + \gamma f(age_{im}) + \mathbf{X}_i m\lambda + \mathbf{Z}_m\zeta + \chi_m + \varepsilon_{im} \tag{D1}$$

Where the notation once again follows that presented in Equation (1). Here $C_{im}$ represents the component of interest and is instrumented with treatment status in a two-stage least squares (2SLS) estimation. As we only have one treatment assignment, including all three treatment components in the same estimation would be under-identified, so we opt to test each component through a separate specification. We present results for the counseling and transport component for both the extensive and intensive margins and again present unadjusted and adjusted specifications.

## D.2 *Results*

We present the relationship between the use of the intervention components and our outcomes of interest in Appendix Table D1. We see that on the intensive margins, each counseling visit is associated with a 0.057-0.064 SD increase in height-for-age and a 0.039-0.047 SD increase in CREDI scores, while each transport visit to the Kauma Clinic is associated with a 0.49-0.59 SD increase in height-for-age and a 0.36-0.42 SD increase in CREDI scores. On the extensive margin, counseling is associated with a 0.30-0.34 SD increase in height-for-age and a 0.20-0.24 SD increase in CREDI scores, while transport use is associated with a 1.00-1.14 SD increase in height-for-age and a 0.89-1.01 SD increase in CREDI scores.

The most striking aspect of the results presented in Appendix Table D1 is the large

Table D1: Intervention Components Analysis

| | Counseling (Intensive Margin) | | Counseling (Extensive Margin) | | Transport (Intensive Margin) | | Transport (Extensive Margin) | |
|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Panel A: Height-for-Age Z-Score | | | | | | | | |
| Treatment Component | 0.0565* | 0.0643** | 0.297* | 0.339** | 0.492* | 0.594** | 1.002* | 1.135** |
| | (0.0309) | (0.0307) | (0.162) | (0.162) | (0.281) | (0.301) | (0.565) | (0.571) |
| Observations | 406 | 404 | 406 | 404 | 406 | 404 | 406 | 404 |
| First Stage F-Statistic | 1976.94 | 2009.36 | 4157.42 | 4726.72 | 40.20 | 44.21 | 71.07 | 67.38 |
| Panel B: CREDI Z-Score | | | | | | | | |
| Treatment Component | 0.0393*** | 0.0469*** | 0.201*** | 0.241*** | 0.363** | 0.418*** | 0.891** | 1.012*** |
| | (0.0148) | (0.0145) | (0.0753) | (0.0742) | (0.154) | (0.153) | (0.360) | (0.343) |
| Observations | 326 | 326 | 326 | 326 | 326 | 326 | 326 | 326 |
| First Stage F-Statistic | 1189.51 | 1226.38 | 3132.59 | 3455.85 | 25.11 | 24.93 | 43.74 | 44.71 |
| Age and Birth Month Controls | X | X | X | X | X | X | X | X |
| Mother and Child Characteristics | | X | | X | | X | | X |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from 2SLS models with the treatment component instrumented with treatment assignment and standard errors in parenthesis. Both adjusted and unadjusted regressions include month of birth fixed effects and a flexible form over the child's age . The adjusted regressions include woman-level controls such as the total number of children who are alive (in seven groups), educational attainment of the woman (primary or less versus secondary and higher), age of the woman (in three age groups), religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut and child-level controls such as biological sex, and birth order (included as dummy variables to relax the linear restriction). Neighbothood fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

difference between components, implying that the transport component was much more effective than counseling. However, the use of the counseling component was much more widespread. Of the 782 treatment women re-interviewed at second-year follow-up, only 39 women had not had a counseling visit past their initial enrollment visit. In contrast, of those 782 women, 599 did not use the transport provided. With these levels of treatment exposure in mind, these results show us that the association with FP counseling is marginal, but because of its widespread use, it is likely to be the component that is driving our main results.

Meanwhile, transport had large gains for women who used it, but fewer women used it. Further, transport use is likely higher among the poorest women, who would benefit the most from increased access. As such, the children born to mothers who used the transport likely would have been worse off than the average child but for the treatment. With this in mind, we suspect the possible gains to treatment were much larger among children born to these women, partially explaining the large association between transport use, height-for-age, and CREDI scores.

# E Appendix E: Heterogeneous Treatment Effects

While the ITT results we find are, in and of themselves, interesting and illuminating, we are also interested in the effects of the intervention on different subgroups. For instance, we are interested to see if our intervention has greater benefits for boys versus girls. While a large literature suggests that there is a gender bias towards boys in parental investments (Alderman and King, 1998; Barcellos et al., 2014), it is possible that induced investments from an inherently gender-sensitive intervention may benefit girls more than boys.

Similarly, Table B2 suggests that women mostly use contraceptives in our context to properly space births rather than prevent them. As such, we may expect our intervention to have little impact on households that do not intend to have births in the future.

We explore the heterogeneous effects of our intervention in Table E1. In each column, we restrict our sample to only those women and children who meet a given condition. For instance, in column (1), we report the estimate for Equation 1, where the sample is restricted to only female children. The estimates for all columns in Table E1 represent a fully adjusted specification.

Table E1: Heterogeneous Treatment Effect of the Intervention

| | Gender | | Wants Another Child? (Baseline) | | Birth Order | |
|---|---|---|---|---|---|---|
| | Female | Male | No | Yes | First Born | Subsequent Birth |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A: Height-for-Age Z-Score | | | | | | |
| Treatment | 0.221 | 0.425* | 0.0892 | 0.511** | 0.0742 | 0.424** |
| | (0.228) | (0.248) | (0.255) | (0.220) | (0.312) | (0.207) |
| Observations | 206 | 198 | 187 | 217 | 118 | 286 |
| Panel B: CREDI Z-Scores | | | | | | |
| Treatment | 0.342*** | 0.175* | 0.349*** | 0.167* | 0.153 | 0.251*** |
| | (0.118) | (0.0959) | (0.117) | (0.0897) | (0.132) | (0.0836) |
| Observations | 182 | 180 | 152 | 210 | 112 | 250 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. Unadjusted regressions include birth month fixed effects and a flexible form over the child's age. Heteroskedasticity-robust standard errors, clustered at the mother level, are reported in parentheses. Sample splits for maternal fertility preference are conducted using the baseline value.

Table E1 shows us that, as discussed in the main text, the point estimates of our specifications suggest that effects are larger on HAZ for boys and cognition for girls. However, we cannot reject the hypothesis that the coefficients are equal. We will not repeat the discus-

sion of these effects provided in Section 5, but we add that HAZ and CREDI Z-Scores were measured at different times. Therefore, the differential investments made in girls may take time to emerge.

In addition to the results across gender and pregnancy, we find strong evidence that the treatment effect on HAZ is larger for children born in households that want more children, and the treatment effect on both HAZ and CREDI Z-scores for children who have older siblings. The results for children born to families with intentions to have future births may be explained by the descriptive analysis presented in Table B2. If women mainly use contraceptives to space births, our FP/RH intervention may have had little effect on women who were not looking to space births. The results for children with older siblings are more obscure, given that we have shown that preceding birth intervals are balanced across treatment arms. One possibility is that firstborn children are born to younger parents who are earlier in their income cycle and do not have the means to act on increased demand for human capital investments. Another is that new parents generally invest heavily in their firstborn child, implying that there is little room to improve human capital investments.

# F Appendix F: Mediation Analysis

We discuss in Section 2, our analysis throughout this paper is motivated by the model developed by Cavalcanti et al. (2021). While our results support their conclusions, it would also be ideal to test the causal pathways they describe. As an extension to our main analysis, we conduct a causal mediation analysis to identify potential underlying causal mechanisms and infer the extent to which any observed impacts of our intervention on child health outcomes support the human capital framework in Section 2. Our approach to causal mediation analysis is motivated by Structural Equation Modeling (SEM) approaches that seek to identify the causal mediation effect, also referred to as the natural indirect effect (Pearl, 2014; Imai et al., 2010; Acharya et al., 2016). The causal mediation effect is given by:

$$\delta(t) = Y_i\left(t, M_i(1)\right) - Y_i\left(t, M_i(0)\right) \tag{F1}$$

where $M(t)$ is a variable affected by the treatment and lies on the causal path between the treatment and the outcome. When defining causal mediation through the principle of temporality, the mediating variable must occur *after* treatment exposure but *before* the outcome is measured, and the causal mediation effect represents the change in the potential outcome that is induced by changes in the state of the mediating variable under treatment regime $t$. Figure 6 presents the causal mediation relationship in a Directed Acyclic Graph

(DAG).

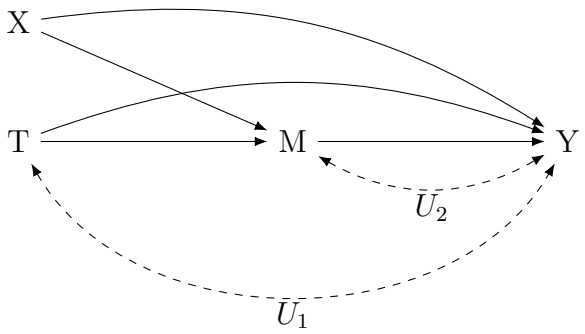



Figure 6: DAG displaying mediator relationship

T: Treatment, M: Mediator, Y: Outcome, X: Set of pre-treatment covariates, $U_1$: Possible confounders between T and Y, $U_2$: Possible confounders between M and Y.

The fundamental problem of causal inference implies that we are unable to observe both $Y_i(t, M(1))$ and $Y_i(t, M(0))$, implying that the causal meditation effect is empirically unidentified. However, Imai et al. (2010) shows that the *average causal mediation effect* (ACME), $\bar{\delta}(t) = E[Y_i(t, M_i(1)) - Y_i(t, M_i(0))]$ can be parametrically identified within a Baron and Kenny (1986) Linear Structural Equation Model (LSEM) structure and under the following sequential ignorability assumption:

*ASSUMPTION 1:* (Sequential Ignorability)

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i | \mathbf{X}_i = \mathbf{x} \tag{F2}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, \mathbf{X}_i = \mathbf{x} \tag{F3}$$

for all $t, t' \in \{0, 1\}$. In our context, condition F2 implies that there exists no characteristic, other than those contained in the covariate set $\mathbf{X}_i$, that influences both treatment assignment and either the mediator (birth spacing or health investments) or the outcome (child health and cognition). Since we are presenting results for a randomized experiment, there is a strong case to be made that condition F2 holds, even if the set of covariates was limited. In contrast, for condition F3 to hold, we must assume the outcome and the mediator are not jointly determined, conditional on treatment status and a set of covariates, $\mathbf{X}_i$. This means we must assume that outside of the covariate set $\mathbf{X}_i$ and treatment status, no observable or non-observable characteristic influences both the mediator and outcome.

In this study, we present analysis for two mediating variables, birth spacing and healthcare use, which we may be concerned are not orthogonal to child health outcomes, thus violating condition F3. For example, if higher household wealth is associated with better health

outcomes and greater healthcare use, we would expect the mediating effect of healthcare use in our results to induce a positive bias. We cope with this possible violation in two ways. First, we present a sensitivity analysis that will be discussed later in this section. Second, we present multiple specifications that include different covariate sets, $\mathbf{X}$. In addition to the naïve estimates, we present specifications that control for a range of maternal and child-level characteristics and a wealth index.[23] Where we do find positive mediation effects, we show that these effects are robust to a range of specifications, moderately alleviating concerns about potential violations to condition F3.

In this study, we use binary mediators and therefore opt for a more flexible model than the LSEM. Specifically, we estimate the treatment-mediator relationship using a binary dependent variable logistic model. Our set of equations are given by:

$$M_i = F(\beta_0 + \beta_1 T_i + \mathbf{X}_i \beta_2) \tag{F4}$$

$$Y_i = \alpha_0 + \alpha_1 T_i + \mathbf{X}_i \alpha_2 \tag{F5}$$

$$Y_i = \gamma_0 + \gamma_1 T_1 + \gamma_2 M_i + \mathbf{X}_i \gamma_3 \tag{F6}$$

where $F(\eta) = \frac{\eta}{1-\eta}$ is the cumulative density function of the logistic distribution. Here, $Y_i$ is the outcome of interest, either height-for-age Z-scores or cognition score for child $i$, $T_i$ is the treatment status of child $i$, and $M_i$ is one of two mediating variables. Finally, $\mathbf{X}_i$ represents a vector of pre-treatment covariates.

To test if changes in birth spacing mediate our results, we conduct a mediation analysis with a binary variable that takes on a value of 1 if the child's mother conceived another child before follow-up. To test if increases in human capital investments drive results, we conduct the analysis with a binary variable that takes on a value of 1 if either the child or mother had received a clinical health visit within the previous 14 months from the time of the survey.[24,25]

We refer to Imai et al. (2010), which extends the identification approach of Imai et al.

[23] This wealth index is calculated using a principle component analysis (PCA) following the Demographic and Health Survey methodology.

[24] Based on data from the 2015-16 Malawi Demographic and Health Survey as well as qualitative observations from our study site, we note that most women jointly seek postpartum care for themselves and perinatal care for their child due to difficulties associated with care seeking (Malawi and ICF, 2017). This pooling behavior alleviates potential concerns related to our measure of joint maternal and child care seeking in our data.

[25] Due to data limitations, we cannot test the effects of more detailed healthcare use such as vaccinations, perinatal care, or treatment of illness. We note that any observed mediation due to such a course measurement of healthcare use likely represents a lower bound for the true mediating effect of healthcare use. Ideally, we would include more "soft" measures of human capital investments, such as time spent on child care and access to print materials. However, this data was not collected at the endline survey.

(2010) under the following algorithm.[26]

*ALGORITHM 1:* [Parametric Inference]

1. Fit models for the outcome and mediator variables.
2. Simulate model parameters from their sampling distribution.
3. Repeat the following three steps:
   (a) Simulate the potential values of the mediator,
   (b) Simulate the potential outcomes given the simulated values of the mediator,
   (c) Compute the (average) causal mediation effects.
4. Compute summary statistics, such as point estimates and confidence intervals.

As noted, one possible threat to validity is that condition F3 is violated in many settings. In our setting, we may be concerned that our mediator cannot be considered ignorable even after conditioning on a comprehensive set of covariates. This concern motivates us to conduct a sensitivity analysis developed by Imai et al. (2010) and extended to the non-linear case by Imai et al. (2010). In our sensitivity analysis, we introduce a sensitivity parameter, $\rho$, that identifies the correlation between the error terms in equations F5 and F6. To test our specification for sensitivity to violations of assumption F3, we re-estimate Algorithm 1 under different values of $\rho$. Here, examining the value of $\rho$ for which the ACME equals 0 tells us how much of the variation between $Y$ and $M$ would need to be explained by a potential confounder for an identified mediation estimate to be invalid.

## F.1 Results

We present the results of our mediation analysis in Table F1. Our results display some evidence of mediating effects of both healthcare usage and birth spacing on height for age z-scores. However, our estimates are only ever significant to the 5 percent level for the unadjusted birth spacing mediation. Most other coefficients are significant to the 10 percent level, with the healthcare usage specification losing significance in the fully adjusted model. Our sensitivity analysis suggests that the results using birth spacing as the mediator are much more robust than those using healthcare usage, suggesting that any omitted variable would need to explain 30 percent of the joint variation between pregnancies and z-scores to

[26] We implement this algorithm and the sensitivity analysis presented below using the **`mediate`** package in Stata 16.

invalidate our results. Taking the estimates of the greatest magnitude for each mediator, we can explain roughly 27.9 percent of the adjusted treatment effect that we observe in Table F1.

Table F1: Mediation Results for Height-for-Age Z-Score, Year 1

| | Mediator: Healthcare Usage | | | Mediator: Second Pregnancy | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ITT on Mediator | 0.102** | 0.114** | 0.0907** | -0.0376** | -0.0378* | -0.0383** |
| | (0.0442) | (0.0455) | (0.0446) | (0.0188) | (0.0194) | (0.0193) |
| ACME | 0.0319* | 0.0385* | 0.0273 | -0.056** | -0.0449* | -0.0411* |
| | [-0.007,0.0911] | [-0.0038,0.0918] | [-0.0139,0.089] | [-0.1394,-0.0025] | [-0.13,0.0077] | [-0.1138,0.014] |
| $\rho$ where ACME =0 | 0.1 | 0.1 | 0.1 | 0.3 | 0.2 | 0.3 |
| Mother and Child Controls | | X | X | | X | X |
| Household Wealth Index | | | X | | | X |
| Observations | 406 | 404 | 398 | 406 | 404 | 398 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. Results presented are for a causal mediation analysis of the ITT results on height-for-age Z-scores. Maternal and child controls include woman-level controls such as the total number of children who are alive, educational attainment of the woman (primary or less versus secondary and higher), age of the woman, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut, and child-level controls such as biological sex, and birth order, month of birth, and a spline fitted to the child's age with knots at 6, 12, 18, 24, and 30 months. These same controls adjust ITT estimations on the mediator. Heteroskedasticity-robust standard errors, clustered by mother, are presented in parentheses and bootstrapped confidence intervals are included in brackets.

The fact that our estimates are attenuated by adding covariates is likely a sign that the sequential ignorability assumption, particularly condition F3, is not met in our analysis. Additionally, the loss of statistical significance accompanying this attenuation likely reflects a lack of statistical power needed for structural equation modeling. While we observe statistically significant and quite large treatment effects on our chosen mediators, the imprecision of our estimates makes it difficult to make definitive statements about the average causal mediation effect.

Nevertheless, we view these results as suggestive evidence that the underlying mechanisms to our main results, at least in part, correspond to the theory presented in Cavalcanti et al. (2021).

We accompany the results on height-for-age z-scores with mediation results for children's CREDI z-scores in Table F2. Here, we do not find any evidence of mediating effects. Further, our estimates are very close to zero relative to the treatment effects we observe on CREDI scores. It is possible that with a larger sample, we may be able to defects, as the point estimates for healthcare usage match those from the previous mediation in direction.[27]

[27]Indeed, in an earlier version of this paper, when we treated all index children as our main sample, we

Table F2: Mediation Results for CREDI Z-Score, Year 2

| | Mediator: Healthcare Usage (2017) | | | Mediator: Second Pregnancy | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| ITT on Mediator | 0.114** | 0.098* | 0.103** | -0.0293 | -0.0309 | -0.0279 |
| | (0.0495) | (0.0512) | (0.0508) | (0.0305) | (0.0315) | (0.032) |
| ACME | 0.0047 | 0.0033 | 0.0041 | 0.0015 | 0.0007 | 0.000413635 |
| | [-0.0092,0.0232] | [-0.0089,0.018] | [-0.0123,0.028] | [-0.0095,0.0145] | [-0.0119,0.0169] | [-0.0116,0.0126] |
| $\rho$ where ACME =0 | - | - | - | - | - | - |
| Mother and Child Controls | | X | X | | X | X |
| Household Wealth Index | | | X | | | X |
| Observations | 332 | 332 | 323 | 362 | 362 | 355 |

$**p < 0.05$, $*p < 0.1$

Notes: For all columns, the unit of observation is a child. Results presented are for a causal mediation analysis of the ITT results on height-for-age Z-scores. Maternal and child controls include woman-level controls such as the total number of children who are alive, educational attainment of the woman (primary or less versus secondary and higher), age of the woman, religion (Christian versus other), tribal ethnicity (Chewa versus other), and age of sexual debut, and child-level controls such as biological sex, and birth order, birth month, and a spline fitted to the child's age with knots at 6, 12, 18, 24, and 30 months. These same controls adjust ITT estimations on the mediator. Heteroskedasticity-robust standard errors, clustered by mother, are presented in parentheses, and bootstrapped confidence intervals are included in brackets.

However, with the current sample, it is difficult to draw inferences on causal mechanisms from this analysis. Instead, this analysis highlights the need for further research on this topic with a larger sample and richer data on child investments.

did find that healthcare usage was able to explain roughly 30 percent of our treatment effect