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

Daniel Maggio† Mahesh Karra‡ David Canning§

November 7, 2021

Abstract

We conduct a randomized control trial that improves pregnant and immediate postpartum women's access to family planning, including counseling, free transport to a clinic, and financial reimbursement for family planning services over a two-year period. We study the effects of our intervention on child growth and development, motivated by a recent extension to Becker's theory of fertility that accounts for the uncertainty in fertility that couples may face. We study 720 children born to participating women and find that children born to mothers assigned to the intervention arm were 0.23-0.25 standard deviations taller for their age and were 6.7-6.9 percent less likely to be stunted within a year of exposure to the family planning intervention. We find that children born to mothers assigned to the intervention arm scored 0.1 standard deviations higher on a caregiver reported measure of cognitive development after two years of intervention exposure. Using a causal mediation analysis, we show that 25 percent of the effects on heights and 30 percent of the effects on cognition can be explained by increases in healthcare use. These results are consistent with the extended model of the quality-quantity trade-off and suggest that reduced fertility uncertainty induces couples to increase investments in the health of their children. Our results also suggest that improved access to family planning may have positive downstream effects on child health that extend beyond outcomes related to contraceptive use and fertility.

*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 makes use of original data collected by Canning and Karra with support from Innovations for Poverty Action (IPA) in Malawi. The authors would 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, which comprised of 22 enumerators and 7 counselors over a three year study period. This project was supported by two grants from the William and Flora Hewlett Foundation and the Human Capital Initiative's 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 Institutional Review Board (protocol number IRB16-0421) and from the Malawi National Health Sciences Research Committee (protocol number 16/7/1628). Informed consent was obtained from all participants in the trial. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily 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.

†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 74 million unwanted pregnancies occur in the developing world each year. About half of these pregnancies will end in an abortion (Bongaarts, 2016). For those pregnancies that result in live births, the family must now raise a child that was unplanned and that they are likely unprepared for. While that child will certainly bring immeasurable benefits to the household, they will also bring costs in terms of both time and goods. Healthcare and education are costly; even in a system with free public services, transportation, school uniforms, and supplies can place a large financial burden on families. This is not to mention foregone earnings from time spent on childcare.

While situations such as these can largely be prevented through the use of modern contraceptives, they continue to occur, in part, because contraceptives are costly and inaccessible to many families in the developing world. This uncertainty in fertility, combined with the potentially high cost of an unwanted pregnancy, may prevent families from investing in their children's health and education. The result is a world where child outcomes are worse than those in a world where families were able to perfectly control their fertility.

In this paper, we are motivated by Cavalcanti et al. (2021), which extends the model presented in Becker (1960) to account for the uncertainty in fertility faced by many poor households and assumes that families can use contraceptive methods to reduce this uncertainty. In line with this model, we postulate that family planning services, which reduce costs and inaccessibility of modern contraceptives should lead to increased investment in child health and education on the part of households. Using a randomized control trial (RCT) conducted in Lilongwe, Malawi, we show that, in line with this theory, children born to women with improved access to postpartum family planning through counseling, transport, and financial support have improved health outcomes. In particular, these children were 0.25 SD taller for their age at first year follow-up, and performed 0.1 SD better on a caregiver reported measure of cognitive development at end line. We also show that the gains in growth patterns do not seem to favor one group, while the gains in cognition are disproportionately made by girls and children born to mothers who were postpartum at the start of the intervention. Finally, we use a causal mediation analysis to show that effects on child growth are partially mediated by increased healthcare usage for the mother and child in households with increased access to family planning, and do not seem to be mediated by changes in birth spacing. Further, we rule out mechanisms related to free transportation and accessibility of care by showing that these mediating effects are robust to the exclusion

of families who utilised the transport. Our results contribute to a large literature on the existence of a quality-quantity trade-off in parental preferences (See Rosenzweig and Wolpin (1980) and Angrist et al. (2010) for well known examples) and a smaller literature specifically relating family planning services to child health improvements (see Joshi and Schultz (2013) and Phillips et al. (2006)).

Our analysis focuses on the Malawi Family Planning Study (Henceforth MFPS), a trial conducted from 2016 to 2018 in which women were eligible to be recruited if they were either pregnant or up to six months postpartum at baseline. Women assigned to treatment received six free family planning counselings, free transport to a maternal and child health clinic, and financial reimbursement for family planning services. The data used in this paper comes from three household surveys conducted throughout the length of the study (one baseline, a mid-line, and an end-line). In particular, we use anthropometric data collected for all children under 5 years old in all surveys, as well as a caregiver reported survey of cognitive ability that was added to the end-line survey and measured cognitive ability for all children under 3 years old.

We focus our analysis on children's heights for a number of reasons. First, childhood stunting[1] remains a great challenge globally. Roughly 144 million children are stunted around the world (WHO, UNICEF, and World Bank Group, 2020; Shekar et al., 2021). In Malawi, the context of our experiment, 37 percent of children are considered stunted (Government of Malawi, 2019). While it may not be immediately obvious how children's heights relate to human capital, childhood stunting is often used as a measure of chronic under-nutrition and has been linked to lower educational attainment (Alderman et al., 2006), lower earnings, and higher poverty rates (Maluccio et al., 2009). One explanation for these associations is that stunted children do not develop cognitively in the same way as well-nourished children. Indeed, research has shown that early life stunting is associated with decreased dendritic density in the hippocampus and occipital lope and reduced myelination of axion fibers (Cordero et al., 1993; Prado et al., 2019). Under this explanation, one might expect any program that decreases stunting rates to also improve cognition. Thus, we also center our analysis on a caregiver reported measure of cognitive ability.

We are not the first to suggest that family planning services may reduce childhood stunt-

[1] A child is considered stunted when their height is two or more standard deviations below the height of an average "healthy" child of the same age, as determined by the World Health Organization's (WHO) Multi-Center Growth Reference Study (de Onis et al., 2004)

ing. A Lancet series on Maternal and Child Malnutrition put forward a framework of international action by Black et al. (2013) that included family planning services as a "nutrition sensitive" intervention. However, despite a solid theoretical backing, there is little empirical evidence to support the claim that increased access to contraceptives improves children's health outcomes. Our experimental design allows us to determine if the authors were correct to include family planning services in their framework. As mentioned previously, we find that the intervention increased the heights of children, supporting the authors' decision.

The rest of this paper proceeds as follows. Section 2 presents our model and conceptual framework. Section 3 describes the design of the MFPS. Section 4 presents our empirical strategy. Section 5 presents the results. Section 6 concludes. In the Appendixes we include analyses that are secondary to the main analysis presented. These analyses include heterogeneous treatment effects of the intervention, an intervention component analysis, and a set of robustness checks.

2 Theoretical Framework

Much of the work done on the economics of fertility to this point has used Becker (1960) as their workhorse model. In Becker's seminal paper on the topic, he acknowledges the impact contraceptives may have on child health outcomes, stating that "...an increase in contraceptive knowledge would raise the quality of children as well as reduce their quantity" (Becker, 1960). However, in his following adaptations of the model, Becker makes no mention of contraceptives (Becker and Tomes, 1976; Becker and Lewis, 1973).

The difficulty in determining the effect of contraceptives on child quality within Becker's model lies in the fact that fertility is assumed to be a deterministic function. In other words, Becker implicitly assumes that families can perfectly target their realized fertility. Under this assumption, families are either perfectly using contraceptives, or have no need for contraceptives. In either situation, this assumption is hard to reconcile with the 74 million unwanted pregnancies that occur in developing countries each year.

To accommodate this discrepancy in Becker's model, Cavalcanti et al. (2021) extends Becker's theory of fertility to allow for a stochastic realization of an underlying fertility preference. The authors begin with a Overlapping Generations Model (OLG) in which individuals live for 3 periods. When couples are young, they consume, and save. When they are old, they consume their savings from the previous period. Further, Couples are only fertile

when they are young, so this is when all pregnancies occur and human capital investments are made. As with traditional quality-quantity models, couples are assumed to gain utility from both birth parity ("quantity"), as well as the human capital accumulation of each child ("quality"). However, child are assumed to be costly in terms of both human capital investments and time.

The defining characteristic of the Cavalcanti et al. (2021) model is that couples do not perfectly choose the number of children they have, as they do in Becker's models. Instead couples face uncertainty in fertility that can be reduced through the use of modern contraceptives. However, modern contraceptive come with costs. Further, these costs may not be purely monetary. If contraceptives make intercourse less enjoyable or violate religious/cultural beliefs, there may also be large psychic costs associated with the use of modern contraceptives.

In our context, the key finding of the model is that both the cost and effectiveness of contraceptive methods enter into the household's decisions on both consumption and human capital investments. The authors' findings show us that when contraceptive cost decreases, or contraceptive effectiveness increases, couples should increase both their consumption when they are young and the investments they make in their children. As our intervention aimed to both decrease the effective price of contraceptives and increase women's knowledge of modern methods, thus increasing the effectiveness of the methods available to women, the model then suggests two channels through which we can expect our intervention to improve child health. First, improving access to contraceptives should directly induce an increase in spending on children's health. As healthcare usage increases, and feeding practices improve, these investments should carry over to health outcomes. Second, we see that improved access to contraceptive should induce households to consume more in the immediate period. If some of that increased consumption is directed towards inputs to the child health production function such as improved maternal nutrition or improved home environments, we would expect an improvement in children's health outcomes.

In Figure 1, we summarize the above mechanisms into our proposed causal framework. In that causal framework, we include one alternative mechanism through which family planning services may improve child health, namely through improved inter-birth intervals and reduced maternal depletion. Healthy inter-birth intervals are one of the main goals of family planning services, and while evidence is incomplete, multiple studies suggest family planning services do assist families in properly timing their births (Bhatia et al., 1980; Cleland et al.,

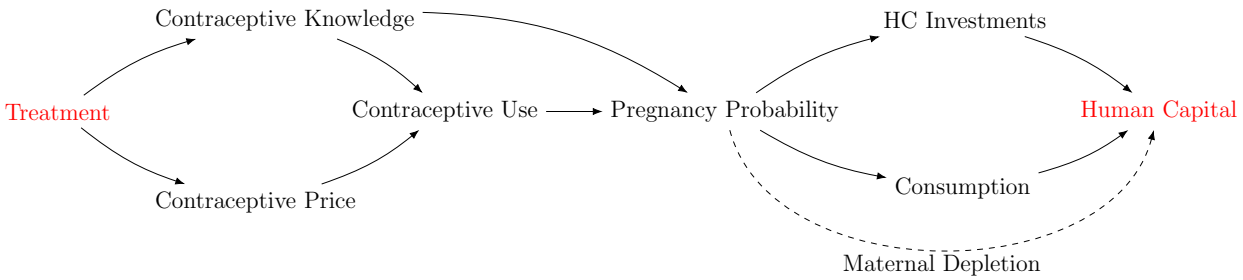


Figure 1: Proposed Causal Framework

2012; Debpuur et al., 2002; Joshi and Schultz, 2013). Indeed, in Karra et al. (2020) we show that pregnant and postpartum women who received our intervention were at lower risk of having a second pregnancy within 24 months (HR:0.575, 95% CI: 0.393-0.843) or giving birth a second time within 33 months (HR: 0.583, 95% CI: 0.346-0.843). Short birth intervals have been associated with increase in risk of neonatal, infant, and child motality (Rutstein, 2006), as well as rates of stunting and wasting (Rutstein, 2006; WHO, 2005; Huttly et al., 1992) so it stands to reason that by helping families better space their births, family planning services may also improve child health outcomes. However, in this paper, we present results for the child that made mothers eligible to be recruited into the MFPS (henceforth referred to as the "index" child). Since women were pregnant or postpartum at baseline, the preceding birth interval for the index child was determined prior to recruitment into our study. Regardless, while the preceding birth interval for index children is plausibly exogenous, from Karra et al. (2020), we know that index children in the treatment group are less likely to have a competing sibling during their first two years of life. To determine if our results are driven by this differential birth spacing, we conduct a causal mediation analysis and find that the mediating effects of birth spacing on children's heights are statistically insignificant and precisely measured, leading us to believe our results are not driven through this channel.

3 Study Design

We focus our empirical work on a Randomized Control Trial (RCT) conducted in Lilongwe, Malawi between November 2016 and November 2018. Below, we provide an abbreviated description of the trial for the sake of continuity and to allow this paper to stand alone. Readers looking for a more detailed description of the study design and implementation should see Karra and Canning (2020) which describes the protocols of the experiment in greater depth.

The study recruited women who were either pregnant or immediately post-partum and randomly assigned them to either a treatment or control arm. Women assigned to receive the intervention received a comprehensive family planning package consisting of postpartum family planning counseling, free transport to a clinic, and financial reimbursement for family planning methods.

In this paper, we present results for a subset of children that resulted from the pregnancy for which the women was enrolled in the study (the "index" children). In some estimations, we also extend this sample to include non-index children in the household who were under 2 years at baseline because these children were within their crucial first 1,000 days since conception and their linear growth patterns could plausibly be improved by changes in parental behavior.

The data used in this paper comes from three separate surveys conducted throughout the length of the trial. A baseline survey was implemented from September 2016 to January 2017 followed by two follow-up surveys. Data collection for the first follow-up survey began in August 2017 and was completed in February 2018, and data collection for the second follow-up survey began in August 2018 and was completed in February 2019.

Study Sample

Eligibility

Women were recruited into the study if at the time of the baseline survey they were:

1. Married
2. Currently pregnant or had given birth within the previous 6 months
3. Between the ages of 18 and 35
4. A permanent resident of Lilongwe, Malawi

In addition to these inclusion criteria, no two women were enrolled from the same household. If two women in the same household were eligible to participate in the study, the younger of the two was enrolled in the study. To avoid spillover effects, women chosen to be enrolled in the study were sufficiently distant from each other (at least five houses apart).

For women who were enrolled in the study, anthropometric data[2] was collected from their children if enumerators were able to obtain consent for measurement and the child was:

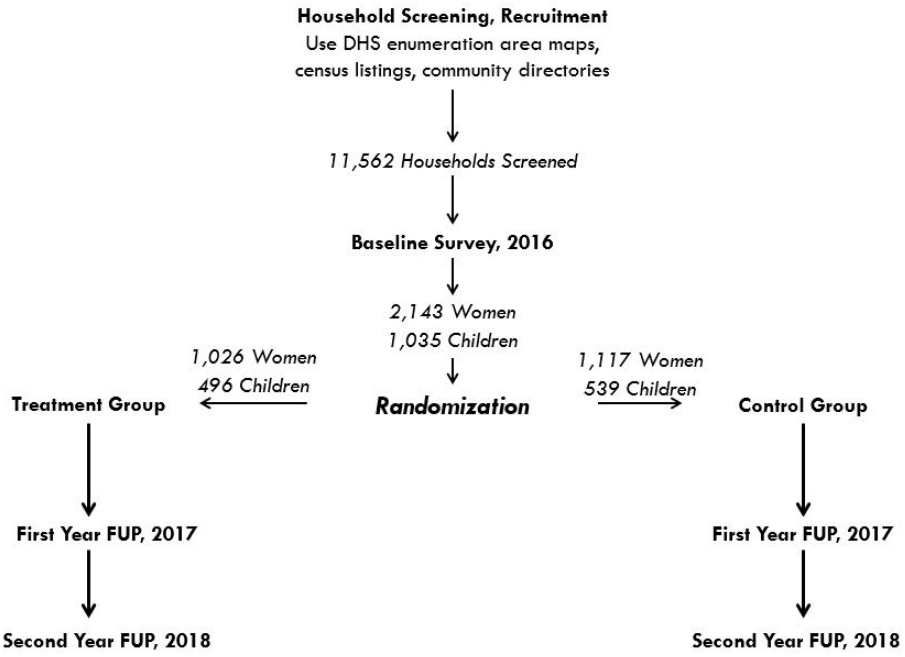
[2]Heights, weights, and anemia statistics were collected during all three survey waves. In the second follow-up survey, a set of questions were added to measure the cognitive development of the child.

1. Under the age of six years at baseline or born after baseline
2. Identified as the biological or adopted child of the woman enrolled in the main study
3. Present in the household at the time of interview

Sample Size and Randomization

At the conclusion of screening and recruitment, 2,143 women were enrolled in the study. Following the baseline survey, women were randomly[3] assigned to treatment following a co-variate balanced randomization procedure (Bruhn and McKenzie, 2009). In total, 1,026 women were assigned to the treatment arm of the study while 1,113 women were assigned to the control group. Figure 2 shows the randomization protocol and follow-up survey scheme.

Figure 2: Randomization Protocol



To accompany mothers, 1,035 index children were enrolled in the study at baseline. Since half of the mothers were pregnant at baseline, 781 children were added to the study sample by first year follow-up leading to a total sample size of 1,816 children. Since some pregnant mothers at baseline were lost to first year follow-up, there were 327 pregnancies at baseline that did not result in a child as part of our sample. Of those 1,816 children, 953 were born to women in the control group and 863 were born to mothers in the treatment group. Table

[3]Women were randomized individually into treatment and control groups. For that reason, the standard errors presented for the estimates are not clustered by any stratification, but rather within family to account to genetic correlations.

1 presents baseline balance for key statistics for women and children who were enrolled at baseline. We see from Table 1 that despite a small difference in family planning experience, women are generally balanced across key characteristics. We also see that there are no significant differences among children in the treatment group relative to the control group. Further, through joint significance tests, we cannot reject that either children or women are systematically different across intervention arms.

Table 1: Baseline Covariates by Treatment Group

	(1) Full Sample	(2) Treatment	(3) Control	(4) Difference (2) - (3)
Sample Women				
Current Use of FP (1 = Yes)	0.237	0.239	0.235	0.003
Long Acting Method Use (1=Yes)	0.034	0.034	0.033	0.001
Injectable Use (1=Yes)	0.187	0.189	0.185	0.004
Implant Use (1=Yes)	0.031	0.031	0.031	0.000
Ever Use of FP (1 = Yes)	0.755	0.775	0.736	0.039**
Woman's Age (Years)	24.580	24.657	24.509	0.148
Total Number of Children	2.382	2.417	2.349	0.068
Primary Education (1 = Yes)	0.587	0.586	0.588	-0.001
Secondary Education (1 = Yes)	0.413	0.414	0.412	0.001
Tertiary Education (1 = Yes)	0.023	0.019	0.028	-0.009
Religion (1 = Christian)	0.827	0.838	0.816	0.023
Ethnicity (1 = Chewa)	0.418	0.414	0.421	-0.007
Woman Works (1 = Yes)	0.096	0.099	0.093	0.007
Age of First Cohabitation (Years)	18.855	18.904	18.810	0.094
Pregnancy Status (1 = Yes)	0.515	0.515	0.516	0.001
Observations	**2,139**	**1,026**	**1,113**	
P-Value (Joint Significance)				**0.556**
Sample Children				
Child's Age (Months)	2.909	2.901	2.915	-0.014
Child's Biological Sex (1=Male)	0.493	0.504	0.482	0.022
Height-for-Age Z-Score	-0.171	-0.157	-0.184	0.027
Weight-for-Age Z-Score	0.319	0.362	0.279	0.083
Observations	**1,035**	**496**	**539**	
P-Value (Joint Significance)				**0.906**

The Intervention

Women assigned to the treatment group of this study were offered a comprehensive, multi-part postpartum family planning package over a two-year period. Designed in coordination with local health authorities, NGOs, and communities, the intervention was designed to overcome multiple barriers to access to care. As such the package combined family planning counseling with free transport and financial reimbursement to overcome barriers due to knowledge barriers, as well as geographic and financial accessibility issues. By offering a comprehensive package, this study aims to determine the true effects of postpartum family planning services, unbiased by accessibility issues.

In the first component of the treatment, women were offered up to six free family planning counselings. The sessions, conducted by trained counselors, were no longer than an hour and took place in the home. During the sessions, women received information on a full range of contraceptive methods and their potential side effects. women also received informational pamphlets at each sessions that focused on the benefits of birth spacing and side effects management. Importantly, women were informed of the other treatment components during the first counseling session, meaning that all women who were properly enrolled into the treatment arm received some form of the treatment.

In addition to free counseling, women were also offered free transportation to a local private family planning clinic, the Good Health Kauma Clinic which offered women a comprehensive listing of family planning services [4]. Transport was conducted by a driver hired exclusively for the project and was always accompanied by a female field manager to ensure that women always had access to transport and mitigate any social stigma surrounding a woman traveling alone in the company of another man.

Finally, women assigned to the treatment arm received up to 17,500 MKW (∼$25.00 USD) in financial reimbursement for any costs incurred while receiving family planning care 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. Though the 17,500 MWK was non-transferable to other services, women were allowed to redeem the reimbursement over multiple visits throughout the two-year intervention period.

Contrary to women assigned to the treatment arm, women assigned to the control arm received publicly available information on family planning methods and their nearest family planning clinic. The women were given this information during the baseline survey and were only contacted again for follow-up surveys.

Follow-Up

During designated one and two year follow-up periods, the entire sample of women was re-interviewed to collect information on follow-up outcomes to be used for data analysis. During the first year follow-up, 1,773 women were re-interviewed, leading to a 82.9 percent

[4]Services offered by the Kauma clinic include insertion and removal of long-acting methods, referral for sterilization, and counseling.

follow rate for sample women. In total, information was collected on 1,547 index children. Of those children, 790 were interviewed at baseline, a 76.32 percent follow rate, and 757 were born during the first year of the intervention.

During the second year follow-up period 1,672 women were re-interviewed, a 77.2 percent follow rate. In total 1,376 children were interviewed including children alive at baseline and those born during the study resulting in a 75.7 percent follow rate. Unfortunately, during 2017 there was a vampire scare in Malawi that forced our partner IPA Malawi to suspend anthropometric measurement midway through our year 2 follow-up (Reuters, 2017).[5] For this reason, we do not have sufficient data on heights to conduct analysis using data from the second year follow-up and the results presented for height-for-age are from the year 1 follow-up survey.[6] Since a child's height is path dependent, it is likely that any treatment effect found in the year 1 follow-up would also be present during the year 2 follow-up. Additionally there was a cognitive development module (discussed in detail in Section 5) added during the year 2 follow-up. Since this cognitive development module consisted of caregiver reported metrics, it was not subject to IPA's suspension of anthropometric measurement. This means that while results for height-for-age are reported from the year 1 follow-up, results for cognition are reported from the year 2 follow-up.

Table 2 compares key covariate characteristics of children surveyed in 2017 to those who were lost to follow-up. We compare differences in remaining participants across treatment and control groups, differences across remaining and lost participants in both groups, and differences in lost participants across both groups. Among remaining mothers, there is a 7 percent difference in contraceptive experience, while treatment mothers lost to follow up were ten percent more likely to be christian than those lost to follow-up in the control group. While these results may pose a threat to the validity of our results, through a joint

[5]According to a Reuters report in October, 2017 "[a] UNDSS report said at least five people had been killed in the area since mid-September by lynch mobs accusing them of vampirism. It said mobs searching for vampires have been mounting roadblocks in the district, raising security concerns." Due to ongoing stories of blood sucking and the existence of vampires, the United Nations made the decision to withdraw many staffers from southern districts. In an effort to protect staffers, IPA Malawi made the decision to suspend any survey module that collected blood samples from children. In our survey, the anthropometric data collection had to be suspended because it included an anemia test for women and children, meaning that children who were interviewed after the suspension were not measured.

[6]Height-for-age is a common measurement of nutritional status used to reflect health and nutrition over the lifespan. Since a child's height is a cumulative measure of health and nutrition, any difference found at first year follow-up may be expected to still be present during second year follow-up. This difference should still be present even if there is no difference in feeding practices between the control and intervention arms. The only situation where one may expect to see a dissipation of the difference is if there is a reversal in the treatment effect between the first and second year, which seems quite implausible.

Table 2: Baseline Table of Key Baseline Covariates Comparing Children Lost to Follow-Up

	Remaining Participants		Lost to Follow-up					
	Treatment (N=381)	Control (N=408)	Treatment (N=115)	Control (N=130)	Difference (1) - (2)	Difference (1) - (3)	Difference (2) - (4)	Difference (3) - (4)
	(1)	(2)	(3)	(4)				
Maternal Characteristics								
Current Use of FP (1 = Yes)	0.496	0.474	0.478	0.500	0.022	0.018	-0.026	-0.022
Long Acting Method Use (1=Yes)	0.068	0.078	0.070	0.046	-0.010	-0.001	0.032	0.023
Injectable Use (1=Yes)	0.383	0.367	0.409	0.400	0.016	-0.025	-0.033	0.009
Implant Use (1=Yes)	0.063	0.071	0.061	0.046	-0.008	0.002	0.025	0.015
Ever Use of FP (1 = Yes)	0.871	0.797	0.809	0.785	0.074***	0.063*	0.012	0.024
Woman's Age (Years)	25.370	25.122	24.400	24.046	0.248	0.970**	1.076**	0.354
Total Number of Children	2.488	2.487	2.391	2.192	0.002	0.097	0.294**	0.199
Primary Education (1 = Yes)	0.533	0.545	0.687	0.646	-0.012	-0.154***	-0.101**	0.041
Secondary Education (1 = Yes)	0.467	0.455	0.313	0.354	0.012	0.154***	0.101**	-0.041
Tertiary Education (1 = Yes)	0.029	0.042	0.017	0.008	-0.013	0.011	0.034*	0.010
Religion (1 = Christian)	0.856	0.836	0.861	0.754	0.019	-0.005	0.082**	0.107**
Ethnicity (1 = Chewa)	0.436	0.408	0.386	0.419	0.027	0.050	-0.010	-0.033
Woman Works (1 = Yes)	0.089	0.098	0.052	0.046	-0.009	0.037	0.052*	0.006
Age of First Cohabitation (Years)	19.150	19.049	18.737	18.703	0.101	0.413	0.346	0.034
Child Characteristics								
Child Age (Months)	2.950	2.924	2.740	2.888	0.026	0.210	0.036	-0.148
Biological Sex (1=Male)	0.517	0.489	0.461	0.462	0.028	0.056	0.027	-0.001
Height-for-Age Z-Score	-0.137	-0.261	-0.219	0.108	0.123	0.082	-0.368	-0.327
Weight-for-Height Z-Score	0.374	0.376	0.322	-0.119	-0.001	0.053	0.494	0.440
P-Value (Joint Significance)					**0.244**	**0.915**	**0.145**	**0.345**

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

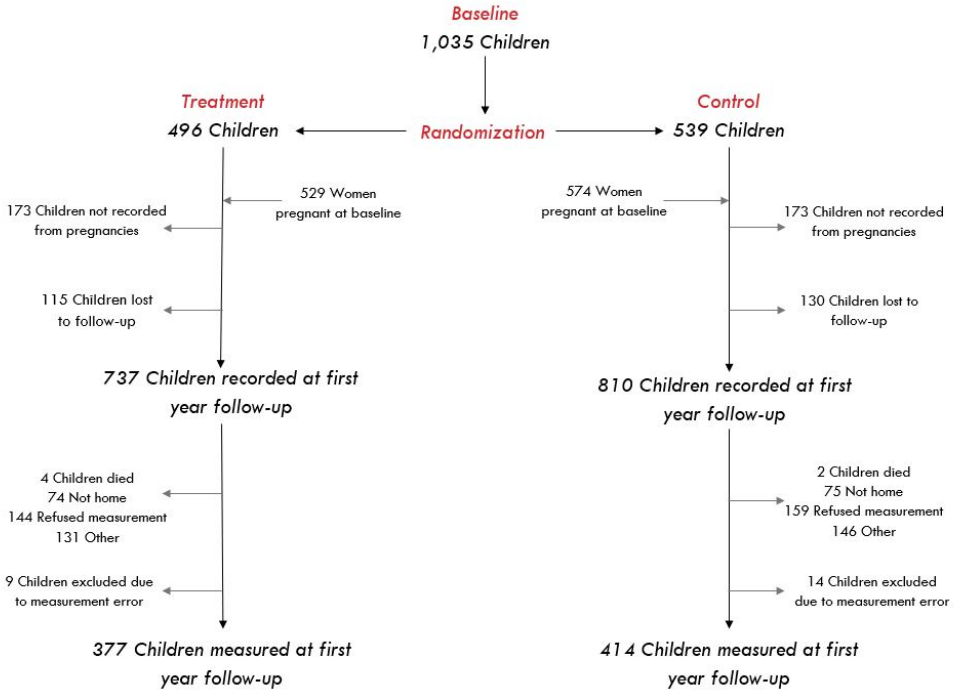
Notes: For all columns, the unit of observation is a child. Covariates are only reported for children who were recorded at baseline as we do not observe on children who were not yet born at baseline and were lost to follow-up. This analysis is therefore incomplete. Stars are based on the critical value from individual t-tests.

significance test, we are unable to reject the hypothesis that mothers and children are not systematically different across treatment groups for either remaining or lost participants. We do see however, that participants lost to follow-up in both intervention arms were older and less educated than those who remained in the sample. While these differences do not pose a threat to our internal validity since they are similar across intervention arms, and the differences are not jointly significant across all covariates, it does mean that we are conducting our analysis on a younger and more educated sample than the full sample of the study. This is important to note because it means that the results within this paper are not presented for the presumed cohort of a scaled up version of our intervention.

Another threat to validity of our results is the low rate of anthropometric measurement among children who were interviewed at follow up. Of the 1,537 children present in our data set, only 791 were measured. Of those that were not measured, 6 had unfortunately died, 149 were not home at the time of interview, 303 did not have consent granted by their parents, and 277 were not measured for other reasons [7]. Figure 3 displays the reasons for attrition and

[7]In addition to those not measured, a small group of 23 children (1.4%) were excluded from the sample because their Z-scores fell outside of the WHO recommended exclusion criteria, likely because of measurement error in age.

Figure 3: Attrition and Non-Measurement by Treatment Group



non-measurement broken down by treatment groups. Such high rates of non-measurement pose a threat to identification if non-measurement is differential across treatment and control. Table 3 compares key covariate characteristics of children measured in 2017 to those who were not in both intervention arms. We compare differences in measured participants across treatment and control group, differences across measured and non-measured participants in both groups, and differences in non-measured participants across both groups. For both measured and non-measured participants, we do not find jointly significant differences in their observable characteristics. However, non-measured children do seem to be slightly older if they were assigned to the treatment group. This difference is driven by the fact that in the treatment group, older children were significantly less likely to be measured. While this result is not robust to a Bonferroni or Hochberg correction (Bonferroni, 1936; Hochberg, 1988), out of an abundance of caution we control for the child's age in all specifications throughout this paper to correct for this imbalance[8].

[8]We will show is Section 4 that we do so using a flexible form that allows the relationship between outcomes and age to change over the lifespan

Table 3: Balance Table Comparing Children Lost to Non-Measurement

	Child Measured		Child Not Measured					
	Treatment (N=377)	Control (N=414)	Treatment (N=360)	Control (N=396)	Difference (1) - (2)	Difference (1) - (3)	Difference (2) - (4)	Difference (3) - (4)
	(1)	(2)	(3)	(4)				
Maternal Characteristics								
Current Use of FP (1 = Yes)	0.249	0.220	0.264	0.260	0.030	-0.015	-0.040	0.004
Long Acting Method Use (1=Yes)	0.032	0.046	0.039	0.033	-0.014	-0.007	0.013	0.006
Injectable Use (1=Yes)	0.196	0.164	0.200	0.207	0.032	-0.004	-0.043	-0.007
Implant Use (1=Yes)	0.029	0.039	0.036	0.033	-0.009	-0.007	0.006	0.003
Ever Use of FP (1 = Yes)	0.806	0.758	0.792	0.753	0.048	0.015	0.006	0.039
Woman's Age (Years)	24.639	25.072	25.100	24.581	-0.433	-0.461	0.492	0.519
Total Number of Children	2.446	2.498	2.494	2.371	-0.052	-0.049	0.126	0.123
Primary Education (1 = Yes)	0.533	0.582	0.550	0.545	-0.049	-0.017	0.037	0.005
Secondary Education (1 = Yes)	0.477	0.430	0.469	0.465	0.048	0.008	-0.035	0.005
Tertiary Education (1 = Yes)	0.021	0.041	0.022	0.035	-0.020	-0.001	0.006	-0.013
Religion (1 = Christian)	0.838	0.819	0.836	0.841	0.019	0.002	-0.022	-0.005
Ethnicity (1 = Chewa)	0.427	0.437	0.425	0.414	-0.010	0.002	0.023	0.011
Woman Works (1 = Yes)	0.101	0.094	0.106	0.111	0.007	-0.005	-0.017	-0.006
Age of First Cohabitation (Years)	18.949	18.968	19.083	18.788	-0.019	-0.134	0.181	0.295
Pregnancy Status (1 = Yes)	0.499	0.473	0.461	0.518	0.025	0.038	-0.044	-0.057
Child Characteristics								
Child Age (Months)	10.794	10.936	11.341	10.848	-0.142	-0.547**	0.088	0.493*
Biological Sex (1=Male)	0.477	0.502	0.536	0.523	-0.025	-0.059	-0.020	0.013
Height-for-Age Z-Score	0.005	-0.197	-0.322	-0.347	0.202	0.327	0.150	0.025
Weight-for-Height Z-Score	0.424	0.392	0.312	0.353	0.032	0.111	0.039	-0.040
P-Value (Joint Significance)					**0.486**	**0.679**	**0.124**	**0.129**

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

Notes: For all columns, the unit of observation is a child. Maternal characteristics are reported for all children surveyed at first year follow-up. Stars are based on the critical value from individual t-tests.

Descriptive Statistics

Table 4 presents summary statistics for children[9] during the baseline survey, first year follow-up, and second year follow-up. Height-for-age Z-scores are not reported during the year 2 follow up because of the lack of data discussed above. We see that the variance on our height-for-age is slightly higher than WHO recommendations (de Onis et al., 2004). This is likely because our sample is over-representative of younger ages when heights are changing rapidly. It may also be because of measurement error in children's ages. However, there is no reason to believe that this measurement error is differential across treatment groups. Therefore, this measurement error should be considered classical and should not bias our treatment effect estimates.

[9]Readers searching for summary statistics for women enrolled in the study should see Karra et al. (2020), which examines the effect of the intervention on first line family planning outcomes.

Table 4: Descriptive Statistics for Sample Children

	(1) Mean	(2) SD	(3) Min	(4) Max	(5) N
Baseline					
Child's Age (Months)	2.909	1.78	0	7.4	1,035
Child's Biological Sex (1=Male)	0.493	0.5	0	1	1,035
Birth Order	2.367	1.391	1	11	1,035
Height-for-Age Z-Score	-0.171	1.874	-5.9	5.9	557
Midline (Year 1 Follow-Up)					
Child's Age (Months)	10.973	3.627	1.2	22.3	1,547
Child's Biological Sex (1=Male)	0.509	0.5	0	1	1,547
Birth Order	2.38	1.344	1	11	1,547
Height-for-Age Z-Score	-1.139	1.721	-5.8	5.5	791
Endline (Year 2 Follow-Up)					
Child's Age (Months)	22.751	3.794	9.8	33.2	1,376
Child's Biological Sex (1=Male)	0.501	0.5	0	1	1,376
Birth Order	2.412	1.357	1	11	1,376
CREDI Score	53.183	2.545	45.5	58.8	727

4 Empirical Strategy

Intent-to-Treat Effects

The main results presented in this paper will follow adjusted and un-adjusted Intent-to-Treat (ITT) specifications. The ITT specification represents the average treatment effect (ATE) on children born to mothers in the treatment group and is preferred here to a local average treatment effect estimation (LATE), which represents the ATE for children born to mothers who utilized the program, for two reasons. First, we are interested in how access to family planning affects childhood development at a population level. Restricting our analysis to only children of mothers who took advantage of the program may overestimate the effect of the program because of bias in program uptake for mothers. It is also important to realize that in a scaled-up implementation of our program, not every women will decide to take part. Therefore, an ITT specification better reflects the policy relevancy of integrating family planning services into a campaign to ameliorate stunting.

To understand the second reason an ITT specification is preferred to a LATE specification, we must return to our earlier discussion of the intervention. Due to our experimental design, all women had to receive the first counseling to receive access to the rest of the intervention. This means that even if a woman did not take full advantage of the intervention, she had increased knowledge of contraceptive efficacy and side effects. While there may be

underlying variation in the treatment exposure and intensity, this design choice means that all women properly enrolled into the study were, at least in part, compliers. Further, it is likely that a woman's characteristics determined her usage of the different intervention components. For example, women from households with fewer assets are more likely to have utilized the transport component as they were less likely to have other means of transportation. For this reason, we err on the side of caution and present the, likely more conservative, ITT estimates.

Beyond these reasons to prefer the ITT over the LATE estimation, there is are also causal challenges that prevent us from estimating the LATE. As this intervention was a combined and comprehensive package, there exist no woman who were offered *only* counseling or transport. Therefore, we only have sufficient variation to estimate the LATE for the entire package, which is inherently less interesting than the LATE for each intervention component. However, getting a sense of which components are most effective is interesting and worthwhile enough to warrant exploration regardless. For the interested reader, we conduct this exploration in Appendix B while acknowledging that this analysis should not be considered causal.

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

$$Y_i = \alpha + \beta_T T_i + \delta MOB_i + \gamma f(age_i) + \lambda X_i + \zeta Z_i + \chi_i + \varepsilon_i \tag{1}$$

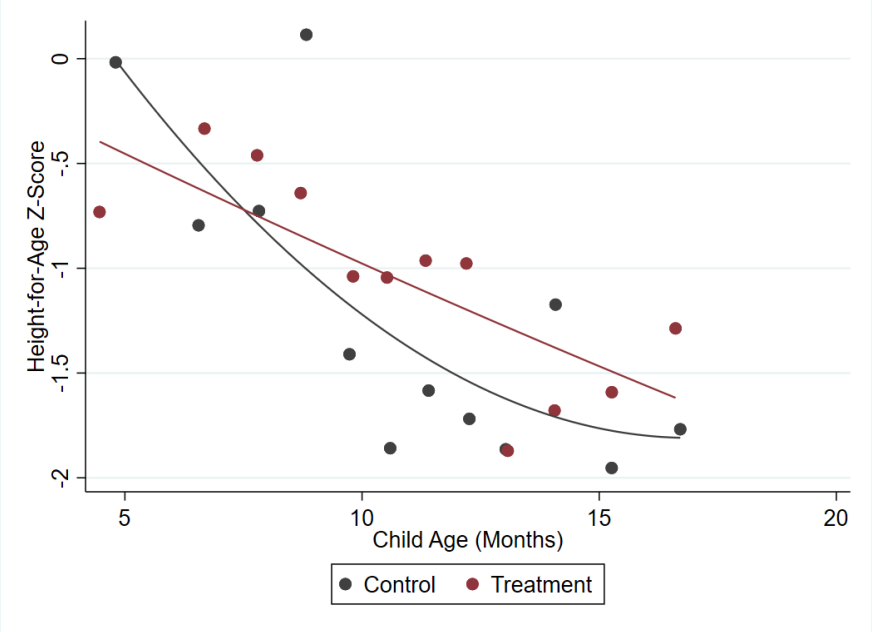
Where Y_i equals the outcome of interest: height-for-age Z-Score, stunting, and cognition scores. T is a dummy variable signifying assignment to treatment, meaning β_T is our coefficient of interest. MOB_i represents a dummy variable for the child's month of birth. X_i is a vector of child characteristics including biological sex and birth order. Z_i is a vector of maternal characteristics including age, employment, education, religion, ethnicity, and total number of children [10]. χ_i is a regional fixed effect defined by the sampling clusters within Lilongwe. $f(age_i)$ represents a flexible specification over the age of the child. Results presented here display a spline with knots at 6, 12, 18, 24, and 30 months. This specification is chosen to allow our model to account for the biological realities of linear growth patterns within the early years of life (Cummins, 2013). Standard errors are clustered by mother to account for any within family genetic correlation.

We present results for "index" children as well as for all children under 24 months of age

[10] In order to utilize all data available to us, any missing covariate is coded to take the value zero and a dummy variable is included that takes the value 1 if that data for that covariate is missing.

at baseline. In our second specification, we include non-index children under the age of 24 months because research has shown that height-for-age Z-scores decrease steadily throughout the first two years of life before becoming reasonably stable (Roche and Himes, 1980). If this relationship holds, we may reasonably expect to see an effect of our treatment on children under 24 months at baseline. Estimates also exclude children who are under 6 months old at the time of follow-up. We exclude these observations because 6 months of age is the WHO recommended age for the transition from exclusive to complementary breast feeding (WHO and PAHO, 2003). If we expect that heights and cognition to be impacted through feeding and care practices, then the exclusion of children under 6 months of age ensures that we are picking up this effect. Additionally, we do find that birth weights in the control group were slightly larger and the exclusion of early age children ensures that our estimates are reflective of early life health and not birth weight. To demonstrate this, Figure 4 presents a bin-scatter of height by age across treatment groups. We see that at early ages, children in the control group are larger, likely reflective of the difference in birth weight, while children in the treatment group become significantly larger once reaching the complementary breastfeeding period.

Figure 4: Height-for-Age by Age and Across Treatment Groups



Causal Mediation Analysis

While finding a large treatment effect on children's heights or cognition scores would be promising, we would not go as far as to say that such a result would provide support for the theory presented in section 2. For that task, we must determine the underlying causal mechanisms driving our results. We attempt to do this using a causal mediation analysis.

Causal mediation analysis is a form of Structural Equation Modeling (SEM) that allows us to identify the *causal mediation effect*, sometimes 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:

$$\delta(t) = Y_i\left(t, M_i(1)\right) - Y_i\left(t, M_i(0)\right) \tag{2}$$

where $M(t)$ is a variable that is affected by treatment and lies on the causal path between treatment and the outcome. Naturally, this means that the mediating variable must occur *after* treatment exposure but *before* the outcome is measured. In other words, the causal mediation effect represents the change in the potential outcome created by changes in the state of the mediating variable, under treatment regime t. Figure 5 displays this relationship in a Directed Acyclic Graph (DAG).

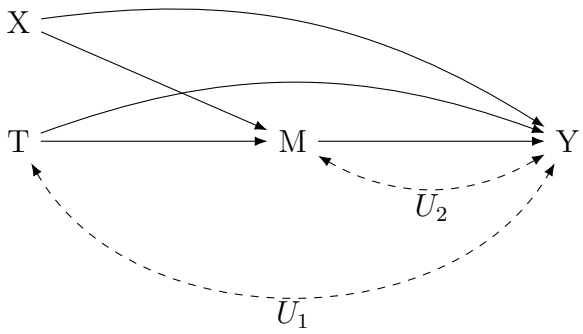


Figure 5: DAG displaying mediator relationship

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.

We are unable to observe both $Y_i(t, M(1))$ and $Y_i(t, M(0))$, meaning the causal meditation effect is unidentified in reality. However, Imai et al. (2010) shows that the *average causal mediation effect*, $\bar{\delta}(t) = E[Y_i\left(t, M_i(1)\right) - Y_i\left(t, M_i(0)\right)]$ is parametrically identified within the Baron and Kenny (1986) Linear Structural Equation Model (LSEM) under a sequential ignorability assumption.

ASSUMPTION 1: (Sequential ignorability)

$$\{Y_i(t', m), M_i(t)\} \perp\!\!\!\perp T_i | X_i = x \tag{3}$$

$$Y_i(t', m) \perp\!\!\!\perp M_i(t) | T_i = t, X_i = x \tag{4}$$

for all $t, t' \in \{0, 1\}$. Since we are presenting results for a randomized experiment, we can easily assume that (3) holds. However, in order for (4) to hold, we must assume the outcome and the mediator are mean independent, conditional on treatment status and a set of covariates, X_i. In this paper, we present analysis for two mediating variables, birth spacing and healthcare usage, which we may be concerned are not orthogonal to child health outcomes, thus violating (4). If, for instance, higher household wealth is associated with better health outcomes and greater healthcare usage, we would expect the mediating effect of healthcare use in our results to have 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 with the inclusion of different covariate sets, X. In addition to the naive estimates, we present estimations that control for maternal and child characteristics, and a principle component analysis (PCA) wealth index[11]. Where we do find positive mediation effects, we show that these effects are robust to different specifications, moderately soothing our concerns about violations of assumption 1.

In this paper, we use binary mediators and therefore opt for a more flexible model than the LSEM. Namely we estimate the treatment-mediator relationship using a binary choice model, rather than a linear probability model. Our set of equations are given by:

$$M_i = \frac{e^{\beta_0 + \beta_1 T_i + \beta_2 X_i}}{1 + e^{\beta_0 + \beta_1 T_i + \beta_2 X_i}} \tag{5}$$

$$Y_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_i \tag{6}$$

$$Y_i = \gamma_0 + \gamma_1 T_1 + \gamma_2 M_i + \gamma_3 X_i \tag{7}$$

Where 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. M_i is one of two mediating variables. To test the birth spacing channel, we mediate with a binary variable that takes the value 1 if the child's mother conceived another child before follow-up. To test if results are driven by increases in human capital investment, we mediate with a binary variable that takes the value 1 if

[11] This wealth index is calculated in the same way as the wealth index found in Demographic and Health Surveys.

either the child or mother had been to a clinical health visit within the previous 14 months [12]. Finally, X_i represents a set of pre-treatment covariates.

To gain inference from our model, we turn to Imai et al. (2010) which extends the identification result of Imai et al. (2010) under the following algorithm[13].

ALGORITHM 1: [Parametric Inference]

1. Fit models for the above 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 causal mediation effects.
4. Compute summary statistics such as point estimates and confidence intervals.

As discussed above, one possible threat to the validity is that (13) is violated in many settings. In our setting, we may may be concerned that our mediator cannot be considered ignorable, even after conditioning on a set of covariates discussed above. This concern drives us to present a sensitivity analysis that was once again developed by Imai et al. (2010) and extended to the non-linear setting by Imai et al. (2010). In our sensitivity analysis, we use a sensitivity parameter, ρ, that represents the correlation between the error terms in equations (15) and (16). To test our model for sensitivity to violations of assumption 1, we re-estimate algorithm 1 under different values of ρ. Here, examining the value of ρ for which the ACME equals zero tells us how much of the variation between Y and M would need to be explained by a potential confounder for our results to be invalid.

[12] In an ideal world, we would have the ability to use a more fine measure of healthcare usage. However, relying on anecdotes from our field team, we understand that most women seek postpartum care for themselves and perinatal care for their child simultaneously due to difficulties associated with care seeking. This reduces our concerns about combining maternal and child care seeking. Additionally, using health seeking over a narrower interval or having more detailed data would likely only increase the association with our outcomes . Therefore, we are confident that this analysis provides a conservative estimate of the ACME.

[13] We conduct the algorithm, and the sensitivity analysis presented below using the mediate package in Stata 16

5 Results

Height-for-Age Z-Score

To measure the effects of family planning services on growth patterns, we report the treatment effect on height-for-age Z-scores. A child's Z-score is the number of standard deviations their height is from a healthy reference [14] child of the same age and sex. Height-for-age Z-scores are a standard proxy for health and nutrition because height is path dependent and determined by, among other things, nutrition and shocks to health such as disease which may temporarily stunt growth. Treatment effects are also reported for moderate and extreme stunting rates. A child is considered to be moderately stunting if his height is more than two standard deviations below the healthy height for a child of their age. They are considered extremely stunted if their height is more than 3 standard deviations that height.

Table 5: ITT Estimates for the Effect of Intervention on Growth Patterns, Year 1

	Index Children		Under 24 months at Baseline	
	(1)	(2)	(3)	(4)
Panel A: Height-for-Age Z-Score				
Treatment	0.231*	0.248**	0.237*	0.243**
	(0.123)	(0.124)	(0.121)	(0.122)
Panel B: Moderate Stunting ($HAZ < -2$)				
Treatment	-0.0665*	-0.0687*	-0.0667*	-0.0677*
	(0.0351)	(0.0357)	(0.0347)	(0.0353)
Panel C: Extreme Stunting ($HAZ < -3$)				
Treatment	-0.0256	-0.0295	-0.0272	-0.0294
	(0.0252)	(0.0265)	(0.0248)	(0.0261)
Age and Birth Month Controls	X	X	X	X
Mother and Child Characteristics		X		X
Observations	720	720	738	738

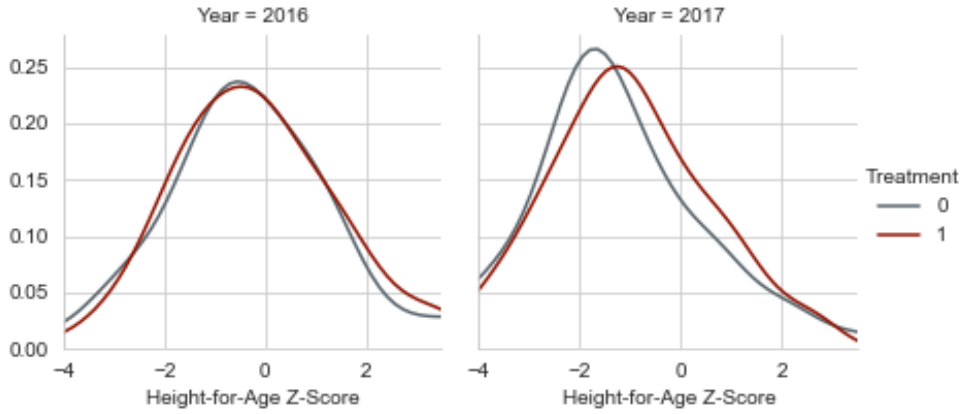
$**p < 0.05$, $*p < 0.1$

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 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 (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 first cohabitation, religion (Christian versus other), and tribal ethnicity (Chewa versus other) and child-level controls such as biological sex, and birth order (included as a set of dummy variables). Sample Cluster fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

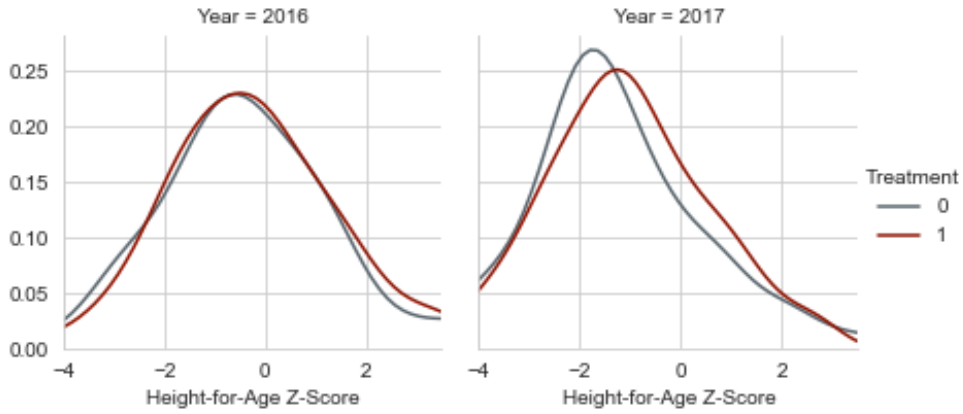
[14] The reference height distribution is taken from the WHO Multicentre Growth Reference Study (MGRS), which measured the heights of healthy children around the world to determine standard growth patterns (de Onis et al., 2004)

Table 5 presents the adjusted and unadjusted ITT estimates for the treatment effect on height-for-age Z-scores, moderate stunting, and extreme stunting during the first year. Treatment effects are reported for index children and all children under 24 months old at baseline. In addition to Table 5, Figure 6 presents non-parametric density estimations for children's height-for-age.

Figure 6: Non-Parametric Density Estimates of Children's Height-for-Age



(a) Index Children



(b) All Children Under 24 Months Old at Baseline

Notes: Density estimations are produced using a Gaussian Kernel and the rule-of-thumb optimal bandwidth. Graphs in the left column show the density estimation for the index children and children under 24 months at baseline. Graphs in the right column show the density estimation for the index children and children under 24 months during first year follow-up.

The combination of Figure 6 and Table 5 suggest that family planning services have a large, causal effect on early childhood growth patterns. We see that estimates do not change significantly when we expand our sample to include non-index children under 24 months

old at baseline. This is likely because index children make up a large majority of children measured at first year follow-up. To accompany the 0.23-0.25 standard deviation treatment effect on children's height for age, we find a 6.7-6.9 percent decrease in moderate stunting. To explain this result, we refer back to Figure 6 which shows that the modal child in the control group is only slightly above the moderate stunting threshold at first year follow-up.[15]. This means that the increase we find in height-for-age pushes a large number of children to the positive side of the stunting threshold. While it is unlikely that there are significant differences in the health of a child directly above the threshold compared to one directly below, this result nonetheless demonstrates that family planning services may be an effective piece of the larger effort to end childhood stunting. In order to expand on these results, we present a subgroup analysis in Appendix Table A1 that estimates the ITT effects disaggregated by the child's biological sex and their mother's pregnancy status at baseline. In this table, we do not observe any substantial differences in the ITT effects across either disaggregation.

To reconcile out ITT results with the model discussed in section 2, we present our causal mediation analysis on children's height-for-age Z-score in Table 6. We present the estimates for the average causal mediation effect using both healthcare usage and birth spacing as mediators. To help understand the underlying pathways, we also report the ITT effect on the mediator using a linear probability model for ease of interpretation. In the instances where we find a significant average causal mediation effect, we also report the value of the sensitivity parameter, ρ, for which the ACME becomes zero.

From Table 6, we find a 0.05 SD average causal mediation effect of healthcare usage on children's height-for-age Z-scores. These results show that increased healthcare usage can explain 20-25 percent of the treatment effect that we find on children's heights. Further, we find large ITT effects on healthcare usage which likely reflect increases in preventative care use, as well as increases in the use of family clinics. The results are remarkably consistent across specifications and a value of $\rho = 0.2$ implies that any relevant covariate would need to explain 20 percent of the combined variance between height-for-age and healthcare use to invalidate these results. While we would be wise to take this sensitivity to possible violations of assumption 1 seriously, we note that the combined R-squared of our ITT models was 0.145 in the most explanatory specification due to the highly genetic nature of children's growth pattern. Therefore, we consider the existence of such a latent variable to be unlikely.

[15]To understand the leftward shift in the modal child from baseline to first year follow-up, refer back to Figure 5 which shows that Z-scores decrease monotonically as children age consistent with findings from previous studies (Roche and Himes, 1980).

Table 6: Causal Mediation Results for Children's Height-for-Age, Year 1

	Mediator: Healthcare Usage			Mediator: Birth Spacing		
	(1)	(2)	(3)	(4)	(5)	(6)
ITT on Mediator	0.100**	0.099**	0.101**	-0.028**	-0.030**	-0.028**
	(0.033)	(0.033)	(0.033)	(0.012)	(0.012)	(0.012)
ACME	0.050**	0.050**	0.052**	-0.018	-0.017	-0.017
	[0.012,0.110]	[0.010,0.103]	[0.013,0.106]	[-0.058,0.011]	[-0.062,0.011]	[-0.059,0.009]
ρ where ACME =0	0.2	0.2	0.2	-	-	-
Mother and Child Controls		X	X		X	X
Household Wealth Index			X			X
Observations	719	719	719	720	720	720

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

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), and tribal ethnicity (Chewa versus other) 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. ITT estimations on the mediator are adjusted by these same controls. heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

Table 6 also shows that the average causal mediation effect for birth spacing is statistically insignificant and precisely measured, allowing us to rule out maternal depletion as the causal pathway through which our intervention improves children's heights. To be clear, this result does not refute the previous literature that cites birth spacing and maternal depletion when explaining the effect of family planning programs on child health outcomes (Cleland et al., 2012). Rather, since the preceding birth interval is exogeously chosen for index children, this result shows that our treatment effects are not driven by a change in timing of the following birth among treatment women.

When taken together, the results of our mediation analysis provide support of the model discussed in section 2. Specifically, they show that family planning services induce an increase in human capital investments among parents that improve the health and human capital growth of their children. While these results are promising, we make two important notes about our results. First, we are unable to test the consumption channel presented above because our study is not powered to detect treatment effects in expenditures, which have high variability. We also do not have data on food consumption that may serve as a proxy for consumption spending. Next, while healthcare usage mediated only a moderately sized portion of our results, we should point out that we are using a very coarse measurement of healthcare usage. We believe with more detailed data on healthcare usage such as infor-

mation on vaccination and antenatal care, we would be able to explain a larger proportion of the treatment effect on children's heights.

Finally, while we believe these results support the theory of Cavalcanti et al. (2021), there is one possible mechanism that may also explain these mediation results. Since one component of our intervention offered women free transport to a clinic, it is possible that this component is driving the mediation results in two ways. The first possibility is that during these trips, women were not only receiving family planning, but also receiving care for their children[16]. A second possibility is that the transport to the clinic introduced to the healthcare system women who had previously not have accessibility. This introduction may have made them more comfortable returning for follow-up visits for themselves and their children. In either case, we would indeed find that healthcare usage has a mediating effect on our ITT estimates but not in a way that supports the relevant theory. In order to address this possibility, we conduct a robustness check in which we exclude children born to mothers who utilized transport from our sub-sample. The results for this robustness check are presented in Appendix Table C1 and show that the mediating results that we find in Table 6 are robust to this exclusion. This implies that our results are indeed driven by changes in behavior due to family planning and not by medical inclusively gained through free transport.

Cognitive Development

In our introduction, we discussed how early life growth patterns are crucial for the development of low and middle income countries, and described a link between childhood stunting and cognitive development. Thus, to compliment our results on children's heights, we report the treatment effects on a score of cognition created by the Caregiver Reported Early Development Instruments (CREDI). The CREDI survey is designed to produce a population level measurement of early childhood development for children under 3 years old. The survey consists of 6 sets of 20 question. Which set is asked depends on which 6-month age bracket the child falls in. The survey focuses exclusively on development milestones and behaviors reported by the caregiver (Fink et al., 2018). Since the survey is exclusively administered to the caregiver and involves no interaction with the child, it was not subject to IPA's halt of anthropometric measurement and we therefore have sufficient data to report results from second year follow-up.

[16]The Good Health Clinic offers a panel of maternal and child health services beyond the family planning services offered.

Table 7 presents the conditional and unconditional estimates of the treatment effects of our intervention on CREDI scores and internally standardized z-scores[17]. The CREDI scores are a scaled, composite score created from the 20 question survey.

Table 7: ITT Estimates for the Effect of Intervention on CREDI Scores, Year 2

	Index Children		Under 24 months at Basline	
	(1)	(2)	(3)	(4)
Panel A: CREDI Short Form Scores				
Treatment	0.301**	0.275*	0.308**	0.276*
	(0.148)	(0.150)	(0.149)	(0.151)
Panel B: CREDI Z-Scores				
Treatment	0.104**	0.0945*	0.106**	0.0947*
	(0.0509)	(0.0515)	(0.0512)	(0.0518)
Age Controls	X	X	X	X
Mother and Child Characteristics		X		X
Observations	727	727	739	739

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

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 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 (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), religion (Christian versus other), and tribal ethnicity (Chewa versus other) and child-level controls such as biological sex, and birth order (included as a set of dummy variables). Sample cluster fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

Our results show that there is roughly a 0.1 standard deviation causal effect of our intervention on the cognitive development of children. Although it is not as large as the effect on height for age, this result is tremendously promising because of the downstream effects it may have on schooling and labor market outcomes. Finding a causal impact on educational attainment or incomes would substantially further the case for family planning as a development strategy.

In Appendix Table A2, we display a subgroup analysis for CREDI Scores disaggregated by child biological sex and pregnancy status at baseline. Unlike the results for height-for-age, we find large and statistically significant differences across the subgroups. In particular, the results seem to be driven by girls and children born to mothers who were postpartum at baseline. These results likely imply that the increases in investments among parents went

[17]Z-scores are standardized off of 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 created by our intervention.

disproportionately towards girls and that family planning counseling was more effective during the postpartum period than during pregnancy. Readers looking for further discussion on these heterogeneous effects should see our discussion in Appendix A.

We once again test for the causal mechanisms underlying our ITT effects by presenting the results of our mediation analysis on the CREDI Z-scores. Again, we mediate our results with healthcare usage and birth spacing. However, in this analysis, we now measure birth spacing as a binary variable that takes the value one if the child's mother has had a second pregnancy (relative to the index pregnancy) by the time of second year follow-up. In contrast, we continue to use our measure healthcare usage from the first year follow-up. We use healthcare usage from the first year follow-up because the visits reported there take place during the child's, extremely formative, first year of life and when their preventative care should include receiving a wide breadth of inoculations including the DPT, MMR, Pneumococcal, polio, and rotavirus vaccines.

Table 8: Causal Mediation Results for Children's CREDI Score, Year 2

	Mediator: Healthcare Usage (2017)			Mediator: Birth Spacing		
	(1)	(2)	(3)	(4)	(5)	(6)
ITT on Mediator	0.100**	0.099**	0.101**	-0.030*	-0.029	-0.030
	(0.034)	(0.034)	(0.034)	(0.018)	(0.019)	(0.019)
ACME	0.031**	0.031**	0.031**	0.003	0.003	0.003
	[0.011,0.057]	[0.010,0.058]	[0.012,0.056]	[-0.005,0.014]	[-0.004,0.015]	[-0.004,0.013]
ρ where ACME =0	0.2	0.2	0.2	-	-	-
Mother and Child Controls		X	X		X	X
Household Wealth Index			X			X
Observations	681	681	681	727	727	727

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

Notes: For all columns, the unit of observation is a child. Results presented are for a causal mediation analysis of the ITT results on CREDI 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), and tribal ethnicity (Chewa versus other) 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. ITT estimations on the mediator are adjusted by these same controls. heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

Table 8 presents the results of this mediation analysis in the same format of the results presented in Table 6. Similar to the results on children's heights, we see that healthcare usage mediates 30 percent of the treatment effect we find on CREDI Z-scores. Once again we find that our results are remarkably consistent across specification and are robust to the exclusion of a covariate that explains up to 20 percent of the covariance between CREDI

scores and care seeking. Unlike the case of height-for-age, our most explanatory estimation for CREDI Z-scores has a combined R squared value of 0.528 which may imply that there exists a latent variable with high influence on both CREDI scores and health seeking. However, as children's cognitive abilities increase throughout the lifespan and the CREDI raw scores increase for each age band, the inclusion of the children's age as a covariate explains a large portion of the variance and therefore we consider the existence of such a latent variable to be unlikely. Finally, we see that the ACME for birth spacing is precisely measured and insignificant. From these results, we can draw similar conclusions about our model as we drew from Table 6. However, these results contribute to that interpretation in that they show the increased investment in human capital leads to an increase in children's cognitive abilities which, more so than their height, should be a predictor of their long term human capital attainment and well-being.

Much like the mediation analysis conducted on height-for-age, the alternative explanation exists that the results are being driven by the transport component. In Appendix Table C2, we present the mediation analysis for CREDI Z-scores on the non-transport sub-sample. We find that the results of our mediation are slightly mediated by this sample exclusion but that the results are still statistically significant with no statistically significant differences from the initial analysis, implying that these results are also not driven by the transportation component.

6 Discussion

Recent extensions to the traditional economic model of fertility account for the uncertainty in fertility that many households in low and middle income countries face due to lack of access to modern contraceptives. These models show that, when fertility is uncertain, human capital outcomes for children are theoretically worse than in a counterfactual world where fertility is certain. These models imply that decreases in the price of contraceptives or an increase in their effectiveness should improve human capital attainment among children. From this theoretical finding, we hypothesize that family planning programs may improve health and educational outcomes among children.

To test this hypothesis, we use data from a randomized experiment in Malawi that gave pregnant and postpartum woman access to a comprehensive family planning package that improved contraceptive knowledge and decreased their effective price. Using a subset of children that made mothers eligible to be recruited into the study, our ITT estimates show that

family planning services have positive effects on children's heights, a common proxy for nutritional status, and on cognitive ability that we consider to be quite large for an intervention that does not explicitly target children. Using a causal mediation, we are able to rule out a biological pathway, namely birth spacing and maternal depletion, as an explanation for our results. Instead, we show that are results are partially driven by increases in healthcare usage among households in the treatment arm. Combining this result with a robustness check showing that the mediation results are robust to the exclusion of households that utilized free transport offered by the trial, we see that our results support the theoretical conclusions of Cavalcanti et al. (2021).

Our results contribute to a large literature on Becker's model 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 the model we use and uses it to discuss educational investment. In their analysis, the authors go as far as to say family planning services may induce larger improvements in educational attainment than most educational programs. We will not be as inflammatory as to say family planning services may out compete most health programs, but our results do show that they may be an overlooked intervention in child-health programming. Additionally, our results contribute to a smaller, more specific, literature on the link between family planning and child health by providing what is to the best of our knowledge the first randomized evidence of this relationship.

We believe our analysis has many strengths but one significant shortfall is that we are unable to test the direct expenditure pathways presented in our theoretical framework. This is because our study is under-powered to detect treatment effects in high variance variables such as expenditures. Further, we lack the data on non-durable and food consumption that we might use as a proxy. Additionally, the mediation effects we find from health care usage are promising but modest. We believe this is because of the coarse nature of our health usage measurement and with more fine data on usage such as vaccination reports or procedure exposure, we may find that human capital investment mediates a larger portion of our results.

Finally, in the absence of our results, the case for expanding family planning services and improving access to contraceptive is strong. Our results add to that case by showing their are positive externalities associated with contraceptive use that have likely not been incorporated into cost-benefit analyses of these family planning services. Our results also

validate the inclusion of family planning services in frameworks to reduce childhood stunting like the one presented in Black et al. (2013) and when our combined with studies that link family planning services to child mortality such as Joshi and Schultz (2013), the story emerges that family planning services are a major welfare improving intervention for children.

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. 17

Alderman, H., J. Hoddinott, and B. Kinsey (2006, July). Long term consequences of early childhood malnutrition. *Oxford Economic Papers 58*(3), 450–474. Publisher: Oxford Academic. 2

Alderman, H. and E. M. King (1998, December). Gender differences in parental investment in education. *Structural Change and Economic Dynamics 9*(4), 453–468. 32

Angrist, J., V. Lavy, and A. Schlosser (2010, October). Multiple Experiments for the Causal Link between the Quantity and Quality of Children. *Journal of Labor Economics 28*(4), 773–824. Publisher: The University of Chicago Press. 2

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. 32

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. 17

Becker, G. (1960). An Economic Analysis of Fertility. NBER Chapters, National Bureau of Economic Research, Inc. 1, 3

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. 3

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. Publisher: The University of Chicago Press. 3

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. 4

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, 29

Bonferroni, C. E. (1936). *Teoria statistica delle classi e calcolo delle probabilità*. Libreria internazionale Seeber. 12

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

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, 3, 4, 24, 28, 32, 37, 38

Cleland, J., A. Conde-Agudelo, H. Peterson, J. Ross, and A. Tsui (2012, July). Contraception and health. *The Lancet 380*(9837), 149–156. 4, 23

Cordero, M. E., E. D'Acuña, S. Benveniste, R. Prado, J. A. Nuñez, and M. Colombo (1993, November). Dendritic development in neocortex of infants with early postnatal life undernutrition. *Pediatric Neurology 9*(6), 457–464. 2

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. 15

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. 2, 13, 20

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. 5

Fink, G., M. Pierre-Louis, and D. C. McCoy (2018, January). CREDI User Guide. pp. 16. 24

Furnham, A. and H. C. Boo (2011, February). A literature review of the anchoring effect. *The Journal of Socio-Economics 40*(1), 35–42. 35

Government of Malawi (2019). Malawi 2019 Floods Post Disaster Needs Assessment. Technical report. 2

Hochberg, Y. (1988, December). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika 75*(4), 800–802. Publisher: Oxford Academic. 12

Huttly, S., C. Victora, F. Barros, and J. P. Vaughan (1992). Birth Spacing and Child Health in Urban Brazil. *Pediatrics 89*(6), 1049–1053. 5

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. 19

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. 17, 19

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. 2, 5, 29

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. 5

Karra, M., D. Maggio, M. Guo, B. Ngwire, and D. Canning (2020). The Causal Effect of Improved Access to Postpartum Family Planning: Evidence from a Single-Blind, Randomized Controlled Trial in Urban Malawi. *Unpublished*. 5, 13, 32, 34

Maluccio, J. A., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and A. D. Stein (2009, April). The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults. *The Economic Journal 119*(537), 734–763. Publisher: Oxford Academic. 2

Pearl, J. (2014, December). Interpretation and identification of causal mediation. *Psychological Methods 19*(4), 459–481. 17

Phillips, J. F., A. A. Bawah, and F. N. Binka (2006, December). Accelerating reproductive and child health programme impact with community-based services: the Navrongo experiment in Ghana. *Bulletin of the World Health Organization 84*(12), 949–955. 2

Prado, E. L., L. M. Larson, K. Cox, K. Bettencourt, J. N. Kubes, and A. H. Shankar (2019, October). Do effects of early life interventions on linear growth correspond to effects on neurobehavioural development? A systematic review and meta-analysis. *The Lancet Global Health 7*(10), e1398–e1413. Publisher: Elsevier. 2

Reuters (2017, October). Vampire scare prompts U.N. pullout from southern Malawi. *Reuters*. 10

Roche, A. F. and J. H. Himes (1980, September). Incremental growth charts. *The American Journal of Clinical Nutrition 33*(9), 2041–2052. 16, 22

Rosenzweig, M. R. and K. I. Wolpin (1980). Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica 48*(1), 227–240. tex.ids= rosenzweig_testing_1980-2 publisher: [Wiley, Econometric Society]. 2

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. 5

Shekar, M., J. Condo, M. A. Pate, and S. Nishtar (2021, March). Maternal and child undernutrition: progress hinges on supporting women and more implementation research. *The Lancet 0*(0). Publisher: Elsevier. 2

WHO (2005). Report of a WHO Technical Consultation of Birth Spacing. Technical report, World Health Organization, Geneva, Switzerland. 5

WHO and PAHO (2003). Guiding Principles for Complementary Feeding of the Breastfed Child. Technical report. 16

WHO, UNICEF, and World Bank Group (2020, March). UNICEF/WHO/The World Bank Group joint child malnutrition estimates: levels and trends in child malnutrition: key findings of the 2020 edition. Technical report. 2

Wilson, T. D., C. E. Houston, K. M. Etling, and N. Brekke (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General 125*(4), 387–402. Place: US Publisher: American Psychological Association. 35

Appendix A: Heterogeneous Effects

While the ITT results we find are in themselves interesting and illuminating towards the theory of Cavalcanti et al. (2021), we are also interested in the effects of the intervention on different subgroups. For instance, we are interested if the increased investment we observe in child care 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.

Since the study registered both pregnant and postpartum we are also interested in possible differential effects based on the pregnancy status of the mother. Indeed in Karra et al. (2020), we find that increases in contraceptive use and improvements in birth space were much larger for women who were postpartum at baseline than those who were pregnant at baseline. Combining this result with the theory used throughout this paper, we may expect similar heterogeneous effects to be observed in the children of these mothers.

Specification

In order to explore the heterogeneous effects of the intervention we alter our ITT Specification to take the form:

$$Y_i = \alpha + \beta_T T_i + \beta_H H_i + \beta_{Het}(T_i \times H_i) + \delta MOB_i + \gamma f(age_i) + \lambda X_i + \zeta Z_i + \chi_i + \varepsilon_i \quad (8)$$

Here, the notation follows the same form presented with the ITT specification in Section 4. However, in this specification, we include H_i which represents a dummy variable representing the relevant subgroup. We present the results for two different values of H_i. In the first, H_i takes the value 1 if the child is male. In the second, H_i takes the value 1 if the child's mother was pregnant at baseline. Under this specification, the treatment effects for female children and those born to postpartum mothers is represented by β_T in their respective specification, while the treatment effects among males and those born to postpartum mothers is represented by $(\beta_T + \beta_{het})$.

Results: Height-for-Age Z-Score

In Table A1 we present the results for equation (8) applied to Height-for-Age Z-Scores. Here, each column represents two specifications, one where H_i represents the gender subgroups and one where H_i represents the pregnancy status subgroups. From this table, we do not detect

any differential effects across subgroups in terms of children's heights. In the unadjusted estimates, there is some suggestive evidence that the treatment effects are driven by male children, but that effect disappears when we control maternal and child characteristics. While these results may suggest that there are not differential effects among the subgroups, we should be quick to note that the main effects themselves are only marginally significant because of our high rate of attrition. Thus, it is entirely possible that there exist differences in the treatment effects that are smaller than our study is powered to detect. Never the less, the absence of large differences in treatment effects suggests that there are not large gender disparities in parental investments.

Table A1: Heterogeneous Treatment Effect of the Intervention on Growth Patterns, Year 1

	Index Children		Under 24 months at Basline	
	(1)	(2)	(3)	(4)
Panel A: Height-for-Age Z-Score				
Treatment X (Gender = Female)	0.123	0.227	0.144	0.225
	(0.169)	(0.168)	(0.166)	(0.166)
Treatment X (Gender = Male)	0.320*	0.269	0.310*	0.263
	(0.178)	(0.185)	(0.176)	(0.183)
Treatment X (Baseline Status = Postpartum)	0.212	0.255	0.207	0.248
	(0.156)	(0.160)	(0.155)	(0.158)
Treatment X (Baseline Status = Pregnancy)	0.250	0.244	0.270	0.194
	(0.200)	(0.198)	(0.196)	(0.194)
Age and Birth Month Controls	X	X	X	X
Mother and Child Characteristics		X		X
Observations	720	720	738	738

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

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. Each column represents two separate specifications. One for the gender subgroup disaggregation and one for the pregnancy status disaggregation. 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), and tribal ethnicity (Chewa versus other) and child-level controls such as biological sex, and birth order (included as dummy variables to relax the linear restriction). Sample Cluster fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

Results: CREDI Scores

In Table A2 we present symmetric results to Table A1 but for CREDI Z-Scores. Here we find very different results than for Height-for-Age Z-Scores. In particular, we find that the treatment effects on CREDI Z-scores are driven by effects on girls and children born

to mothers postpartum at baseline. Further, we do find significant differences within the subgroups. The results within the gender subgroup, when contrasted with the results from Table A1, suggest that the induced increased investment in children goes towards girls more than boys but may take time to be realized (since the results for height-for-age are from year 1 and the results for CREDI are from year 2), or may go towards inputs that effect cognition but not growth status, such as play time, or other cognitive stimulants.

Table A2: Heterogeneous Treatment Effect of the Intervention on CREDI, Year 2

	Index Children		Under 24 months at Basline	
	(1)	(2)	(3)	(4)
Panel A: CREDI Z Scores				
Treatment X (Gender = Female)	0.162**	0.145*	0.170**	0.152**
	(0.0743)	(0.0761)	(0.0743)	(0.0761)
Treatment X (Gender = Male)	0.036	0.045	0.034	0.039
	(0.069)	(0.071)	(0.070)	(0.071)
Treatment X (Baseline Status = Postpartum)	0.206***	0.229***	0.208***	0.230***
	(0.0664)	(0.0692)	(0.0663)	(0.0686)
Treatment X (Baseline Status = Pregnancy)	0.001	-0.041	0.005	-0.038
	(0.075)	(0.079)	(0.077)	(0.080)
Age Controls	X	X	X	X
Mother and Child Characteristics		X		X
Observations	727	727	739	739

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

Notes: For all columns, the unit of observation is a child. The results presented are from OLS models with standard errors in parenthesis. Each column represents two separate specifications. One for the gender subgroup disaggregation and one for the pregnancy status disaggregation. 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), and tribal ethnicity (Chewa versus other) and child-level controls such as biological sex, and birth order (included as dummy variables to relax the linear restriction). Sample Cluster fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

Meanwhile, we find that increases in cognition are almost entirely centered on children born to mothers who were postpartum at baseline, which matches our result from Karra et al. (2020). One possible explanation for this differential effect is that these children are inherently older than those born to mothers who were pregnant at baseline and therefore may have had more time to develop a greater variance in cognitive abilities, allowing the increased investments to shine through. However, another explanation that may explain this combination of results is that family planning counseling during pregnancy is not as effective as counseling during the amenorrheic period. This could be because when they are

immediately postpartum, women are facing the physical, emotional, and financial costs of childbirth and are thus anchored to their preference to prevent subsequent births (Furnham and Boo, 2011; Wilson et al., 1996).

Appendix B: Intervention Component Analysis

The combined nature of our intervention naturally begs the question "well, which part is most effective?". However, as we partially discussed in Section 4, there are a number of causal identification challenges that come with this question. First, there are numerous latent characteristics that likely determine women's use of each intervention component. While we are able to navigate this challenge using exogenous variation from treatment assignment, in an ideal world, we would have multiple sources of said variation. This second complication is driven by the combined nature of our intervention. As women assigned to the treatment arm received the full combined intervention, there is no way to compare women who *only* got counselling to women who *only* used transport. In this appendix section, we will present instrumental variable estimates for the counseling and transport component that in a single component intervention would identify the LATE of treatment assignment. However, as we are looking at individual components of a multiple component intervention this estimation is under-identified and we must back off of causal claims, treating this evidence as suggestive.

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:

$$Y_i = \alpha + \beta_C C_i + \delta MOB_i + \gamma f(age_i) + \lambda X_i + \zeta Z_i + \chi_i + \varepsilon_i \tag{9}$$

Where the notation once again follows that presented in Equation (1). Here C_i 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. We once again present unadjusted and adjusted specifications.

Results

We present the relationship between the use of the intervention components and our outcomes of interest in Table B1. We see that on the intensive margins, each counseling visit is associated with a 0.045-0.048 SD increase in height-for-age and a 0.009-0.01 SD increase in CREDI scores, while each transport to the Kauma Clinic is associated with a 0.45-0.47 SD increase in height-for-age and a 0.09-0.1 SD increase in CREDI scores. Meanwhile on the extenisive margin, counseling is associated with a 0.24-0.26 SD increase in height for age and a 0.04-0.05 SD increase in CREDI scores and transport use is associated with a 0.81-0.85 SD increase in height for age and a 0.21-0.24 SD increase in CREDI scores.

Table B1: Intervention Component 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	0.0458*	0.0489**	0.243*	0.260**	0.445*	0.467**	0.806*	0.849**
	(0.0241)	(0.0237)	(0.128)	(0.126)	(0.240)	(0.234)	(0.430)	(0.420)
Observations	720	720	720	720	720	720	720	720
First Stage F-Statistic	3476.74	3560.59	6859.82	7234.90	76.64	77.63	129.93	130.93
Panel B: CREDI Z-Score								
Treatment Component	0.0103**	0.00929*	0.0529**	0.0482*	0.106*	0.0930*	0.240**	0.210*
	(0.00503)	(0.00496)	(0.0259)	(0.0257)	(0.0543)	(0.0514)	(0.119)	(0.112)
Observations	727	727	727	727	727	727	727	727
First Stage F-Statistic	2395.44	2479.92	5844.86	5874.31	45.96	46.11	82.94	84.63
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), and tribal ethnicity (Chewa versus other) and child-level controls such as biological sex, and birth order (included as dummy variables to relax the linear restriction). Sample Cluster fixed effects are included and heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

The most striking aspect of the results presented in Table B1 are the large difference between components, implying that the transport component was much more effective than counseling. However, it is important to note that use of the counseling component was much more wide spread. Of the 782 treatment women who were 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 of them did not use the transport provided by the MFPS. With these levels of treatment exposure in mind, these results show us that the association with FP counseling is marginal but because of it's widespread use, it is likely the component driving our main results. Meanwhile, the transport had large gains for those

women who used it, but was used by few women. Further, transport use is likely higher among the poorest women who benefit the most from increased access to transport. As such, the children born to mothers who used the transport likely would have been worse off than the average child but for 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.

Appendix C: Robustness Checks

While the results of our mediation analysis are illustrative of the possible mechanisms behind our ITT results, supporting the theory of Cavalcanti et al. (2021), there does exist one alternative explanation for the mediating effect of healthcare usage. Since one component of our intervention offered women free transport to a clinic, it is possible that this component is driving the mediation results in two ways. The first possibility is that during these trips, women were not only receiving family planning, but also receiving care for their children[18]. A second possibility is that the transport to the clinic introduced to the healthcare system women who had previously not have accessibility. This introduction may have made them more comfortable returning for follow-up visits for themselves and their children. In either case, we would indeed find that healthcare usage has a mediating effect on our ITT estimates but not in a way that supports the relevant theory.

In order to address this possible alternative explanation, we once again conduct the mediation analysis described in Section 4 on a subsection of children that we refer to as the "Non-Transport Sub-sample". These children are those who were born to women that did not take advantage of the transport component of the intervention. While mothers and children within this sub-sample likely have very different characteristics than the full sample, finding a mediating effect within this sample would imply that healthcare usage mediates our ITT results even in the absence of the transport component.

Table C1 displays this mediation analysis among the non-transport sub-sample for our results on height-for-age Z-scores. The table follows the same format as Table 6. We find that within this sub-sample, healthcare usage has a significant ACME of 0.035-0.037 SD and birth spacing once again does not have a significant mediating effect. In addition, our sensitivity analysis implies that our results are robust to the exclusion of any relevant co-

[18]The Good Health Clinic offers a panel of maternal and child health services beyond the family planning services offered.

Table C1: Mediation Analysis Among Non-Transport Women for Growth Patterns, Year 1

	Mediator: Healthcare Usage			Mediator: Birth Spacing		
	(1)	(2)	(3)	(4)	(5)	(6)
ITT on Mediator	0.097**	0.099**	0.097**	-0.021	-0.026*	-0.021
	(0.036)	(0.037)	(0.036)	(0.013)	(0.014)	(0.014)
ACME	0.037**	0.035**	0.037**	-0.008	-0.011	-0.008
	[0.004,0.087]	[0.0002,0.084]	[0.002,0.084]	[-0.048,0.02]	[-0.054,0.014]	[-0.048,0.017]
ρ where ACME =0	0.1	0.1	0.1	-	-	-
Mother and Child Controls		X	X		X	X
Household Wealth Index			X			X
Observations	616	616	616	617	617	617

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

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), and tribal ethnicity (Chewa versus other) 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. ITT estimations on the mediator are adjusted by these same controls. heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

variate that explains less than 10 percent of the combined variance between height-for-age and healthcare usage.

Table C2 mirrors Table C1 with the results for CREDI Z-Scores. Among this sub-sample, we see that healthcare usage in 2017 mediates our ITT results on cognition to the tune of 0.025-0.026 SD while birth spacing continues to not display a mediating effect. In the case of CREDI Z-scores, our sensitivity analysis implies that our results are robust to the exclusion of any relevant covariate that explains less than 20 percent of the combined variance between CREDI scores and healthcare usage.

Overall, the results for this robustness check imply that our mediation results are robust to the exclusion of mothers who utilized the transport component of our intervention. This result implies that there is a mechanism through which family planning influences healthcare usage, even in the absence of free transport, supporting the theory of Cavalcanti et al. (2021). While our results remain statistically significant with this exclusion, we should note that the point estimates are smaller than in the full analysis, now explaining only 15 percent of the effect on height-for-age, and 25 percent of the effect on CREDI scores. Acknowledging that these differences are not statistically significant, we see two possible reasons for the attenuation. The first is that there is indeed a mechanism that works through the transport and our main analysis is capturing the mix of two separate effects. The second is that, as

Table C2: Mediation Analysis Among Non-Transport Women for CREDI, Year 2

	Mediator: Healthcare Usage (2017)			Mediator: Birth Spacing		
	(1)	(2)	(3)	(4)	(5)	(6)
ITT on Mediator	0.122**	0.121**	0.123**	-0.024	-0.023	-0.024
	(0.036)	(0.036)	(0.036)	(0.02)	(0.021)	(0.02)
ACME	0.025**	0.026**	0.026**	0.002	0.002	0.002
	[0.007,0.05]	[0.006,0.053]	[0.007,0.051]	[-0.005,0.011]	[-0.005,0.012]	[-0.005,0.011]
ρ where ACME =0	0.2	0.2	0.2	-	-	-
Mother and Child Controls		X	X		X	X
Household Wealth Index			X			X
Observations	616	616	616	659	659	659

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

Notes: For all columns, the unit of observation is a child. Results presented are for a causal mediation analysis of the ITT results on CREDI 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), and tribal ethnicity (Chewa versus other) 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. ITT estimations on the mediator are adjusted by these same controls. heteroskedasticity-robust standard errors are clustered by mother to account for genetic correlation.

discussed in Appendix B, the children born to mothers who utilized transport were born into poorer households and thus would likely have been less developed but for our intervention. Such a trajectory difference would imply that these children had the greatest potential for improvement and their exclusion eliminates the largest gains made by treatment, attenuating the measured effect.