Improving the Early Childhood Environment: Multi-Generational Effects on Human Capital *

Tania Barham[1], Brachel Champion[2], Gisella Kagy[3], and Jena Hamadani[4]

[1]University of Colorado Boulder

[2]U.S. Air Force Academy

[3]Vassar College

[4]icddr,b

February 14, 2023

Abstract

This paper examines the effect of improving the early childhood health environment on human capital in adulthood and on the next generation. We exploit a quasi-randomly placed mother and child health and family planning program in Bangladesh using individual level panel data linked across three generations. Individuals who were eligible during childhood are taller with less short-stature, and men have higher grades attained. Effects are concentrated among those with the lowest health endowment at birth, indicating a reduction in inequality. In the next generation, daughters of eligible women are taller, less stunted, and have better cognition. Findings suggest ignoring intergenerational effects leads to underestimating the benefits of programs that improve early childhood health and and misses their effects on inequalities.

JEL Classification: I15, J13, O12

Keywords: human capital, intergenerational, vaccination, family planning, Bangladesh

*Barham: University of Boulder Colorado, Economics Building Rm 212, 256 UCB, Boulder, CO 80309 (email: tania.barham@colorado.edu); We could not have done this project without icddr, a special thanks to the whole team but especially Abdur Razzaque for their partnership and data access. The data collection for this project was generously funded by the National Institutes of Health, Population Research Bureau, the International Initiative for Impact Evaluation, and CU Population Center.

1 Introduction

There is mounting evidence that strong correlations exist for human capital and economic measures between parents and their children.[1] Determining the causes of these correlations is important for understanding the role of public policy in addressing the persistent disadvantage across generations. Theoretical models depict how positive investments, or disadvantage, in-utero and early childhood can have lasting effects on human capital formation in adulthood and for future generations, due to behavioral and biological mechanisms.[2] To date, empirical research on the long-term and intergenerational effects of the early childhood environment has primarily focused on early-life insults.[3] However, research on interventions designed to improve the early childhood environment is necessary to understand the role for policy. While positive interventions, such as preventative health, are common worldwide, few interventions have both the required data and program design to allow for rigorous long-term and intergenerational assessment.

We bring new individual level panel data that spans three generations to provide evidence that preventive health interventions have important long-term and intergenerational effects on several measures of human capital: height, stunting, cognitive functioning and grades attained. To identify the effects, we take advantage of a quasi-randomly placed Maternal and Child Health and Family Planning Program (MCH-FP) in rural Bangladesh. The interventions provided are typical of preventive health interventions world-wide, with the provision of vaccinations against measles, DPT, polio, and tetanus, and modern contraception among other components. Interventions were phased-in starting in 1977 with a focus on family planning and expanded in 1982 to include intensive child health interventions (Bhatia et al., 1980). Treatment and comparison areas were built into the program design and were also chosen ex-ante to be economically and socially similar. Similar childhood interventions became available in health clinics in the comparison area in 1989 providing an approximately 10-year evaluation period.

Previous literature documents that the MCH-FP program had important effects on human capital attainment in early adolescence. Children born during the roll-out of the child health intervention had

1. For examples on income and education see Black and Devereux (2011), Hertz et al. (2008), Richter and Robling (2013) for health Ahlburg (1998), Bhalotra and Rawlings (2013), and Venkataramani (2011), and for cognition Grönqvist, Öckert and Vlachos (2017).
2. For theoretical models of early-life skill formation in economics see Cunha and Heckman (2007); Heckman (2007*a*); Attanasio (2015), and for models of developmental origins of health and disease in the biomedical literature see (Gluckman, Buklijas and Hanson, 2016).
3. See Almond and Mazumder (2011), Almond, Currie and Duque (2018), and Currie and Vogl (2013) for reviews of the literature and East et al. (2019) for further information.

improved human capital as measured by height (0.22 SD), cognitive functioning (0.39 SD), and years of education (0.17 SD) at ages 8-14 (Barham, 2012). The increase in schooling for these children was concentrated in males (Joshi and Schultz, 2013).[4] However, there was no improvement in these same measures for children born prior to the intensive child health interventions (Barham, 2012).

In this paper, we examine the long-term and intergenerational impacts of the MCH-FP program using two cohorts which we refer to as the first and second generation. For the first generation, we focus on the cohort eligible for the intensive early child health interventions at birth because adolescent human capital improvements from the MCH-FP program were concentrated in this cohort. This cohort is approximately 24–30 years-old when their long-term human capital outcomes are measured. The second-generation sample is composed of first-born children of females in the first-generation cohort. These children are aged 0-14 when their human capital outcomes are measured.

We take advantage of the well-designed treatment and comparison areas to estimate intent-to-treat (ITT) effects on human capital measures including height, grades attained and cognition. For the first generation, we estimate double-difference models that use individuals born between 1947-1969 as a baseline cohort as the human capital measures examined in this paper are fairly stable for this cohort after the introduction of the MCH-FP program. For the second generation, we use single-difference models as our primary specification because it is not possible to identify a similarly aged cohort that is truly unaffected by the program.[5] We demonstrate that the comparison area provides a good counterfactual: there is pre-program balance in human capital, employment, migration trends, fertility and individual and household characteristics. We also find no evidence of spatially correlated errors or have knowledge of a disease outbreak that affected only the treatment or comparison area from the decades of demographic surveillance data on mortality and disease (Fauveau 1994). Further, we demonstrate results are unlikely to be biased due to attrition for either generation or a result of fertility selection for the second generation.

Results for the first-generation demonstrate human capital gains documented in childhood (Barham, 2012; Joshi and Schultz, 2013) persist into adulthood for height and grades attained but not cognition. Both males and females are almost one centimeter taller, leading to a reduction in short stature of almost 50 percent. In addition, grades attained increased by almost one year for men, but there is no effect on

4. The program also reduced family size by almost one child (Phillips et al., 1984; Joshi and Schultz, 2013) though the reduction in family size depends on length of exposure of the mother to the program (Barham et al., 2021).

5. As robustness, we show results for the second generation are similar for double-difference models where the baseline cohort is similarly aged, any birth order children, of older mothers whose human capital was less affected by the MCH-FP program.

education for first generation females, which is not surprising since a nationwide girl's secondary school scholarship existed in both the treatment and comparison areas.

Intergenerational effects are concentrated in female children. Second generation females in the treatment area, compared to the comparison areas, were taller (1.6 cm or 0.33 SD), experienced a 50 percent reduction in stunting, and cognitive functioning was higher (0.26 SD) among 7-14 year olds.

We explore effects on inequality by examining the heterogeneity of the impacts by an individual's health endowment at birth. We proxy the health endowment with the height of the mothers of the first generation cohort (the grandmothers of the second generation cohort). Results highlight that the program effects for the first generation are largest for those born of the shortest mothers where gains from health improvements were potentially the largest, indicating a reduction in inequality. For the second generation, results are strongest for females whose matrilineal grandmothers were in the top tercile for height in the sample.

We examine a variety of mechanisms to explore the potential underpinnings of the patterns of results by gender for the second generation. Females did not benefit more than males because they were worse off to start. We also find few program effects on behavioral mechanisms that could affect a child's human capital, including pre-natal care, delivery by trained personnel, post-natal inputs such as early childhood vaccination, or mother's agency as measured through level of education or an index of empowerment. None of these behavioral mechanisms differ by gender of the second generation individual. While examination of all possible behavioral mechanisms is not possible, and it is difficult to disentangle behavioral from biological mechanisms, we find no evidence of behavioral mechanisms that could explain the intergenerational results and the pattern by sex, indicating a possible role for biological mechanisms.

The existing body of literature on the long-term effects of positive early childhood interventions is growing and generally positive.[6] However, data constraints and design limitations have curbed the number of empirical studies on the intergenerational effects of positive early childhood health interventions on human capital. Three notable papers that do explore intergenerational linkages are Behrman et al. (2009), East et al. (2019), and Barr and Gibbs (2017). Behrman et al. (2009) examines intergenerational effects of

6. Evidence on the long-term effects of early child health on adult human capital outcomes in the U.S. focuses on the expansions of large federally funded programs such as Medicaid (Miller and Wherry, 2019), food stamps (Hoynes, Schanzenbach and Almond, 2016) and show improved educational attainment and adult health including reductions in chronic disease and the incidence of metabolic syndrome. However, effects on cognition are sparse and evidence from Head Start showed early program impacts on cognition fade by adolescence (Deming, 2009). In developing countries, (Maluccio et al., 2009; Walker et al., 2021) show an early childhood nutrition intervention in rural Guatemala lead to higher educational attainment, reading comprehension and cognition in adulthood, though attrition was high (40 percent).

an early childhood nutrition intervention in rural Guatemala[7] and find male children of eligible mothers were taller and had greater weight-for-age. Research on two large federally funded programs in the US document intergenerational reductions in low birth weight from Medicaid (East et al., 2019) and intergenerational improvements in higher educational attainment from Head Start (Barr and Gibbs, 2017).[8] While these works begin to frame our overall understanding of causal intergenerational linkages, they still leave open questions about intergenerational effects in lower income country contexts where individuals may face many competing health risks and concern over effects fading out is pertinent. Similarly, it is unknown if intergenerational health effects persist past birth, or if the effects differ by human capital endowment.

This paper answers several of these open questions in the literature. First, we estimate the long-term and intergenerational effects of common preventive health interventions in a low income country context, Bangladesh, where the relationship between parent and child outcomes may be different than the US context. Second, we examine a wider set of human capital indicators on children past birth and empirically explore potential behavioral mechanisms. Examining effects on different measures of human capital is informative as the sensitive periods for development, and effects over the life course, may differ by outcomes. Third, we document the heterogeneity of effects by the pre-program health endowment to examine effects on inequality across generations and if there are non-linearities in health production. Finally, outcomes are drawn from a survey with an attrition rate of less than 10 percent, which is very low, minimizing attrition selection which plagues long-term and intergenerational research.

The remainder of the paper is structured as follows. Section 2 presents information on the program and a conceptual framework to motivate possible mechanisms. Section 3 describes the data and Section 4 the empirical strategy. Section 5 and 6 presents our results. Section 7 discusses heterogeneity by initial health endowment, section 8 discusses potential mechanisms and section 9 discusses our interpretation of the results.

7. The program provided a protein enriched beverage in two randomly chosen villages out of four with the other villages receiving a sugary beverage. 60 percent of children eligible for the intervention before age three were followed into adulthood.
8. Also in the US context, Garcia, Heckman and Ronda (2021) find intergenerational effects on educational attainment which are concentrated in males from an early child education intervention for disadvantaged youth (The Perry Preschool Project).

2 MCH-FP Program and Conceptual Framework

2.1 The Maternal Child Health and Family Planning Program

The MCH-FP program was initiated in October 1977 in the Matab district of Bangladesh. The program was administered by icddr,b, (formally known as International Center for Diarrheal Disease Research, Bangladesh) and was a demonstration project for the national family planning program. To aid evaluation, MCH-FP was administered as an experiment and treatment and comparison areas were built into the design of the program. The program covered about 200,000 people in 149 villages, with the population split fairly evenly between the two areas. Individuals living in the treatment areas were eligible for health and family planning interventions provided by the program while those living in the comparison areas were not (Fauveau, 1994). A key feature of the program is that interventions were free and administered in the beneficiary's home during monthly visits made by local female health workers hired and trained by the program (Bhatia et al., 1980).

During the first phase spanning October 1977 through April 1982, program interventions focused on family planning and maternal health through the provision of modern contraception, tetanus toxoid vaccinations for pregnant women (starting June 1978), and iron and folic acid tablets for women in the last trimester of pregnancy (Bhatia et al., 1980). Tetanus toxoid was expanded to all women of reproductive age in 1982. During these visits, the female health workers also provided counseling about contraceptives, nutrition, hygiene, and breastfeeding, and instructions on how to prepare oral rehydration solution. These services were supported by well-developed follow-up and referral systems to ensure management of side effects and continued use of contraceptives (Phillips et al., 1984).

Starting in March 1982, child health interventions were intensified for children under the age of five. The measles vaccine was introduced in half the treatment area and expanded to the other half in November 1985 (Koenig et al., 1990). Additional child health interventions were phased in between 1986 and 1988. In January 1986, DPT, polio, and tuberculosis immunizations were added, and later in that year vitamin A supplementation. Curative care, such as nutrition rehabilitation or those who were nutritionally at risk was introduced in the late 1980s. In the comparison area, then-standard government health and family planning services were available. Services included modern family planning, but these services were only available at clinics, not in the home, and childhood vaccinations were not readily available until 1989 or later, providing an approximately 10-year experimental period, 1977–1988, to evaluate the

program. Differences between the treatment area and the rest of the country, including the comparison area, narrowed after 1988 as the lessons of the Matlab success were incorporated into the national plan (Cleland et al. 1994). The number of governmental community health workers delivering in-home services increased throughout the country over time reducing the client-worker ratio from 1 per 8000 in 1987/88 to 1 per 5,000 in 1989/90 in the comparison (Cleland et al. 2004). The ratio was still lower in the treatment area at 1 per 1,300 in 1990.

Program implementation followed the planned timeline and uptake was rapid. For example, Figure 3 indicates that the contraceptive prevalence rate (CPR) for married females 15-49 was similarly low prior to the program (<6%) in both areas. There was a large increase in the CPR to 30% in the treatment area during the first year of the project and it increased steadily reaching almost 50% by 1988. Due to availability of contraceptives from government services, the CPR did increase in the comparison area over time (though not as quickly) and rates remained below 20% in 1988.

The measles vaccination rate also rose rapidly to 60% in 1982 after it was introduced in half of the treatment area and in 1985 when it was introduced in the other half (Figure 3). By 1988, coverage rates for children aged 12–23 months living in the treatment area were 93 percent for the BCG vaccine against tuberculosis, 83 percent for all three doses of DPT and polio, 88 percent for measles, and 77 percent for all three major immunizations (icddr,b 2007). Data on vaccination rates in the comparison area do not exist, but are believed to be near zero. Government services did not regularly provide measles vaccination for children until around 1989, so the comparison area was viewed as a largely unvaccinated population (Koenig, Fauveau and Wojtyniak, 1991).[9]

2.2 The Importance of MCH-FP Interventions for Human Capital Formation

In low-income countries, children face a myriad of interrelated obstacles such as infectious diseases, malnutrition and high levels of fertility in reaching their full human capital potential. At the time of the MCH-FP rollout, Matlab was a poor rural area characterized by high rates of disease, stunting and fertility. It is estimated that in the late 1970's rates of stunting in Bangladesh were as high as 71% (Bangladesh Bureau of Statistics 1985). The MCH-FP program provided a package of interventions, including vaccination, family planning, and maternal health, that are known to improve the early childhood health environment.

9. Nationally, measles vaccination for children under the age of five was less than 2% in 1986 (Khan and Yoder, 1998), and was below 40% in the comparison area in 1990 (Fauveau, 1994).

Dissemination of early childhood vaccines marked a turning point in public health in the twentieth century. Prior to the widespread accessibility of vaccines, rates of morbidity and mortality were high from contagious diseases such as measles.[10] Measles, in particular, is known to lead to severe morbidity because it affects the child's nutritional status through secondary complications such as pneumonia and diarrhea (Reddy, 1987). Undernutrition weaken the immune system leaving the child more susceptible to and slower to recover from disease. In addition, the measles virus is damaging because of the longevity of the impact of the virus.[11] It is now understood that the measles virus causes immunosuppression for up to five years, and effectively resets previously acquired immunity (Gadroen et al., 2018). This "immune amnesia," coined by Mina et al. (2015), creates a highly compromised immune system during childhood and is associated with increased non-measles related mortality for a period of two years after a measles infection, and increased non-measles related morbidity for up to 5 years after a measles infection (Gadroen et al., 2018; Mina et al., 2015).

While children's physical and developmental growth may catch up once the illness has passed, in high-disease environments children may experience a number of episodes of illness in combination or close succession, reducing the time for catch-up growth. Both non-randomized and randomized studies show that undernutrition—especially before the age of three—affects the growth and cognition of young children (Grantham-McGregor, Fernald and Sethuraman, 1999*a*,*b*; Walker et al., 2007). In addition, morbidity from vaccine-preventable diseases causes general malaise, and apathetic children typically receive less stimulation from adults and learning opportunities, which also hinders human capital development (Walker et al., 2007). The negative impact of disease on human capital development may be higher in lower-income countries because they have higher levels of undernutrition.

The non-child health components of the MCH-FP program may also have an indirect effect on human capital. Reduction in family size can lead to a quantity-quality trade-off, with low-fertility parents investing more in their children both in terms of time or access to resources such as better nutrition or more schooling. In addition, longer birth spacing resulting from the family planning interventions and the maternal health inputs (e.g., iron and folate supplementation, nutritional counseling, referral for pregnancy complications) may also directly affect human capital of a child through the improved

10. For example, prior to the measles vaccine it is estimated that by age sixteen 95% of a population would have contracted measles at some point in their lives (Miller, 1964; Perry and Halsey, 2004)

11. A key characteristic of measles is that the virus causes immune suppression, and this had previously been thought to last a few weeks to several months after infection.

nutrition and health of the mother while the child is in utero (Walker et al., 2007; Almond and Mazumder, 2011).

2.3 Conceptual Framework

To understand the role of the improved childhood health environment resulting from the MCH-FP program in the process of human development, we specify a model to illustrate the channels through which the program could affects human capital in the first and second generations. We follow Heckman (2007*b*) and Attanasio, Meghir and Nix (2020) and model the dynamic evolution of human capital, but restrict the model to a three-stage framework, encompassing early childhood ($t = 1$), late childhood ($t = 2$), and adulthood ($t = A$). We include j types of human capital in the model since our analysis focuses on multiple dimensions of human capital. The production function for adult human capital of type j, θ_{jA}, can be written as:

$$\theta_{jA} = f_j(\theta_{j1}, \theta_{-j1}, I_1, I_2, Z_1, \theta_{jp}, \theta_{-jp}) \tag{1}$$

where θ_{j1} are initial conditions which we simplify to be a child's endowment of type j of human capital, θ_{-j1} a child's endowment of all other measures of human capital except j, I represents either public or private investments made during early childhood ($t = 1$) or late childhood ($t = 2$), Z is a vector of background characteristics (such as gender and religion), θ_{jp} is parental human capital of type j and θ_{-jp} is parental endowment of all other measures of human capital except j. Given conditions outlined in Heckman (2007*b*), this framework allows for a dynamic human capital production function that embodies qualities of self-reinforcement, dynamic complementary, and sensitive periods of development.

Self-reinforcement means that for a given level of investment, higher levels of human capital in one period create higher levels of human capital in the next period, within and across human capital types, i.e., $\partial f_j(\cdot)/\partial\theta_{jt} > 0$. Dynamic complementarity arises when investments are more productive because the previous period's stock of human capital is higher i.e., $\partial^2 f_j(\cdot)/\partial\theta_{j1}\partial I_2 > 0$. Finally, a period is defined as sensitive if a given level of investment has a higher return in this period than any other period. The relevance of these three concepts in any period could vary by background characteristics such as gender and human capital type.

For the first generation, the MCH-FP program represents an increase in public investment, I_1, in early

childhood, the first period of the model. Early childhood is viewed as a sensitive period for human capital measures such as height and cognition. For the same sample considered in this paper, Barham (2012) shows, that the first period investment was successful at increasing human capital in the second period, late childhood, for three measures of human capital, height, cognition, and grades attained. However, whether these impacts on human capital caused by investments made in the first period persist into adulthood may depend on dynamic complementarity, and may require further second period investment. The persistence of impacts in adulthood may vary by the type of human capital, and may depend on whether period 1 is the only sensitive period for that type of human capital. Investments made in period 1 may also affect adult human capital through epigenetics as the biomedical literature shows one's environment in period 1 can alter gene expression in adulthood.[12]

For the second generation, the framework provides several pathways for intergenerational transmission of MCH-FP benefits. First, parental human capital (θ_{jp} and θ_{-jp}) is altered because of the MCH-FP program. Parents from the first generation may pass on height on health-determining genes to their children (the second generation), including epigenetic marks. This represents a biological pathway through which higher human capital of the first generation may impact the second generation.

Second, there may be a correlation between parental human capital and the level, and productivity, of investments made in early and late childhood (I_1 and I_2). Higher human capital parents may be healthier and able to spend more time actively engaging with their children, thus providing a more stimulating environment. Tastes and preferences for health and health-promoting behaviors may also be altered given higher parental human capital. In addition, higher parental human capital could lead to income or financial support for investments in children. These parental investments may vary by the sex of the child and the effect of the program on the agency of women, since women are primarily responsible for child care in this setting.

This framework conceptualizes how the human capital of three generations is related. For the first generation, their parent's human capital (θ_{jp} and θ_{-jp}) is a direct input into the their human capital production function. Effects may be heterogeneous depending on how the parental health endowment interacts with the increased public investment, I_1. For the second generation, individuals are impacted by their grandparent's human capital through the impact on the first generation (the parents) level of human capital. This model illustrates how pre-program levels of human capital may impact both the first and

12. See Gluckman, Hanson and Beedle (2007) and Gluckman, Buklijas and Hanson (2016) for a survey of this literature.

second generation.

3 Sample, Data, and Treatment Status

3.1 Samples

This paper analyzes how MCH-FP impacted the early childhood health environment of two samples—the first and second generation—to determine the long-term and intergenerational effects, respectively. The first generation sample includes those born during the experimental phase of MCH-FP when both family planning and child health interventions were provided in the treatment area. This cohort was born between 1982–1988 and are ages 24–30 during MHSS2. We include individuals born between 1947–1969 (ages 44–65 in 2012) as a baseline cohort to control for any differences between treatment and comparison areas prior to the program in the double difference model when examining height, education, and cognition measures. The 1947–1969 cohort was too old when the MCH-FP interventions were implemented for the program to affect their height, education or cognition directly and not likely to be a sibling of those impacted directly, thus minimizing sibling competition effects.

The second generation sample is comprised of firstborn children of females from the 1982–1988 first generation cohort. The second generation sample were born between 2002–2015 and are aged 0-14 in MHSS2. We identify children based on pregnancy histories collected in MHSS2 which includes a listing of all live births, still born children, or lost pregnancies.[13] We follow the children of the females in the first generation cohort because they are the cohort that experienced human capital gains from the MCH-FP program as children (Barham, 2012) and analysis from this paper finds gains persisted into adulthood. These females are also typically married to older men who were born prior to the roll out of the intensive child health interventions in 1982. The MCH-FP program did not affect the human capital of people born prior to 1981 on average (Barham, 2012), so the age difference between spouses helps isolate from which spouse any intergenerational effects on humcan capital from the program are likely being passed. We focus specifically on firstborns because first generation females are early in their fertility. While 83 percent of the sample have a first-born child, only 47 percent have a second born child by the time of the MHSS2 survey. Who has a second born child is likely to be selected so we do not include this group of children.

13. While the pregnancy history provides the main way to identify children, we check for any children missing from the listing, perhaps because they died when young, using the DSS data and the MHSS2 siblings module.

While we would like to examine effects on children born to fathers in the 1982–1988 cohort, only 34 percent have a first child because men have children later than females in this context. In addition, the majority of these men married females who were born after 1988 when the interventions became increasingly available in the comparison area, making it difficult to isolate the treatment effects as spouses of these men in both the treatment and comparison area are likely treated. While it is possible that we miss important intergenerational effects by not examining first generation fathers, previous research shows intergenerational correlations are stronger between mother and child than father and child (Thompson, 2014; Bhalotra and Rawlings, 2013).

3.2 Data Sources

This paper draws on panel data from the Matlab district. Unique identification numbers allow the merging of individuals throughout time from four main data sources: the 2012–2015 Matlab Health and Socioeconomic Survey wave 2 (MHSS2), the 1996 Matlab Health and Socioeconomic Survey wave 1 (MHSS1) (Rahman et al., 1999), periodic censuses of the study area conducted by icddr,b in 1974 and 1982 (icddrb, 1974, 1982), and the 1974–2014 Matlab demographic surveillance site (DSS) data on vital events (e.g., births, marriages, deaths, in- and out-migrations) collected by icddr,b.[14] MHSS1 and MHSS2 are random samples of the study area, while the periodic censuses and DSS data cover the entire study area. The census and DSS data are known for their high quality as they were collected bi-weekly or monthly, and allow determination of exact birth dates, treatment status, migration status, and testing of pre-program balance. These data sets are further linked, usually at the village level, to data on potentially confounding programs such as access to microcredit, primary schools, health facilities, flood mitigation, and arsenic exposure. More details on data construction are in Appendix B and potential confounders in Appendix C.

Outcomes are drawn from MHSS2 which is a large socio-economic survey designed to be a panel follow-up of all individuals in the MHSS1 primary sample and their descendants. The MHSS1 primary sample is representative of the study area's 1996 population, but does not include individuals who migrated between the start of MCH-FP and MHSS1. To address this bias, MHSS2 also includes individuals born to an MHSS1 household member between 1972 and 1989 who had migrated out of Matlab between 1977

14. MHSS2 data linked with the baseline data will be publicly available in the future. It was collected by the authors together with a team of researchers from the University of Colorado Boulder, Brown University, and icddr,b. Specific icddr,b census and DSS data must be requested from the organization. More information can be found at http://www.icddrb.org/component/content/article/10003-datapolicies/1893-data-policies

and 1996, which we refer to as pre-1996 migrants.

MHSS2 was collected between 2012 and 2015 throughout Bangladesh and included interviews of international and difficult-to-track migrants when they returned to the study area to visit family, particularly during Eid celebrations. A phone survey was also administered to international migrants who did not return to Bangladesh during data collection. Phone survey respondents are all men and represent fifteen percent of men in our sample. Non-phone survey respondents were interviewed and tested in their homes, so the survey not suffer from the same type of attrition that occurs when test measures are collected in institutional setting such as schools. Individual test and health measurements were collected by enumerators who were rigorously trained by psychologists and public health professionals working for icddr,b. The phone survey was significantly shorter than the main survey, but includes self-reported measures of height, weight, and education. However, other health and cognitive tests were not included in the phone survey so the sample size for men is smaller for cognition than height.

Attrition rates for MHSS2 are less than 10 percent of the target sample which is extremely low for a 35-year follow-up survey. Attrition, including death, for the first generation sample is 10 percent for men and 7 percent for women (Table B2). For the second generation sample attrition is less than 7 percent, with attrition from mortality accounting for about 4 percent (Table B3). Attrition is balanced between the treatment and control groups for all age and sex cohorts.

3.3 Outcomes

This paper focuses on four main measures of human capital: height, short stature or stunting, cognition index, and grades attained. These outcomes are referred to as stock measures of human capital because they tend to be stable by adulthood since early childhood is a sensitive period of physical growth and brain development. As a result, height and cognition are known to reflect changes in early childhood health and nutrition. We also present program effects on a measure of health, metabolic syndrome, which is a cluster of conditions that indicate an increased risk of heart disease, stroke and type 2 diabetes. These results are presented in the appendix because the prevalence of these risk factors, while growing, is still low among the less than thirty population in Bangladesh. The metabolic syndrome index includes an indicator for hypertension, overweight, and underweight. Measures of weight and blood are usually thought of as flow measures of health because they can change over time to reflect current behaviors and environment, and don't necessarily reflect changes in the early childhood environment. However, the

literature on the Development Origins of Health and Disease posits that metabolic syndrome can also reflect on childhood health since a child's environment (e.g. nutrition, health, stress, or exposure to toxins) can create a predisposition to poor health in adulthood, such as risk of obesity, high blood pressure and disease, through epigenetic mechanisms (Gluckman, Buklijas and Hanson, 2016).

To help with interpretation and account for differences in outcomes with age and gender, outcomes are standardized into z-scores using WHO international standards when available. Otherwise, z-scores are created by internally standardizing by age and sex using the comparison group mean with six-month age bins for those age 6 and younger, and one year age bins for those age 6 and older. Cognition measures are all internally standardized into z-scores because norms for the cognitive tests do not exist for Bangladesh. Details on how the outcomes are defined are in Appendix B.3.

MHSS2 includes a battery of cognitive tests that include number of domains of cognition. To reduce concerns regarding multiple hypotheses, we create a cognitive index by averaging the z-scores across the cognitive tests. For the first generation, the cognitive index includes the Mini Mental State Exam (MMSE), Digit Span Forward and Backwards, and Raven's Colored Progressive Matrices (Ravens). Digit Span tests memory which is thought to increase with intelligence; Ravens is a non-verbal and simple measure of general intelligence and perhaps the most common and popular test for people above the age of five (Kaplan and Saccuzzo, 2009); the MMSE was developed as a global assessment of cognitive status and is also a frequently used brief cognitive screening test (Ismail, Rajji and Shulman, 2010). Digit Span Forward is the only cognitive test available in both the in-person and the phone surveys, so we do not create a cognitive index for people who responded to the phone survey resulting in a lower sample size for men for this outcome.

For the second generation, cognitive tests for children under age seven are drawn from The Denver Developmental Screening Test (Frankenburg et al., 1975). We use the subcomponents on language, fine motor, and gross motor skills to create a Child Development Index. For respondents age seven and older, the cognitive index is comprised of the MMSE, digit span forward and backward, Ravens, a test of memory drawn from a subtest from the Woodcock Johnson IV Test of Cognitive Abilities (Memory for Names), and a three minute timed visual matching test to measure perceptual speed.

Finally, we examine effects on grades attained for the first generation but not the second generation. About half the second generation sample are too young to attend school, and for those who are school age, most have not yet reached the age of school drop out.

3.4 Intent-to-Treat Indicator and Linking to Baseline Variables

Access to the MCH-FP program was based on the village of residence of the individual during the program period. Because a person's residence after program start is potentially endogenous, we use DSS and census data to create an intent-to-treat indicator based on the village of residence for an individual's first household head prior to 1977.[15] We take advantage of the fact that each individual has a unique ID that allows us to link the MHSS1/2 data with the DSS and census data, and use the following sequence of linkages. First, we link our respondents to the 1974 census through the household head of their first residence in the DSS area. If their household head was not present in the 1974 census, we identify that person's first household head in the DSS area and link that new person to the 1974 census. Finally, remaining unlinked individuals are assigned a treatment status using the location of their household head in the DSS area after the 1974 census, but before the inception of MCH-FP in 1977.[16] The intent-to-treat variable, Treat, takes the value of 1 if the 1974 census-linked household head was living in a village in the treatment area in 1974 or migrated into a village in the treatment area between 1974 and 1977.

Baseline characteristics from the 1974 census are linked to individuals in the same manner used to construct treatment status. For the few individuals that could not be linked to the 1974 census, missing baseline characteristics are assigned means based on treatment status, sex, and cohort.[17] Finally, the village from the 1974 census link is used to cluster standard errors in our analysis.

4 Estimation Strategy

4.1 Program Design

Treatment and comparison areas were built into the design of the MCH-FP program. Randomization was not used to determine treatment status. Instead, villages were assigned to treatment and comparison areas based on blocks of contiguous villages (Figure 2)[18] that were viewed as socially and economically similar and geographically insulated from outside influences at the time (Phillips et al., 1982). The block

15. The treatment indicator would be nearly identical if individuals were linked to 1974 through their fathers and grandfathers. Less than 0.5% of the sample would have been assigned a different treatment status. We use household head because this sequence of linkages results in more direct links to the 1974 census, and therefore fewer missing baseline characteristics.
16. We link over 96% of individuals in our sample to the 1974 census through their first household head. An additional 3 percent link to the 1974 census through that person's first household head. The remaining less than 1 percent link through their household head's location in the DSS after the 1974 census, but before program inception in October 1977.
17. Only 12 male and 13 female respondents have missing baseline data.
18. Villages were assigned to one of 6 blocks, 4 treatment blocks, and 2 control blocks.

design was important to mitigate potential spillovers into the comparison area of information about the family planning interventions (Huber and Khan, 1979) or from the positive externalities generated by vaccination. Past research shows that the treatment and comparison areas are similar with respect to potentially targeting outcomes including rates of mortality and fertility (Koenig et al., 1990; Menken and Phillips, 1990; Joshi and Schultz, 2013). This is important since it means the program was not placed first in areas that had poor child health or high fertility. In addition, most pre-intervention household and household head characteristics, migration stocks and flows, and employment sector of the household head were well balanced.[19] Barham (2012) also shows that cognitive functioning was similar between the treatment and comparison area in 1996 for those whose cognitive functioning was not likely to have been affected by the program.

For the sample used in this paper, we use the 1974 census to determine if the baseline characteristics are balanced after attrition. Table 1 reports differences in means between the treatment and comparison area and normalized differences in means (difference in the means divided by the standard deviation of the mean for the comparison group) because it is not influenced by sample size (Imbens and Wooldridge, 2009). Normalized differences bigger than 0.25 standard deviations are generally thought to be substantial.

Individual and household characteristics are similar across the treatment and comparison areas for the 1982–1988 first generation cohort, with only 4 out of the 23 baseline characteristics showing significant differences below the 5 percent level for the pooled sample of all ages and sex and all normalized differences are small (less than 0.12) suggesting the differences are not substantial.[20]

The two largest normalized differences are religion and access to tubewell water. The treatment area had fewer Muslims than the comparison area (84 versus 95 percent of household heads), the remainder being Hindu. All specifications control for if an individual is Muslim, and robustness analysis shows that the results remain similar if only Muslims are included in the analysis. In the treatment area, 30 percent of households use tubewells compared to only 16 percent of households in the comparison area. The difference in tubewell water in 1974 is the result of a government program, so does not reflect household income, propensity to drill a tubewell, or a household's concern about child health or potentially other unobservables that could be correlated with a person's anthropometrics, or cognition. Regardless, the difference in access to tubewell water for drinking is concerning since tubewell water is often thought to

19. See Barham (2012); Barham and Kuhn (2014); Barham, Kuhn and Turner (2021).
20. Results are similar if examined by cohort and sex, though most have only three significant differences.

be cleaner than other sources of water. However, there is widespread groundwater arsenic contamination in the tubewells in Bangladesh (Chowdhury et al., 2000) which is a serious health concern and has been shown to led to reduce IQ among school-aged Bangladeshi children (Wasserman et al., 2004). As a result, the direction of the bias is unclear. In the robustness analysis, we control for the level of arsenic in the tubewells as part of the extended controls.

These findings, together with previous results strongly suggest that the two areas had very similar observable characteristics. To account for the few differences in baseline characteristics we include the observables listed in Table 1, as well as provide robustness analysis to further explore the imbalance between religion and tubewell water as explained above.

4.2 Identification Strategy and Empirical Specification

Our analytical approach builds on Barham (2012) and exploits the quasi-random nature of the treatment and comparison areas to estimate intent-to-treat effects of the MCH-FP program on first and second generation human capital. Ideally, we would like to show that human capital was similar between the treatment and comparison areas prior to the intervention. This is not possible given that individual level human capital outcomes were not collected prior to MCH-FP. Instead, for the first generation, we estimate a double-difference model that uses an older cohort, born between 1947-1969, whose human capital measures were unlikely to be impacted by the program to account for any pre-existing differences in human capital.

For the first generation, the double difference model for individual i, from household h located in village v in 1974, is specified as follows:

$$Y_{iv} = \beta_0 + \beta_1 \text{Treat}_{iv} \times (1982\text{–}1988_i) + \beta_2 \text{Treat}_{iv} + \beta_3 1982\text{–}1988_i + \alpha_t + X_{ih}\Gamma + \varepsilon_{iv} \tag{2}$$

where Y_{iv} is the outcome of interest, Treat_{iv} is the binary treatment status variable. $1982\text{–}1988_i$ is one if person i is was born between 1982–1988 and zero if between 1947-1969. Birth year fixed-effects, α_t, are included to control for differences in the outcome due to year of birth as well as other events that may be correlated with birth year. X_{ih} is a vector of individual (sex and religion), baseline household and household head characteristics found in Table 1 and the interaction of the baseline variables with the age cohort. Standard errors are clustered at the pre-program village level to account for likely intracluster

correlation in the error term. All models are weighted to account for attrition from birth to MHSS2. Appendix E provides details on the weights.

In this model, β_1 is the double-difference estimator and provides the long-run effect of the MCH-FP for those born between 1982–1988 when the are adults approximately aged 24-30 in MHSS2. This is the combined effect of all program interventions, including intensive child health interventions. The double difference model assumes that the treatment and the comparison group would have had the same trend in the outcome variable in the absence of the MCH-FP program. This is not a testable assumption, but seems reasonable given the similarity between treatment and comparison area prior to the program. However, to the extent that people lose height and cognitive functioning as they become elderly differentially between the treatment and comparison area in the pre-program cohort, then the double difference estimates may be biased. In section 5.2 we discuss potential threats to causal identification and perform a number of robustness checks to show the sensitivity of our results.

For the second generation, we use a single-difference model as the main specification and a double-difference model as a robustness check. We do not use the double-difference as a main specification because determining a group to measure the pre-program difference is difficult. The 1947-1969 cohort cannot be used because outcomes and measurement of the variables differ between adults and children for most variables. Instead, we identify a less affected cohort of children at a similar point in development, and hence the same age, as the children in the single difference estimator (see section 6.2 for more details).

The single-difference model for individual i, born to mother m from village v is specified as follows:

$$Y_{imv} = \beta_0 + \beta_1 \text{MomTreat}_{imv} + \alpha_{6\text{mo}} + X'Z + \varepsilon_{imv} \tag{3}$$

where Y_{imv} is the outcome of interest, MomTreat_m is the binary treatment status variable of the mother (who is in the first generation). Six month age fixed-effects, $\alpha_{6\text{mo}}$, are included to control for differences in the outcome due to age as well as other events that may be correlated with age. X is a vector of individual (sex and religion), baseline household and household head characteristics of the mother found in Table 1. Standard errors are clustered at the village level to account for likely intracluster correlation in the error term and are weighted for attrition from birth to MHSS2 of their mother. The coefficient β_1 is the average difference in the outcome for children whose mothers were from a treatment village versus those who were from a comparison village. Similarly to the first generation, standard errors are clustered at the

mother's pre-program village to account for likely intracluster correlation in the error term. All models are weighted to account for attrition from birth to MHSS2 using the mother's weight. Appendix E provides details on the weights.

5 First Generation Results

5.1 ITT Effects on Human Capital

ITT effects of the MCH-FP program on stock measures of human capital— height, cognition index, and grades attained— are presented in Table 2. Pooled results are in Panel A, and results disaggregated by sex in Panels B and C. Comparison group means are presented at the bottom of each panel for variables that are not z-scores along with a p-value of a test for random inference.[21] Based on the comparison group means, short-stature is still an issue for approximately 8 percent of men and 11 percent of women, and average grades attained, while growing, is still low compared to developed countries at an average of 7 years.

The point estimate on the variable *Treat* is the difference in means between the treatment and comparison areas for the pre-program 1947-1969 cohort. The point estimate on this variable is close to zero and not significantly different for any of the outcomes in Table 2, providing further evidence of the similarity in the treatment and comparison areas prior to MCH-FP.

Pooled results demonstrate important effects for height and grades attained, but not for cognition.[22] Specifically, the program led to a 0.97 cm increase in height (significant at the 1 percent level), 4 percentage point decrease in short stature (significant at the 10 percent level), 0.42 increase in grades attained (significant at the 5 percent level). Effects are close to zero for the cognition index (-0.011 SD) and for each of the components of the index (Table 3).

Effects are similar between the sexes for height and short-stature but not for education, and there is no impact for either sex on the cognition. For height, males experienced a 1.05 cm (significant at the 10 percent level) increase and females 0.99 cm (significant at the 5 percent level) increase. The standard

21. With any assignment of village-level treatment status, significant treatment effects could occur simply by chance. Following Athey and Imbens (2017), we simulate the distribution of treatment effects that would occur from randomly assigning a fixed number of villages to treatment.

22. For completeness, we also present results of the impact of the MCH-FP program on the cohort that were born when only family planning services were provided, those born between 1977-1981, who are age 31-34 during the MHSS2 survey. As found in Barham (2012), this cohort did not experience any medium term impacts and Table A2 shows there are still no impacts on this cohort.

errors for males may be larger than females due to height being self-reported for those interviewed in the phone survey. For education, effects are driven solely by males. who experienced 0.96 increase years for grades attained (significant at the 5 percent level). The lack of effect on education for females is not surprising because there was a nationwide program for girls' education operating at the time in both the treatment and comparison area. The Bangladesh Female Secondary Education Stipend Program started in Matlab in 1984 (see Appendix C for more details) and continued to operate when the 1982-1988 cohort was in secondary school. This program provided a stipend for females in both the treatment and comparison areas to attend secondary school, as well as tuition vouchers. This program is documented to have increased female educational attainment by 1.2 years (Shamsuddin, 2015).

Finally, we present effects on metabolic health, a flow measure of health, in Table A1. We use a single-difference estimator since this is a flow variable and can change in adulthood, making it possible for the MCH-FP program to affect the metabolic index for the pre-program cohort in the double-difference model. There is no affects on the metabolic index, or its components, which may be a reflection of the relatively young age of the sample population.

5.2 Robustness

This section summarizes robustness checks to test the validity of the first generation results, weighting scheme, and attrition bounds. Given the treatment and comparison areas are in contiguous geographic areas, one key threat to identification is if a shock occurred in one of the areas but not the other, such as a disease outbreak or placement of another program. Fortunately, this concern is limited because the study area is relatively small and homogeneous (almost all people in the comparison area are located within 5 km of the treatment area), and the programs and services available in both areas tend to be similar because they are in the same district.

Pre-Trend Analysis—In Table D1, we examine if there are any pre-MCH-FP differences between the treatment and comparison areas that could be associated with lifecycle effects by examining effects on a similarly aged sample (aged 24-30) in 1996 using MHSS1. This cohort was born more than five years prior to the MCH-FP intervention. We find no program effects for height, cognition (as measured by the Mini Mental State Exam), or years of education. Effect sizes are on the whole small, negative and statistically insignificant.

Spatially Correlated Errors—We find no evidence of spatially correlated errors across villages (results

not reported). This could arise, for example, if there was a health shock such as a disease outbreak in a given year in several villages in the treatment area but not any of the other villages. We also have no knowledge of a disease outbreak that affected just one area from the decades of demographic surveillance data on mortality and disease even in the early years (Fauveau, 1994).

Extended Controls—We include an extended set of controls in Table D2 Panel B to account for prominent changes that took place in Matlab during 1982-2012 that could potentially bias the results. These changes include the introduction of a river embankment for irrigation and flood protection in 1987, introduction of micro-credit and other programs through BRAC in the 1990s, increased support for education, construction of village healthcare facilities, and discovery of arsenic in some deep tube-wells (program details are in Appendix C). Again, results are qualitatively the same.

Unbalanced Controls—We control for a limited set of baseline variables that only includes variables that are unbalanced for any age or sex cohort. These controls include religion, tubewell water, household head and spouse age and education. Table D2 Panel C shows results are qualitatively similar, though effects on grades attained are slightly attenuated but still statistically significant at the five percent level for males.

Muslims Only—Given the imbalance of religion between the treatment and comparison areas prior to the MCH-FP program, we test the main results for only those that identify as Muslim in Table D2 Panel D. Results are similar, though point estimates on height for females are slightly smaller (0.82 cm) and standard errors for both sexes higher due to the smaller sample sizes.

Unweighted and Attrition Bounds— Unweighted results are presented in Table D2 Panel E. Unweighted results are similar, suggesting that the weights are not driving the results. In Table D3 we present Kling-Leibman attrition bounds (Kling, Liebman and Katz, 2007) and find qualitatively similar results, suggesting attrition is not an issue.

Self-Reported Height—Finally, we report results where males who self-reported their height because they were interviewed via phone are excluded. Table D4 shows results remain similar.

5.3 ITT effects on Marriage, Fertility and Child Mortality of First Generation Women

We next examine how an improved childhood health environment may have impacted first generation women's fertility, marriage patterns, and the mortality rates of their children in Table 4. These are important mechanisms that may be linked to biological pathways that could lead to selection in the

second generation. Panel A considers all first generation females, and Panel B and C first generation mothers with a firstborn male and female respectively.

There are no significant differences in marriage or fertility outcomes between the treatment and comparison areas for females in the 1982-1988 cohort based on marital status, age at menarche, number of live births, if the female lost a pregnancy (stillborn, miscarriage, or induced abortion), and if a firstborn child born live later died. These results indicate that second generation impacts are unlikely to be affected by fertility selection.

The majority of the 1982-1988 females are married (93%), and 83% have at least one child providing a sample of 677 first born children. Table 4, columns 6-8 focus specifically on 1982-1988 females who have at least one live birth. The ITT effects indicate that, between the treatment and comparison areas, the sex composition of firstborn children is balanced (column 6), that females had their first child at similar ages (column 7), and that there is no difference in the number who have a second born child (column 8; 56 percent of those with a firstborn child). Panels B and C further shows that fertility and child mortality are also similar between treatment and comparison areas by the sex of firstborn child. Together these results indicate the second generation results are not likely driven by marriage, fertility or mortality selection.

6 Second Generation Results

6.1 ITT Effects on Human Capital

Single-difference ITT effects on the second generation are presented in Table 5. These children range from ages 0 to 14 and are the firstborn children of first generation women. Results are reported for height-for-age (WHO z-score), stunting, and a cognition index.[23] Results for both sexes are in Panel A, and for males and females in Panel B and C, respectively. Comparison group means of any variables that are not internally standardized z-scores are at the bottom of the table together with the same test of random inference as the first generation.

Results pooled by sex in Panel A show there are no intergenerational effects. However, these pooled results mask important differences by sex. While effects for males are generally small, negative and statistically insignificant, first born females experienced a 0.33 SD increase in height-for-age (significant at

23. Results on all outcomes are presented for the entire age group with the exception of cognition as there are not suitable tests for the entire age range. For the 0-6 year-olds, we use the Denver Developmental Screening Test for a measure of child development. For the 7-14 year-olds, we create a cognition index based on the six cognitive tests discussed in the data section.

the 5 percent level), a 50 percent decrease in the likelihood of stunting (significant at the 1 percent level), and 0.27 SD increase in the cognitive index (significant at the 10 percent level) for the 7-14 year-olds, but not the 0-6 year olds. Table A4 indicates that there are sizable effects for each component of the cognitive index for the 7-14 year-olds, except those associated with memory.

6.2 Robustness

To examine the sensitivity of the findings for the second generation, we conduct a number of robustness checks including constructing a comparison cohort for a plausible double difference model, accounting for changes over time in the health and labor supply in the area and other potentially confounding programs. Our set of findings is robust to these alternative specifications.

Double-Difference—To test the validity of the single difference results we present double difference results in Table D5. Identifying a cohort of similarly aged children that is not affected by the program is challenging as women born prior to the first generation have older children. Instead, we identify a less affected cohort of children at a similar point in development, and hence the same age. The double difference cohort includes children age 0-14 from any birth order of women born before the 1982, the first birth year of the first generation. We focus on these women because their human capital was not affected by MCH-FP program (Barham, 2012). While we would like to restrict the less affected cohort to also be first birth because birth order can affect human capital, there is not a large enough sample. The double difference results are qualitatively similar to the single difference results. Based on effects for height, the single difference estimator provides a slightly more conservative estimate of the program effects.

Changes Over Time—In Table D6 Panel B, we include a set of extended controls to account for changes in the health and education supply by the time of birth of each child as well as two measures of arsenic in the drinking water. One that was used for mothers and arsenic in MHSS2, which may be more relevant for children. Again, results are similar.

Confounding Programs—One potentially confounding program is the Maternal, Neonatal and Child Health (MNCH) program. It began in 2007 and the targeting of this program overlapped with the treatment area of the MCH-FP program, potentially biasing the intergenerational effects. The goal of the program was to encourage delivery at a facility with skilled obstetric care rather than at home and provide counseling on newborn and maternal well-being (see Rahman et al. (2011) for a detailed description of the program). For robustness, we control for mother's exposure to the MNCH program based on her residence in 2005

interacted with a dummy indicating whether the child was born on or after 2007. Results presented in Table D6 Panel C shows MCH-FP program effects again remain similar after controlling for the MNCH program. Another well-known program in the study area is the Maternal and infant Nutritional Intervention (Minimat). This is an RCT that enrolled children born prior to the second generation cohort in this study, so should not affect the single-difference results, but could bias the double-difference results (see Appendix C for more details).

Unbalanced Controls—We control for a limited set of baseline variables by only including baseline variables that are unbalanced for any age or sex cohort. These controls include religion, tubewell water, household head and spouse age and education. Table D6 Panel D shows results are qualitatively and statistically similar.

Muslims Only—Given the imbalance of religion between the treatment and comparison areas prior to the MCH-FP program, we restrict results to only those who are Muslim in Table D6 Panel E. Results are qualitatively and statistically similar.

Weights and Attrition Bounds—Unweighted results are presented in Table D6 Panel F and results using Kling-Leibman attrition bounds in Table D7. Results are qualitative and statistically similar demonstrating the results are not driven by the weights or attrition.

7 Heterogeneity by Matrilineal Health Endowment

To explore how the program affected inequity, we examine the interaction between the initial human capital endowment (θ_{jp} in our conceptual framework) and the improved childhood health environment resulting from the MCH-FP program. As noted in the conceptual framework, the MCH-FP program increased public investment, I_1, for the first generation, and the effects of this improvement may impact adult human capital formation deferentially depending on the human capital of an individual's parents, which in-turn may affect the human capital endowment and program effects of the second generation.

We use the pre-program matrilineal health endowment as a measure of the initial health capital endowment. We proxy the endowment with terciles of height of the first generation's mother (second generations' maternal grandmother) since it was not affected by the MCH-FP program. We use adult height as it is a stock measure of health that is known to reflect genetics as well as the early life health environment.

For the first generation, impacts of the MCH-FP program are strongest for those with the lowest matrilineal health endowment leading to a reduction in inequality as indicated by height and stunting. Table 6 demonstrates that program impacts are mostly concentrated in the lowest tercile of the first generation mother's height, or shortest mothers. These results indicate the program reduced the gap between those with the lowest and highest health endowment as measured by matrilineal height. For the shortest tercile, the ITT effect on height for the pooled sample is 1.85cm (significant at the 1 percent level) and is accompanied by a 50 percent reduction in short stature (significant at the 1 percent level). Effects sizes by gender are similar for height, but males in the shortest tercile also experience an increase in the cognition and education indices of 0.2 and 0.46 SD respectively. There were also height gains in the tallest tercile for the pooled sample, but they were smaller, 0.97cm (significant at the 10 percent level), indicating that the while the program reduced inequalities, it still lead to improvements for those with the highest matrilineal health endowment.

In contrast, second generation effects are strongest for the tallest tercile of the maternal grandmother's height (Table 7). Effects are concentrated in females and show an increase in height of 0.54 SD (significant at the 1 percent level) and a reduction in stunting of 27 percentage points (significant at the 1 percent level) on a comparison group mean of 25 percent, indicating the stunting may have been eliminated in the grandchildren of the the tallest grandmothers. Females who were aged 7-14 at the time of the survey, also experienced an increase in the cognition index of at 0.71 SD (significant at the 1 percent level). An increase in height of 0.25 SD, a reduction in stunting of 12 percentage points, and an increase in the cognitive index by 0.06 SD were also experienced among granddaughters of the shortest tercile grandmothers, but effects are not statistically significant from zero, or between the shortest and tallest grandmothers, with the exception of cognition. It is unclear why effects for the second generation would be driven by the tallest grandmothers, but could indicate pre-program health inequalities based on endowments were partially addressed in the first generation, and the second generation differences reflect differential investments of first generation mothers in their daughters.

8 Exploration of Mechanisms

The conceptual framework outlined in Section 2.3 identifies a number of potential behavioral and biological mechanisms that may be governing the impact of the MCH-FP program on human capital

accumulation in the second generation. We are unable to examine biological pathways as we do not have any biological markers or proxies, such as birth weight. However, for individuals in the second generation, we have several measures of investments made during different points in early childhood to illuminate potential behavioral mechanisms. We group our measures into four categories: (1) pre-natal health and delivery, (2) post-natal health and human capital, (3) parental engagement, and (4) mother characteristics.[24]

We explore if investments are different between MCH-FP treatment and comparison individuals in the second generation using a single difference ITT estimator. We graph the point estimate and the 95 percent confidence interval in Figure 6. Results are presented separately by sex and the difference between the two can be determined by the overlap of the confidence intervals. The comparison group mean is in parentheses next to the point estimates.

We consider three prenatal care outcomes: if the mother had a prenatal checkup, if the birth occurred at a skilled delivery location, and if a trained individual assisted with the delivery. Second generation firstborns of both sexes were more likely to have been delivered in a skilled location and with a trained delivery assistant. While these impacts are statistically different from zero, there is not a statistical difference between males and females. To the extent that male and female human capital development react differently to these investments, with females responding more than males, it could explain the pattern of results for the second generation. However, there is not strong empirical proof of this pathway.

Post-natal health investments are measured by the number of vaccinations received, and if the individual attended preschool. None of the point estimates are statistically different from zero, and again, there is no difference in health promoting investments between males and females. Parental engagement is measured by an interview observation of positive parenting. These measures are also not statistically different from zero for males or females.

Lastly, we consider if the empowerment of the mother is differentially impacting males versus females in the second generation and find the program had no effect on female empowerment. We measure empowerment through an index of survey questions related to women's decision making power, attitudes toward gender equivalence in social issues, attitudes toward husband violence and women's mobility.

Overall, we find few behavioral effects and no statistically significant differences in investment by sex that could potentially explain the sex differences in human capital for the second generation. The

24. See Appendix B for more details on the measures).

lack of effects on investments by sex suggest that either we don't have data on the pertinent investments generating our results or that biological mechanisms play an important role. Given the importance of parenting and early childhood stimulation (Black et al., 2017), it is possible that mothers who benefited from the program as children, invested more in their girls relative to boys.

9 Conclusion

Exposure to infectious diseases, high rates of fertility, and undernutrition are common and related obstacles faced by low-income individuals that limit reaching their human capital potential (Prendergast and Humphrey, 2014; Grantham-McGregor et al., 2007). Mounting correlational evidence shows that disadvantage in human capital experienced in one generation can persist to the next. It is imperative to examine the causal effects of interventions designed to improve the health environment in childhood on adult human capital and on the next generation to understand the potential role of policy in reducing the spread of inequalities over time.

Exploiting the quasi-random placement of a mother and child health and family planning program in Matlab, Bangladesh, results highlight that increased public investment during early childhood leads to improvements in human capital not only for the first generation that directly benefited from the interventions, but also their children, the second generation. We find that improving the child health environment increased adult height in the first generation by approximately one centimeter for both sexes, and grades attained for males, but there was no effect on cognitive functioning.

The results also demonstrate important intergenerational impacts for females but not males. Females in the second generation are taller and have higher levels of cognition. While not statistically different, effects on height and cognition are even larger for the second generation than the first, suggesting improvement may grow over time. In addition, the difference in results between height and cognition, and the resurgence of the effect on cognition in the second generation highlight the importance of considering many dimensions of human capital development. Human capital measures are affected differently by one's environment and sensitive ages differ, making it important to examine effects at different points in the life cycle.

A unique strength of this paper is the ability to link data from three generations to explore how pre-program health endowments (proxied by height of generation one's mother) interact with a policy to

affect human capital over time and generations. The finding that the largest improvements in adult human capital for the first generation are experienced by those that have the lowest levels of pre-program health endowment suggests that even though the MCH-FP program was not targeted to low endowment individuals it did reduce inequality. This reduction in first generation inequality manifested in second generation impacts that were more equal by pre-program health endowments. To our knowledge, previous research has not been able to examine how impacts of an improved childhood health environment are differentially transmitted by human capital endowment. However, understanding this process is important for policy aimed at reducing inequalities.

Behavioral and biological pathways link human capital between the generations. We find no effects on child mortality or fertility selection that may be linked to biological pathways, nor effects on early childhood health investments or maternal empowerment that can explain the pattern of second generation results by sex. Gender effects are also not due to females initially having worse human capital than males, leaving more room for improvement.

These finding suggests policy makers may be underestimating benefits of programs that improve the childhood health environment if they do not consider intergenerational effects and the effect on reducing health inequalities over generations. While we examine many mechanisms, a limitation of the study is that it was not designed to examine biological mechanisms or the many behavioral pathways. Further research is needed to understand the key mechanisms and gender dimensions of the intergenerational transmission of human capital.

References

Ahlburg, Dennis. 1998. "Intergenerational Transmission of Health." *American Economic Review*, 88(2): 265–270.

Almond, Douglas, and Bhashkar Mazumder. 2011. "Health Capital and the Prenatal Environment: The Effect of Ramadan Observance during Pregnancy." *American Economic Journal: Applied Economics*, 3(4): 56–85.

Almond, Douglas, Janet Currie, and Valentina Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature*, 56(4): 1360–1446.

Athey, Susan, and Guido W Imbens. 2017. "The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments*. Vol. 1, 73–140. Elsevier.

Attanasio, Orazio, Costas Meghir, and Emily Nix. 2020. "Human Capital Development and Parental Investment in India." *The Review of Economic Studies*, 87(6): 2511–2541.

Attanasio, Orazio P. 2015. "The Determinants of Human Capital Formation During the Early Years of Life: Theory, Measurement, and Policies." *Journal of the European Economic Association*, 13(6): 949–997.

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

Barham, Tania, and Randall Kuhn. 2014. "Staying for Benefits The Effect of a Health and Family Planning Program on Out-Migration Patterns in Bangladesh." *Journal of Human Resources*, 49(4): 982–1013. Publisher: University of Wisconsin Press.

Barham, Tania, Brachel Champion, Andrew D Foster, Jena D Hamadani, Warren C Jochem, Gisella Kagy, Randall Kuhn, Jane Menken, Abdur Razzaque, Elisabeth Dowling Root, et al. 2021. "Thirty-five Years Later: Long-term Effects of the Matlab Maternal and Child Health/Family Planning Program on Older Women's Well-being." *Proceedings of the National Academy of Sciences*, 118(28).

Barham, Tania, Randall Kuhn, and Patrick Turner. 2021. "No Place Like Home: Long-Run Impacts of Early Child Health and Family Planning on Labor and Migration Outcomes." *Working Paper*.

Barr, Andrew, and Chloe Gibbs. 2017. "Breaking the cycle? Intergenerational effects of an anti-poverty program in early childhood."

Behrman, Jere R, Maria C Calderon, Samuel H Preston, John Hoddinott, Reynaldo Martorell, and Aryeh D Stein. 2009. "Nutritional Supplementation in Girls Influences the Growth of Their Children: Prospective Study in Guatemala." *The American Journal of Clinical Nutrition*, 90(5): 1372–1379.

Bhalotra, Sonia, and Samantha Rawlings. 2013. "Gradients of the Intergenerational Transmission of Health in Developing Countries." *Review of Economics and Statistics*, 95(02): 660–672.

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

Black, Maureen M, Susan P Walker, Lia CH Fernald, Christopher T Andersen, Ann M DiGirolamo, Chunling Lu, Dana C McCoy, Günther Fink, Yusra R Shawar, Jeremy Shiffman, et al. 2017. "Early childhood development coming of age: science through the life course." *The Lancet*, 389(10064): 77–90.

Black, Sandra, and Paul Devereux. 2011. "Recent Developments in Intergenerational Mobility." In *Handbook of Labor Economics*. 1487–1541.

Chowdhury, Uttam K, Bhajan K Biswas, T Roy Chowdhury, Gautam Samanta, Badal K Mandal, Gautam C Basu, Chitta R Chanda, Dilip Lodh, Khitish C Saha, Subhas K Mukherjee, et al. 2000. "Groundwater Arsenic Contamination in Bangladesh and West Bengal, India." *Environmental Health Perspectives*, 108(5): 393–397.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review*, 97(2): 31–47.

Currie, Janet, and Tom Vogl. 2013. "Early-Life Health and Adult Circumstance in Developing Countries." *Annual Review of Economics*, 5(1): 1–36.

Deming, David. 2009. "Early Childhood Intervention and Life-cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics*, 1(3): 111–34.

East, Chloe N., Sarah Miller, Marianne Page, and Laura R. Wherry. 2019. "Multi-generational Impacts of Childhood Access to the Safety Net: Early Life Exposure to Medicaid and the Next Generation's Health." National Bureau of Economic Research 23810.

El Arifeen, Shams, Eva-Charlotte Ekström, Edward A Frongillo, Jena Hamadani, Ashraful I Khan, Ruchira T Naved, Anisur Rahman, Rubhana Raqib, Kathleen M Rasmussen, Katarina Ekholm Selling, et al. 2018. "Cohort Profile: The Maternal and Infant Nutrition Interventions in Matlab (MINIMat) Cohort in Bangladesh." *International Journal of Epidemiology*, 47(6): 1737.

Fauveau, Vincent. 1994. *Matlab: Women, Children and Health*. Dhaka; icddr,b.

Frankenburg, W, J Dodds, AW Fandal, E Kauk, and M Cohrs. 1975. "The Denver Developmental Screening Test–Revised (DDST)." *Denver, CO: University of Colorado Medical Center*.

Gadroen, Kartini, Caitlin N Dodd, Gwen MC Masclee, Maria AJ De Ridder, Daniel Weibel, Michael J Mina, Bryan T Grenfell, Miriam CJM Sturkenboom, David AMC Van De Vijver, and Rik L De Swart. 2018. "Impact and longevity of measles-associated immune suppression: a matched cohort study using data from the THIN general practice database in the UK." *BMJ Open*, 8(11): e021465.

Garcia, Jorge Luis, James J. Heckman, and Victor Ronda. 2021. "The Lasting Effects of early Childhood Education and Promoting the skills and Social Mobility of Disadvantaged African Americans." *NBER Working Paper No 202183*.

Gluckman, Peter D, Mark A Hanson, and Alan S Beedle. 2007. "Early Life Events and Their Consequences for Later Disease: A Life History and Evolutionary Perspective." *American Journal of Human Biology*, 19(1): 1–19.

Gluckman, Peter D, Tatjana Buklijas, and Mark A Hanson. 2016. "The Developmental Origins of Health and Disease (DOHaD) Concept: Past, Present, and Future." In *The Epigenome and Developmental Origins of Health and Disease*. 1–15. Elsevier.

Grantham-McGregor, Sally M, Lia C Fernald, and Kavita Sethuraman. 1999*a*. "Effects of Health and Nutrition on Cognitive and Behavioural Development in Children in the First Three Years of Life: Part 1: Low Birthweight, Breastfeeding, and Protein-Energy Malnutrition." *Food and Nutrition Bulletin*, 20(1): 53–75.

Grantham-McGregor, Sally M, Lia C Fernald, and Kavita Sethuraman. 1999*b*. "Effects of Health and Nutrition on Cognitive and Behavioural Development in Children in the First Three Years of Life: Part 2: Infections and Micronutrient Deficiencies: Iodine, Iron, and Zinc." *Food and Nutrition Bulletin*, 20(1): 76–99.

Grantham-McGregor, Sally, Yin Bun Cheung, Santiago Cueto, Paul Glewwe, Linda Richter, Barbara Strupp, International Child Development Steering Group, et al. 2007. "Developmental Potential in the First 5 Years For Children in Developing Countries." *The Lancet*, 369(9555): 60–70.

Grönqvist, Erik, Björn Öckert, and Jonas Vlachos. 2017. "The Intergenerational Transmission of Cognitive and Noncognitive Abilities." *Journal of Human Resources*, 52(4): 887–918.

Heckman, James J. 2007*a*. "The economics, technology, and neuroscience of human capability formation." *Proceedings of the National Academy of Sciences*, 104(33): 13250.

Heckman, James J. 2007*b*. "The Economics, Technology, and Neuroscience of Human Capability Formation." *Proceedings of the National Academy of Sciences*, 104(33): 13250–13255.

Hertz, Tom, Tamara Jayasundera, Patrizio Piraino, Sibel Selcuk, Nicole Smith, and Alina Verashchagina. 2008. "The Inheritance of Educational Inequality: International Comparisons and Fifty-Year Trends." *The B.E. Journal of Economic Analysis & Policy*, 7(2).

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review*, 106(4): 903–34.

Huber, Douglas H, and Atiqur Rahman Khan. 1979. "Contraceptive Distribution in Bangladesh Villages: the Initial Impact." *Studies in Family Planning*, 10(8/9): 246–253.

icddrb. 1974. "Matlab 1974 Census."

icddrb. 1982. "Matlab 1982 Census."

Imbens, Guido W, and Jeffrey M Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature*, 47(1): 5–86.

Ismail, Zahinoor, Tarek K Rajji, and Kenneth I Shulman. 2010. "Brief Cognitive Screening Instruments: An Update." *International Journal of Geriatric Psychiatry*, 25(2): 111–120.

Joshi, Shareen, and T Paul Schultz. 2013. "Family Planning and Women's and Children's Health: Long-term Consequences of an Outreach Program in Matlab, Bangladesh." *Demography*, 50(1): 149–180.

Kaplan, RM, and DP Saccuzzo. 2009. "Standardized Tests in Education, Civil Service, and the Military." *Psychological testing: Principles, Applications, and Issues*, 7: 325–327.

Khan, M Mahmud, and Richard Yoder. 1998. *Expanded Program on Immunization in Bangladesh: Cost, Cost-effectiveness, and Financing Estimates*. Bethesda, MD: Partnerships for Health Reform Project, Abt Associates Inc., 1998.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*, 75(1): 83–119.

Koenig, M. A., M. A. Khan, B. Wojtyniak, J. D. Clemens, J. Chakraborty, V. Fauveau, J. F. Phillips, J. Akbar, and U. S. Barua. 1990. "Impact of Measles Vaccination on Childhood Mortality in Rural Bangladesh." *Bulletin of the World Health Organization*, 68(4): 441–447.

Koenig, Michael A., Vincent Fauveau, and Bogdan Wojtyniak. 1991. "Mortality Reductions from Health Interventions: The Case of Immunization in Bangladesh." *Population and Development Review*, 17(1).

Maluccio, John A, John Hoddinott, Jere R Behrman, Reynaldo Martorell, Agnes R Quisumbing, and Aryeh D Stein. 2009. "The Impact of Improving Nutrition During Early Childhood on Education Among Guatemalan Adults." *The Economic Journal*, 119(537): 734–763.

Menken, Jane, and James F Phillips. 1990. "Population Change in a Rural Area of Bangladesh, 1967-87." *The Annals of the American Academy of Political and Social Science*, 510(1): 87–101.

Miller, DL. 1964. "Frequency of Complications of Measles, 1963." *British Medical journal*, 2(5401): 75.

Miller, Sarah, and Laura R Wherry. 2019. "The Long-Term Effects of Early Life Medicaid Coverage." *Journal of Human Resources*, 54(3): 785–824.

Mina, Michael J, C Jessica E Metcalf, Rik L De Swart, ADME Osterhaus, and Bryan T Grenfell. 2015. "Long-Term Measles-Induced Immunomodulation Increases Overall Childhood Infectious Disease Mortality." *Science*, 348(6235): 694–699.

Perry, Robert T, and Neal A Halsey. 2004. "The Clinical Significance of Measles: a Review." *The Journal of Infectious Diseases*, 189(Supplement_1): S4–S16.

Phillips, James F., Ruth Simmons, J. Chakraborty, and A. I. Chowdhury. 1984. "Integrating Health Services into an MCH-FP Program: Lessons from Matlab, Bangladesh." *Studies in Family Planning*, 15(4).

Phillips, James F, Wayne S Stinson, Shushum Bhatia, Makhlisur Rahman, and Jyotsnamoy Chakraborty. 1982. "The Demographic Impact of the Family Planning–Health Services Project in Matlab, Bangladesh." *Studies in Family Planning*, 131–140.

Prendergast, Andrew J, and Jean H Humphrey. 2014. "The Stunting Syndrome in Developing Countries." *Paediatrics and International Child Health*, 34(4): 250–265.

Rahman, Anisur, Allisyn Moran, Jesmin Pervin, Aminur Rahman, Monjur Rahman, Sharifa Yeasmin, Hosneara Begum, Harunor Rashid, Mohammad Yunus, Daniel Hruschka, et al. 2011. "Effectiveness of an Integrated Approach To Reduce Perinatal Mortality: Recent Experiences From Matlab, Bangladesh." *BMC Public Health*, 11(1): 1–14.

Rahman, O, J Menken, A Foster, C Peterson, MN Khan, R Kuhn, and P Gertler. 1999. "The Matlab Health and Socio-economic Survey: Overview and User's Guide, 1999."

Reddy, Vinodini. 1987. "Interaction Between Nutrition and Measles." *The Indian Journal of Pediatrics*, 54(1): 53–57.

Richter, André, and Per Olof Robling. 2013. "Multigenerational Effects of the 1918-19 Influenza Pandemic in Sweden." *Working paper / Swedish Institute for Social Research (SOFI)*.

Shamsuddin, Mrittika. 2015. "Labour Market Effects of a Female Stipend Programme in Bangladesh." *Oxford Development Studies*, 43(4).

Thompson, Owen. 2014. "Genetic Mechanisms in the Intergenerational Transmission of Health." *Journal of Health Economics*, 35: 132–146.

van Ginneken, Jeroen, Radheshyam Bairagi, Andres de Francisco, Asdul Mazid Sardar, and Patrick Vaughan. 1998. *Health and demographic surveillance in Matlab: past, present and future*. International Centre for Diarrhoeal Disease Research, Bangladesh.

Venkataramani, Atheendar S. 2011. "The Intergenerational Transmission of Height: Evidence From Rural Vietnam." *Health Economics*, 20(12): 1448–1467.

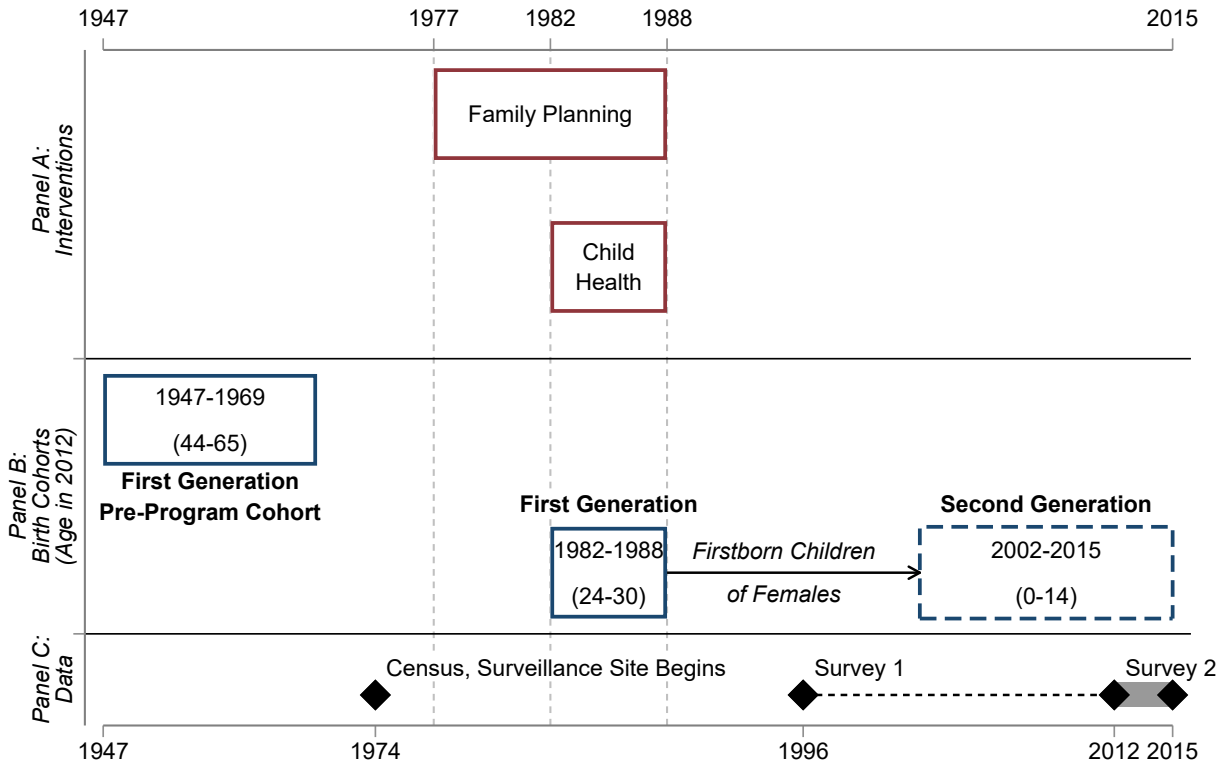
Walker, Susan P, Susan M Chang, Amika S Wright, Rodrigo Pinto, James J Heckman, and Sally M Grantham-McGregor. 2021. "Cognitive, Psychosocial, and Behaviour Gains at Age 31 Years From the Jamaica Early Childhood Stimulation Trial." *Journal of Child Psychology and Psychiatry*.

Walker, Susan P, Theodore D Wachs, Julie Meeks Gardner, Betsy Lozoff, Gail A Wasserman, Ernesto Pollitt, Julie A Carter, International Child Development Steering Group, et al. 2007. "Child Development: Risk Factors for Adverse Outcomes in Developing Countries." *The Lancet*, 369(9556): 145–157.

Wasserman, Gail A, Xinhua Liu, Faruque Parvez, Habibul Ahsan, Pam Factor-Litvak, Alexander van Geen, Vesna Slavkovich, Nancy J Lolacono, Zhongqi Cheng, Iftikhar Hussain, et al. 2004. "Water Arsenic Exposure and children's Intellectual Function in Araihazar, Bangladesh." *Environmental Health Perspectives*, 112(13): 1329–1333.

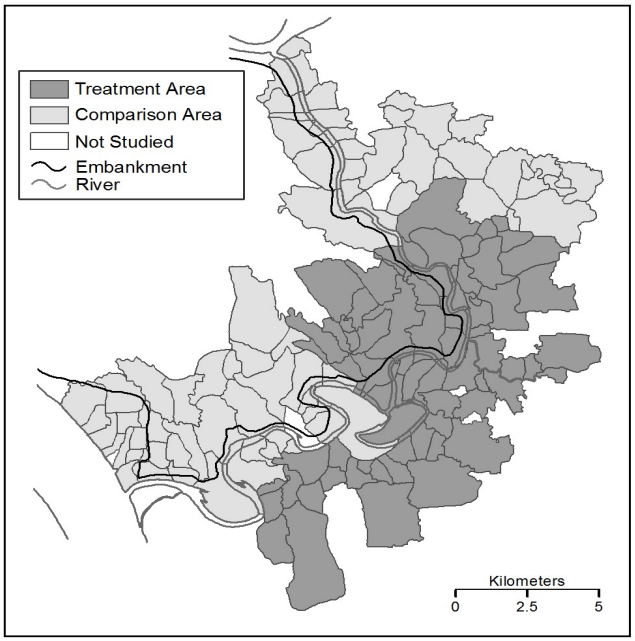
Figures and Tables

FIGURE 1 — TIMELINE OF INTERVENTIONS, BIRTH COHORTS AND DATA COLLECTION



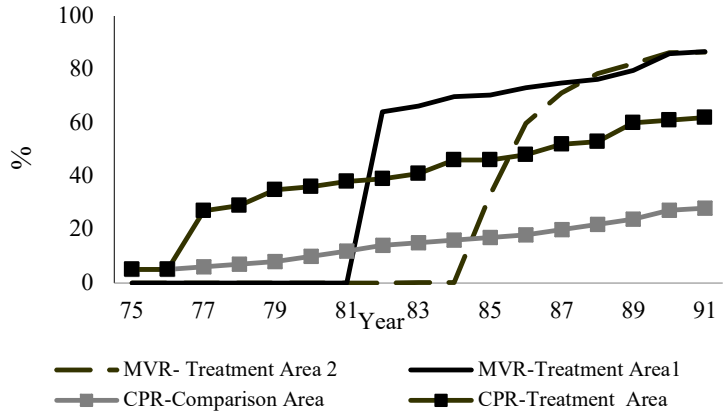
Notes: This figure shows when the timeline of the program, birth cohorts used and data collection events. Panel A shows the experimental period of each arm of the program. Panel B shows the birth cohorts of interests identified by their generation, birth years and ages in 2012. The second generation consists of only firstborn children from females in the 1982-1988 cohort. Panel C shows the timing of data collection events. The 1974 census provides pre-program baseline characteristics of first generation individuals' household and household head. The demographic surveillance site (DSS) recorded life events (marriage, birth, death, migration) from 1974 onward. The first round of the MHSS panel survey was conducted in 1996 (Survey 1) with intense tracking and second round surveys conducted between 2012-2015 (Survey 2). For more detailed descriptions, see Sections 2.1, 3.1, 3.2 and Barham (2012).

FIGURE 2 — MAP OF MATLAB STUDY AREA



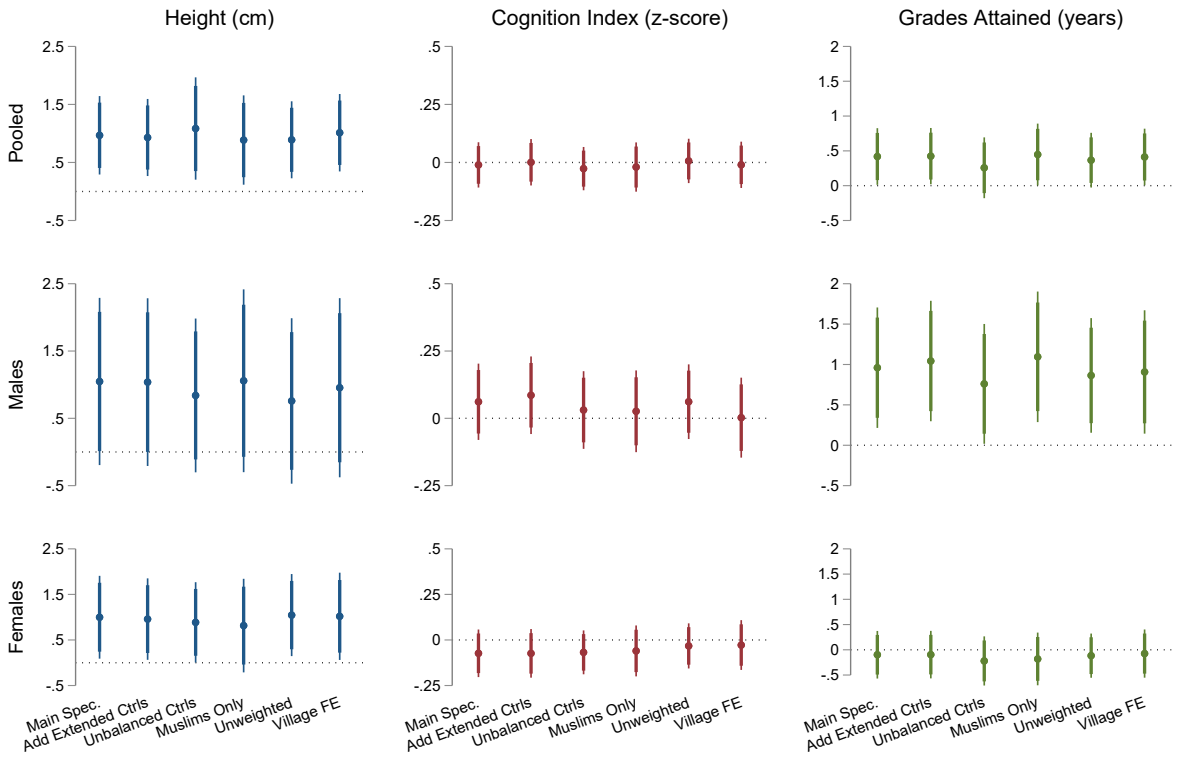
Notes: This figure shows the Matlab study site in Bangladesh where the MCH-FP program was implemented. Subdivisions indicate distinct villages where an individual's household head lived in 1974 used to create the intent-to-treat measure.

FIGURE 3 — TRENDS IN CONTRACEPTIVE PREVALENCE RATE (CPR) AND MEASLES VACCINATION RATES (MVR)



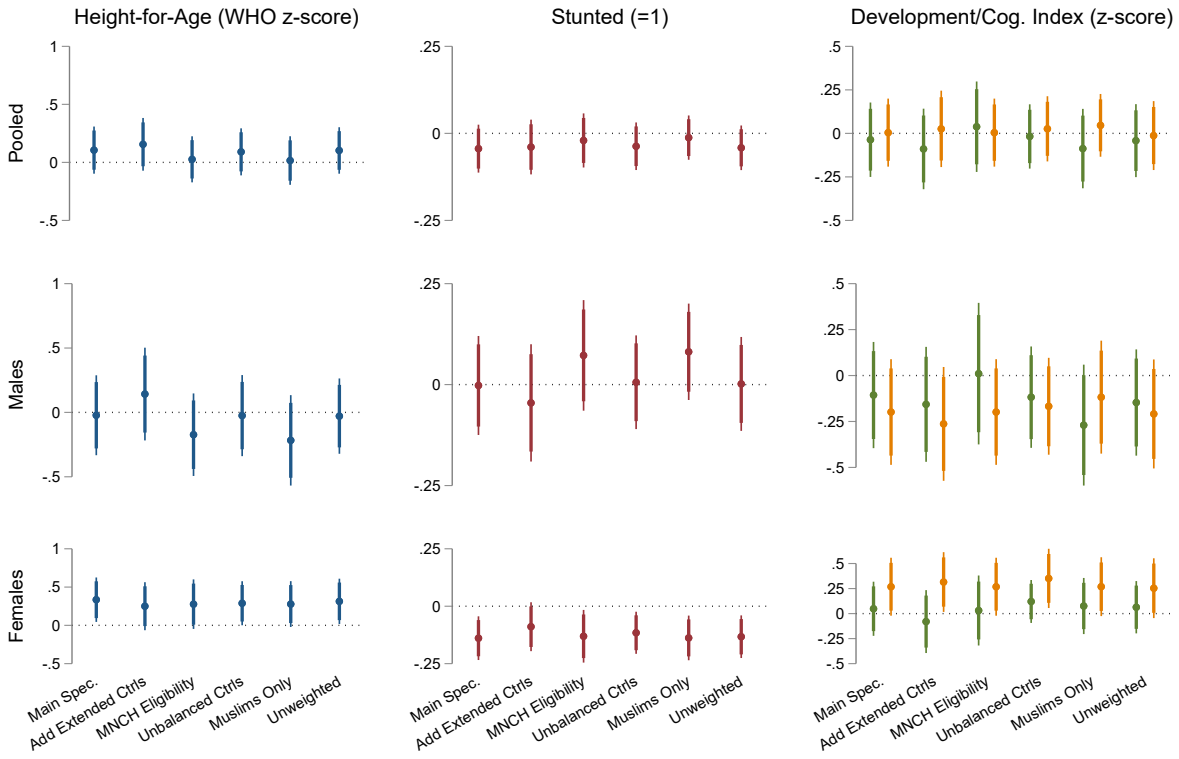
Notes: This figure shows the Contraceptive Prevalence Rate (CPR) and Measles Vaccination Rate (MVR) for children aged 12-59 months by year. Contraceptive use data is from van Ginneken et al. (1998); Measles vaccination data from icddr,b Record Keeping System.

FIGURE 4 — FIRST GENERATION: SELECTED ROBUSTNESS CHECKS



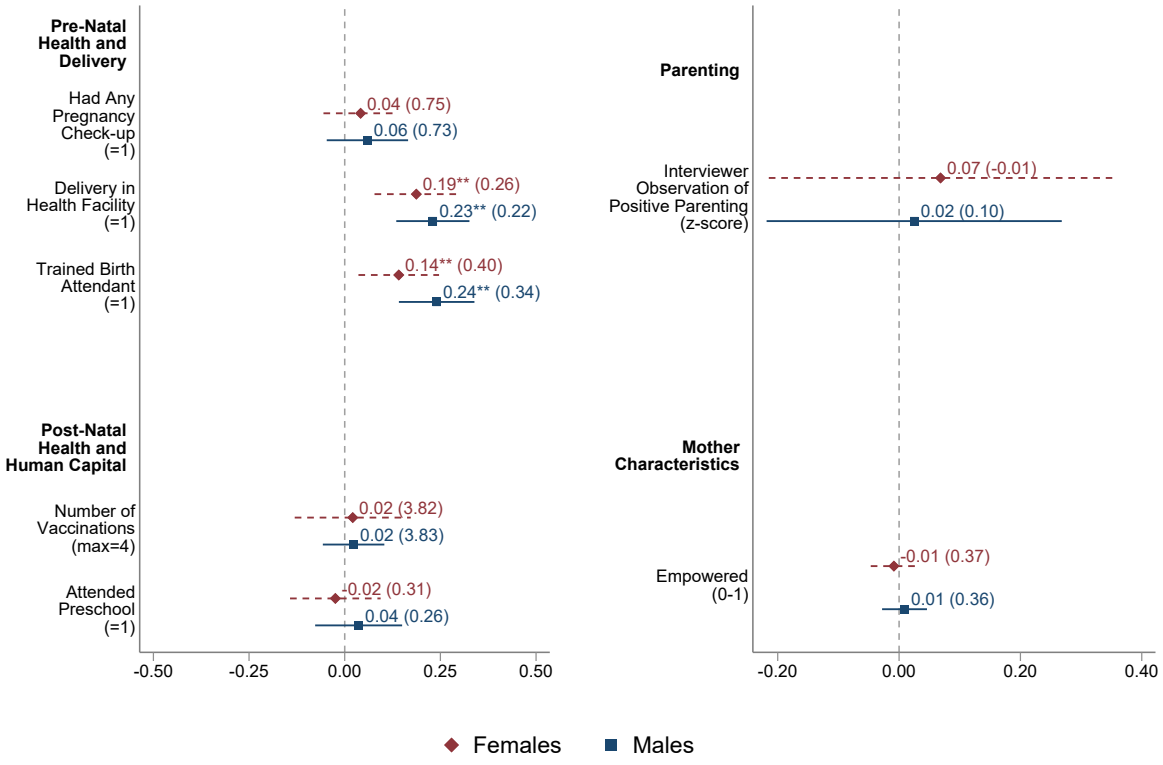
Notes: This figure shows ITT effects on first generation human capital are robust to alternate specifications. Thinner bars represent the 95% confidence interval while thicker bars represent the 90% confidence interval. *Main Spec* shows our preferred model for context. *Add Extended Ctrls* includes additional controls to account for changes over time in the supply of health and education. *Unbalanced Ctrls* limits the pre-program baseline control set to only those that are unbalanced. *Muslims Only* limits the sample to Muslims which are overrepresented in the comparison area. *Unweighted* does not reweight the sample for attrition between 1974 and 2012. *Village FE* adds fixed effects for pre-program village to account for village-specific differences. See Appendix D and Appendix E for a full description of each model and additional robustness checks.

FIGURE 5 — SECOND GENERATION: SELECTED ROBUSTNESS CHECKS



Notes: This figure shows ITT effects on second generation human capital are robust to alternate specifications. Child Development Index (left, green) is available for children aged 0-6 while Cognition Index (right, yellow) is only available children aged 7-14 (see Appendix B.3 for a description of indexes). *Main Spec* shows our preferred model for context. *Add Extended Ctrls* includes additional controls to account for changes over time in the supply of health and education. *MNCH Eligibility* adds a control for the 2007 Maternal, Neonatal and Child Health Project (MNCH). *Unbalanced Ctrls* limits the pre-program baseline control set to only those that are unbalanced. *Muslims Only* limits the sample to Muslims which are overrepresented in the comparison area. *Unweighted* does not reweight the sample for attrition between 1974 and 2012. See Appendix D and Appendix E for a full description of each model and additional robustness checks.

FIGURE 6 — SECOND GENERATION: POTENTIAL MECHANISMS



Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This figures shows the ITT effects for firstborn children born to 1982-1988 mothers from Equation 3, separately by males and females. The mean of the comparison group is displayed in parentheses and the bars represent 95% confidence intervals. Delivery in Health Facility equals one if the pregnancy was delivered in a certified medical center. Trained Birth Attendant equals one if the delivery was performed by a physician, nurse or other healthcare worker. Vaccinations is the total types of vaccinations received out of four (BCG, Polio, DPT and MMR). Interviewer Observation of Positive Parenting is the enumerator's measure of how positive the parent's interactions with their children were. Z-scores were calculated using the comparison group mean and standard deviation by age and sex. Empowerment is an index ranging 0 to 1 of survey questions related to women's decision making power, attitudes toward gender equivalence in social issues, attitudes toward husband violence and women's mobility. See Appendix B for a fuller description of each mechanism.

TABLE 1 — FIRST GENERATION: BALANCE OF 1982-1988 COHORT'S PRE-PROGRAM CHARACTERISTICS

	Treatment Area		Comparison Area		Difference		
	Mean	SD	Mean	SD	Mean	T-stat	Mean/SD
Individual Characteristics							
Male (=1)	0.51	0.46	0.50	0.46	0.01	0.61	0.02
Birth year	1984.97	1.76	1984.93	1.89	0.03	0.38	0.01
Islamic (=1)	0.84	0.81	0.95	0.31	-0.12	-3.81	-0.14
Mother's height (cm)	149.47	14.33	149.29	6.47	0.18	0.29	0.01
Father's height (cm)	160.91	8.46	160.68	8.09	0.22	0.45	0.02
Mother's years of education	1.83	3.30	1.61	3.73	0.22	1.18	0.04
Father's years of education	3.79	4.31	3.29	5.55	0.50	1.74	0.07
Household Characteristics							
HH Bari size	8.63	10.25	8.03	10.40	0.61	1.18	0.04
HH Family size	6.79	3.51	6.54	3.08	0.25	1.53	0.05
Latrine (=1)	0.81	0.67	0.88	0.73	-0.07	-2.03	-0.07
Owns a lamp (=1)	0.62	0.65	0.59	0.74	0.03	0.92	0.03
Owns a watch (=1)	0.14	0.42	0.15	0.45	-0.01	-0.41	-0.01
Owns a radio (=1)	0.08	0.34	0.07	0.28	0.01	0.35	0.01
Number of cows	1.41	2.25	1.33	1.98	0.08	0.71	0.03
Number of boats	0.65	0.95	0.65	0.99	-0.01	-0.16	-0.01
Wall tin or tinmix (=1)	0.29	0.56	0.29	0.60	-0.01	-0.23	-0.01
Tin roof (=1)	0.82	0.50	0.83	0.54	-0.01	-0.40	-0.01
Number of rooms per capita	0.22	0.11	0.22	0.12	0.00	0.80	0.03
Drinking water, tubewell (=1)	0.30	0.88	0.16	0.74	0.15	3.59	0.13
Drinking water, tank (=1)	0.40	1.07	0.33	1.31	0.07	1.11	0.04
HH age	48.20	16.84	45.93	16.35	2.27	2.75	0.10
HH <2 years of education (=1)	0.60	0.56	0.60	0.77	0.00	0.00	0.00
HH works in agriculture (=1)	0.61	0.67	0.56	0.73	0.04	1.21	0.04
HH works in fishing (=1)	0.07	0.47	0.07	0.35	0.00	-0.08	0.00
HH spouse's age	37.50	15.62	35.62	15.03	1.88	2.46	0.09
HH spouse <2 years of education (=1)	0.86	0.43	0.87	0.48	-0.01	-0.59	-0.02
1982 Land size	10.02	19.28	10.99	19.81	-0.97	-1.00	-0.04
Obs	742		896				

Notes: This table shows balance of pre-program individual and household characteristics between treatment and comparison areas. The sample includes male and female respondents in the 1982-1988 age cohort who have education data in MHSS2. There are 1,431 and observations with non-missing mother's height and education, and 1,119 with non-missing father's height and education. Unless otherwise noted, household characteristics come from the 1974 census. Standard deviations (SD) are clustered at the treatment village level. Household head and spouse age are reported, but these variables are likely affected by the family planning program which increased birth intervals and decreased family sizes in the treatment area Barham et al. (2021). Observations are weighted to correct for attrition between birth and the 2012 MHSS2 survey.

TABLE 2 — FIRST GENERATION: ITT EFFECTS ON HUMAN CAPITAL

	Pooled				Males				Females			
	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treat*(1982-1988)	0.968	-0.041	-0.011	0.419	1.047	-0.037	0.061	0.960	0.999	-0.042	-0.073	-0.098
	(0.342)**	(0.022)+	(0.049)	(0.206)*	(0.629)+	(0.029)	(0.072)	(0.377)*	(0.460)*	(0.034)	(0.066)	(0.239)
Treat	0.034	0.021	0.018	0.029	-0.022	0.023	-0.026	-0.049	-0.011	0.019	0.063	0.084
	(0.280)	(0.018)	(0.043)	(0.159)	(0.421)	(0.022)	(0.057)	(0.262)	(0.337)	(0.025)	(0.046)	(0.164)
Rand. Inf. P-value	0.03	0.03	0.77	0.09	0.25	0.13	0.45	0.11	0.03	0.11	0.23	0.85
1982-1988 Mean	157.32	0.09		7.11	163.40	0.08		7.00	151.31	0.11		7.23
Obs	4213	4213	4039	4340	1893	1893	1723	1952	2320	2320	2316	2388

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows double difference ITT effects on the stock of human capital of the first generation. Standard errors are clustered at the pre-program village level and reported in parentheses. Random inference p-values are calculated for 1982-1988 cohort effects from a distribution of test statistics constructed by reassigning treatment status to villages over 10,000 permutations while maintaining geographic contiguity. 1982-1988 means are for the comparison group. All regressions include individual and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Individual characteristics include year of birth fixed effects, age cohort fixed effects, and religion. Pre-intervention characteristics include all characteristics in Table 1. Short stature is defined as height less than 155cm for males and 145cm for females. The Cognition Index includes MMSE, Digit Spans Forward and Backward, and Ravens Progressive Matrix scores. Male Cognition Index sample sizes are smaller than female because tests were not administered in the phone survey.

TABLE 3 — FIRST GENERATION: ITT EFFECTS ON COGNITION INDEX AND ITS COMPONENTS

	Index (z-score)	MMSE (z-score)	Digit Spans (z-score)	Ravens (z-score)
	(1)	(2)	(3)	(4)
Panel A: Pooled				
Treat*(1982-1988)	-0.011	-0.033	0.003	-0.006
	(0.049)	(0.065)	(0.070)	(0.064)
Treat	0.018	0.026	0.020	0.009
	(0.043)	(0.059)	(0.056)	(0.048)
Obs	4039	4034	4036	4014
Panel B: Males				
Treat*(1982-1988)	0.061	0.068	0.027	0.067
	(0.072)	(0.110)	(0.094)	(0.104)
Treat	-0.026	-0.015	-0.003	-0.045
	(0.057)	(0.090)	(0.073)	(0.069)
Obs	1723	1722	1721	1714
Panel C: Females				
Treat*(1982-1988)	-0.073	-0.130	-0.019	-0.062
	(0.066)	$(0.074)^{+}$	(0.095)	(0.095)
Treat	0.063	0.074	0.054	0.053
	(0.046)	(0.062)	(0.062)	(0.056)
Obs	2316	2312	2315	2300

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows double difference ITT effects on the components of the Cognition Index of the first generation. Standard errors are clustered at the treatment village level. All regressions include individual characteristics and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Individual characteristics include year of birth fixed effects, age cohort fixed effects and controls for religion. Pre-intervention characteristics include all individual and household characteristics in Table 1. Cognition sample sizes are smaller for men because no cognitive tests, with the exception of digit span forward, were administered in the phone survey. All females in the sample were interviewed in person.

TABLE 4 — FIRST GENERATION: ITT EFFECTS ON MARRIAGE, FERTILITY AND CHILD MORTALITY

	All Females					Females with ≥1 Live Birth		
	Married (=1)	Age at Menarche	Number of Live Births	Lost Pregnancy (=1)	Firstborn Died (=1)	Male Firstborn (=1)	Age at First Birth	Has Secondborn (=1)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: All 1982-1988 Females								
Treat	0.002	-0.055	-0.039	-0.001	-0.014	-0.006	-0.065	-0.019
	(0.017)	(0.072)	(0.065)	(0.031)	(0.012)	(0.044)	(0.245)	(0.038)
Mean	0.932	13.462	1.516	0.173	0.045	0.505	21.535	0.559
Obs	811	805	811	811	811	677	677	677
Panel B: 1982-1988 Mothers With Male Firstborn								
Treat		-0.080	-0.010	0.046	-0.012		-0.172	0.041
		(0.101)	(0.078)	(0.049)	(0.026)		(0.343)	(0.056)
Mean		13.428	1.766	0.186	0.080		21.612	0.548
Obs		340	341	341	341		341	341
Panel C: 1982-1988 Mothers With Female Firstborn								
Treat		-0.013	-0.114	-0.035	-0.007		-0.041	-0.059
		(0.126)	(0.108)	(0.038)	(0.018)		(0.391)	(0.064)
Mean		13.429	1.853	0.174	0.027		21.457	0.571
Obs		334	336	336	336		336	336

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects on marriage, fertility and child mortality outcomes of first generation females in the 1982-1988 cohort. Standard errors are clustered at the pre-program village level. Means are for the comparison group. All regressions include individual characteristics and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Individual characteristics include year of birth fixed effects and religion. Pre-intervention characteristics include all individual and household characteristics listed in Table 1.

TABLE 5 — SECOND GENERATION: ITT EFFECTS ON HUMAN CAPITAL

	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)
	Ages 0-14		Ages 0-6	Ages 7-14
	(1)	(2)	(3)	(4)
Panel A: Pooled				
Mom Treat	0.106	-0.044	-0.036	0.004
	(0.103)	(0.035)	(0.108)	(0.098)
Rand. Inf. P-value	0.243	0.192	0.910	0.958
Mean	-1.262	0.258		
Obs	629	629	334	262
Panel B: Males				
Mom Treat	-0.022	-0.002	-0.106	-0.198
	(0.157)	(0.062)	(0.145)	(0.144)
Rand. Inf. P-value	0.867	0.974	0.678	0.133
Mean	-1.221	0.242		
Obs	311	311	159	135
Panel C: Females				
Mom Treat	0.335	-0.139	0.048	0.268
	(0.146)*	(0.048)**	(0.136)	(0.145)+
Rand. Inf. P-value	0.008	0.072	0.815	0.048
Mean	-1.302	0.273		
Obs	318	318	175	127

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects on the stock of human capital of second generation children of females in the 1982-1988 cohort using mother's treatment assignment. Standard errors in parentheses are clustered at the mother's pre-program village level. Random inference p-values are calculated from a distribution of test statistics constructed by reassigning mother's treatment status to villages over 10,000 permutations while maintaining geographic contiguity. Means are for children of comparison group mothers. All regressions include mother's individual and pre-intervention characteristics and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Height-for-age is calculated using WHO standards. Stunted equals one if the child's height-for-age z-score is less than -2. Child Development and Cognition indexes are internally standardized by sex and age using the comparison group's mean and standard deviation. Child Development Index is the average of the standardized language, fine motor, and gross motor Denver scores. Cognition Index is the average of standardized MMSE, memory, digit spans, Ravens, and matching test scores. All models include child's six-month age fixed effects, mother's year of birth fixed effects, mother's religion and mother's pre-intervention characteristics listed in Table A3.

TABLE 6 — FIRST GENERATION: HETEROGENEOUS ITT EFFECTS BY MOTHER'S HEIGHT

	Pooled				Males				Females			
	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treat*(1982-1988)*Shortest Tercile	1.850	-0.091	0.081	1.111	1.998	-0.070	0.202	1.880	1.927	-0.109	-0.051	0.381
	(0.501)**	(0.034)**	(0.068)	(0.340)**	(0.784)*	(0.042)	(0.102)*	(0.570)**	(0.711)**	(0.052)*	(0.095)	(0.374)
Treat*(1982-1988)*Middle Tercile	0.501	-0.020	-0.030	-0.061	0.496	-0.018	-0.023	0.391	0.555	-0.010	-0.019	-0.537
	(0.460)	(0.028)	(0.075)	(0.290)	(0.756)	(0.037)	(0.112)	(0.461)	(0.643)	(0.042)	(0.101)	(0.384)
Treat*(1982-1988)*Tallest Tercile	0.970	-0.015	-0.064	0.048	0.975	-0.013	0.051	0.526	0.762	-0.010	-0.162	-0.383
	(0.554)+	(0.023)	(0.075)	(0.348)	(0.865)	(0.030)	(0.100)	(0.590)	(0.774)	(0.038)	(0.095)+	(0.430)
Treat	0.002	0.022	0.018	0.030	-0.063	0.024	-0.025	-0.049	-0.032	0.020	0.062	0.087
	(0.279)	(0.018)	(0.043)	(0.158)	(0.426)	(0.022)	(0.058)	(0.262)	(0.336)	(0.025)	(0.046)	(0.164)
P(Shortest Tercile = Tallest Tercile)	0.184	0.029	0.120	0.026	0.228	0.200	0.241	0.077	0.258	0.071	0.378	0.150
1982-1988 Means												
Shortest Tercile	154.68	0.18		6.64	160.75	0.15		6.42	148.87	0.21		6.86
Middle Tercile	158.14	0.04		7.14	163.72	0.04		6.96	151.88	0.05		7.35
Tallest Tercile	160.52	0.02		7.91	167.41	0.01		7.91	154.29	0.03		7.91
Obs	4122	4122	3954	4249	1847	1847	1683	1906	2275	2275	2271	2343

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows double difference ITT effects on the stock of human capital, breaking the 1982-1988 cohort effects by terciles of their mother's height. Terciles are defined by a mother's height less than 147.9cm (shortest), 147.9-152.7cm (middle) or greater than 152.7cm (tallest). 55 percent of respondents in the shortest tercile had a mother of short stature (height less than 145cm). P(Shortest Tercile = Tallest Tercile) is the p-value from a two-sided test that the treatment effects among the shortest and tallest tercile are equal. Tercile means are the comparison group. All regressions include individual and pre-intervention characteristics and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Standard errors are clustered at the pre-program village level and reported in parentheses.

TABLE 7 — SECOND GENERATION: HETEROGENEOUS ITT EFFECTS BY MATERNAL GRANDMOTHER'S HEIGHT

	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)
	Ages 0-14		Ages 0-6	Ages 7-14
	(1)	(2)	(3)	(4)
Panel A: Pooled				
Mom Treat*Shortest Tercile	0.031	-0.002	-0.014	-0.157
	(0.135)	(0.049)	(0.150)	(0.121)
Mom Treat*Middle Tercile	0.074	0.036	-0.227	0.080
	(0.167)	(0.062)	(0.147)	(0.136)
Mom Treat*Tallest Tercile	0.371	-0.202	0.113	0.140
	(0.160)*	(0.060)**	(0.132)	(0.161)
P(Shortest Tercile = Tallest Tercile)	0.051	0.004	0.506	0.045
Means				
Shortest Tercile	-1.414	0.280		
Middle Tercile	-1.207	0.227		
Tallest Tercile	-1.112	0.259		
Obs	594	594	315	247
Panel B: Males				
Mom Treat*Shortest Tercile	0.110	0.007	0.017	-0.221
	(0.195)	(0.080)	(0.187)	(0.217)
Mom Treat*Middle Tercile	-0.004	0.034	-0.289	-0.031
	(0.269)	(0.110)	(0.267)	(0.196)
Mom Treat*Tallest Tercile	0.091	-0.146	-0.171	-0.262
	(0.240)	(0.111)	(0.239)	(0.254)
P(Shortest Tercile = Tallest Tercile)	0.933	0.207	0.518	0.879
Means				
Shortest Tercile	-1.535	0.308		
Middle Tercile	-1.032	0.170		
Tallest Tercile	-1.101	0.278		
Obs	292	292	150	125
Panel C: Females				
Mom Treat*Shortest Tercile	0.246	-0.122	-0.019	0.062
	(0.242)	(0.079)	(0.166)	(0.208)
Mom Treat*Middle Tercile	0.205	0.023	-0.123	0.119
	(0.234)	(0.091)	(0.172)	(0.132)
Mom Treat*Tallest Tercile	0.541	-0.276	0.186	0.708
	(0.188)**	(0.062)**	(0.162)	(0.199)**
P(Shortest Tercile = Tallest Tercile)	0.332	0.137	0.356	0.001
Means				
Shortest Tercile	-1.283	0.250		
Middle Tercile	-1.369	0.281		
Tallest Tercile	-1.121	0.244		
Obs	302	302	165	122

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects on the stock of human capital by terciles of their maternal grandmother's height. Tercile definitions from the first generation sample are used to be consistent across generations (see Table 6 for exact tercile cutoffs used). P(Shortest Tercile = Tallest Tercile) is the p-value from a two-sided test that the treatment effects among the shortest and tallest tercile are equal. Means are for children of comparison group mothers. All regressions include mother's individual and pre-intervention characteristics and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Standard errors in parentheses are clustered at the treatment village level.

ALL APPENDICES ARE FOR ONLINE PUBLICATION

Appendix A Appendix Figures and Tables

TABLE A1 — FIRST GENERATION: ITT EFFECTS ON THE METABOLIC INDEX AND COMPONENTS OF THE INDEX

		Components		
	Metabolic Syndrome Index (z-score)	Underweight (=1)	Overweight (=1)	Stage 1 Hypertension (=1)
	(1)	(2)	(3)	(4)
Panel A: Pooled				
Treat	-0.008	-0.009	0.025	-0.008
	(0.027)	(0.020)	(0.027)	(0.017)
Mean		0.185	0.324	0.120
Obs	1500	1425	1425	1492
Panel B: Males				
Treat	0.044	0.007	0.054	0.002
	(0.037)	(0.027)	(0.039)	(0.030)
Mean		0.186	0.293	0.165
Obs	696	695	695	694
Panel C: Females				
Treat	-0.052	-0.015	-0.004	-0.022
	(0.037)	(0.028)	(0.036)	(0.018)
Mean		0.183	0.354	0.080
Obs	804	730	730	798

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects on the flow of human capital of first generation people in the 1982-1988 cohort. Standard errors are clustered at the pre-program village level and reported in parentheses. Means are for the comparison group. All regressions include individual and pre-intervention characteristics and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Metabolic Syndrome is the average of the standardized components in subsequent columns. All variables are internally standardized by sex using the comparison group's mean and standard deviation. Original component indicators are shown in columns 2-4. Underweight equals one if BMI is less than 18.5 while overweight equals one if BMI is greater than 23. Stage 1 Hypertension equals one if systolic blood pressure is above 130 mm Hg or diastolic blood pressure is above 80 mm Hg. Individual characteristics include year of birth fixed effects and religion. Pre-intervention characteristics include all characteristics in Table 1.

TABLE A2 — ITT EFFECTS ON HUMAN CAPITAL FOR COHORT BORN 1977-1981

			Cognition Index & Components				
	Height (cm)	Short Stature (=1)	Index (z-score)	MMSE (z-score)	Digit Spans (z-score)	Ravens (z-score)	Grades Attained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Pooled							
Treat*(1977-1981)	-0.418	-0.028	-0.036	-0.096	-0.096	0.078	0.009
	(0.473)	(0.025)	(0.072)	(0.087)	(0.086)	(0.098)	(0.264)
Treat	0.006	0.021	0.026	0.031	0.031	0.018	0.043
	(0.277)	(0.018)	(0.043)	(0.059)	(0.056)	(0.049)	(0.159)
1977-1981 Mean	155.406	0.150					3.408
Obs	3444	3444	3334	3330	3331	3308	3566
Panel B: Males							
Treat*(1977-1981)	-0.825	0.001	-0.073	-0.121	-0.187	0.076	-0.519
	(0.697)	(0.029)	(0.105)	(0.144)	(0.135)	(0.137)	(0.483)
Treat	-0.006	0.021	-0.012	0.000	0.014	-0.035	-0.058
	(0.415)	(0.022)	(0.058)	(0.090)	(0.073)	(0.070)	(0.265)
1977-1981 Mean	162.144	0.105					4.398
Obs	1521	1521	1415	1414	1413	1405	1577
Panel C: Females							
Treat*(1977-1981)	-0.023	-0.058	0.008	-0.077	0.001	0.110	0.487
	(0.683)	(0.043)	(0.089)	(0.101)	(0.107)	(0.135)	(0.351)
Treat	-0.070	0.020	0.068	0.074	0.060	0.061	0.122
	(0.341)	(0.025)	(0.047)	(0.062)	(0.063)	(0.056)	(0.167)
1977-1981 Mean	150.089	0.186					2.625
Obs	1923	1923	1919	1916	1918	1903	1989

Notes: $^{+}\ p < 0.10$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$. This table shows double difference ITT effects on the stock human capital of first generation 1977-1981 cohort using the 1947-1969 cohort as the comparison group. Standard errors are clustered at the pre-program village level and reported in parentheses. 1977-1981 means are for the comparison group. All regressions include individual and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. Individual characteristics include year of birth fixed effects, age cohort fixed effects, and religion. Pre-intervention characteristics include all characteristics in Table 1. Short stature is defined as height less than 155cm for males and 145cm for females. Cognition sample sizes are smaller for men because no cognitive tests, with the exception of digit span forward, were collected in the phone survey. All females in the sample were interviewed in person. Math sample sizes are smaller because the math test was not administered to anyone over age 50.

TABLE A3 — SECOND GENERATION: BALANCE OF 1982-1988 COHORT MOTHERS' PRE-PROGRAM CHARACTERISTICS

	Treatment Area		Comparison Area		Difference		
	Mean	SD	Mean	SD	Mean	T-stat	Mean/SD
Mother's Characteristics							
Birth Year	1984	1.97	1984	1.69	-0.16	-1.10	-0.06
Islamic (=1)	0.85	0.59	0.95	0.27	-0.11	-2.79	-0.16
Household Characteristics							
Bari size	9.04	7.13	8.08	6.88	0.97	1.73	0.10
Family size	6.73	2.56	6.36	2.53	0.38	1.87	0.11
Asset Index (max=6)	2.83	1.62	2.77	1.68	0.06	0.47	0.03
Wall tin or tinmix (=1)	0.27	0.47	0.30	0.43	-0.04	-1.02	-0.06
Tin roof (=1)	0.81	0.45	0.82	0.46	-0.01	-0.41	-0.02
Number of rooms per capita	0.22	0.10	0.23	0.11	-0.01	-0.74	-0.04
Drinking water, tubewell (=1)	0.25	0.52	0.14	0.53	0.11	2.72	0.15
Drinking water, tank (=1)	0.44	0.84	0.33	0.90	0.11	1.59	0.09
HH age	48.42	15.30	45.32	15.58	3.11	2.53	0.14
HH <2 years of education	0.64	0.50	0.61	0.55	0.03	0.70	0.04
HH works in agriculture (=1)	0.58	0.44	0.54	0.57	0.03	0.77	0.04
HH works in fishing (=1)	0.07	0.35	0.10	0.33	-0.03	-0.94	-0.05
HH spouse's age	37.31	13.21	35.23	13.24	2.08	1.98	0.11
HH spouse <2 years of education	0.87	0.34	0.88	0.33	0.00	-0.14	-0.01
1982 Land size	9.64	13.83	10.11	15.44	-0.47	-0.40	-0.02
Obs	293		343				

Notes: The sample includes females in the 1982-1988 cohort who have a firstborn child in the second generation analytic sample. Unless otherwise noted, household characteristics come from the 1974 census. Standard deviations (SD) are clustered at the treatment village level. Household head and spouse age are reported, but these variables are likely affected by the family planning program increasing birth intervals and decreasing family size in the treatment area. Observations are weighted to correct for attrition between birth and the 2012 MHSS2 survey.

TABLE A4 — SECOND GENERATION: ITT EFFECTS ON INDEX COMPONENTS

	Child Development Index (Ages 0-6)				Cognition Index (Ages 7-14)						
	Index	Language	Fine Motor	Gross Motor	Index	MMSE	Memory	Digits Forwards	Digits Backwards	Ravens	Matching
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Pooled											
Mom Treat	-0.036	0.042	-0.005	-0.078	0.004	-0.138	0.118	0.000	0.066	0.010	-0.145
	(0.108)	(0.119)	(0.143)	(0.127)	(0.098)	(0.148)	(0.138)	(0.141)	(0.143)	(0.155)	(0.151)
Obs	334	353	348	342	262	266	263	266	266	265	266
Panel B: Males											
Mom Treat	-0.106	0.096	-0.085	-0.241	-0.198	-0.426	0.053	-0.132	-0.251	-0.143	-0.506
	(0.145)	(0.171)	(0.200)	(0.151)	(0.144)	(0.279)	(0.258)	(0.185)	(0.176)	(0.214)	(0.206)*
Obs	159	167	167	164	135	137	135	137	137	137	137
Panel C: Females											
Mom Treat	0.048	0.023	0.085	0.027	0.268	0.363	0.239	-0.083	0.353	0.313	0.307
	(0.136)	(0.152)	(0.177)	(0.186)	(0.145)+	(0.201)+	(0.200)	(0.193)	(0.213)	(0.284)	(0.282)
Obs	175	186	181	178	127	129	128	129	129	128	129

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects on components of the Child Development and Cognition indexes of second generation children of 1982-1988 cohort females using their mother's treatment assignment. Standard errors in parentheses are clustered at the mother's pre-program village level. All variables are internally standardized by sex and age (6 month bins for ages 0-6, 12 months for ages 7+) using the comparison group's mean and standard deviation. Indexes are the average of the standardized components in subsequent columns. All models include six-month age fixed effects, mother's age fixed effects, mother's religion and mother's pre-intervention characteristics listed in Table A3.

Appendix B Data and Construction of Selected Variables

This appendix describes the data sources, attrition, and the creation of the intent-to-treat and main outcome variables.

B.1 Data Sources

MHSS1/2.—The main outcomes variables used in this paper are from MHSS2. It is a large socioeconomic survey comprised of several instruments including an individual survey, a household survey, village survey, facility surveys, and market price survey of major markets areas throughout Bangladesh where MHSS2 respondents lived. The key anthropometric, cognitive, and educational outcomes were collected in the individual instrument and are not proxy reports as is the case in many surveys. Most of the data were collected during face-to-face interviews, though a subset of data was collected in a phone survey of international migrants who did not return to Bangladesh during the data collection period (about 15 percent of our male sample). The MHSS2 phone survey instrument was shorter than the in-person survey instrument, as a result, there are smaller sample sizes for some variables such as the cognitive family.

MHSS2 was conducted between 2012 and 2014 and was designed to be a panel to MHSS1 (icddr,b 1996). MHSS1 is a seven percent random subsample of household compounds (called baris) from the Matlab area living in both the treatment and comparison areas and was designed to be representative of the study area's 1996 population. In MHSS1, two households were interviewed in each bari: a primary household, selected randomly, and a secondary household, selected purposively. Within a household, individuals age six and older were randomly selected to be personally interviewed, but basic information, including education, was collected on all household members via proxy.

The MHSS2 sample includes all individuals selected for personal interview in MHSS1 primary households creating panel data for these individuals.[25] To limit migration selection for key age groups, the MHSS2 sample also includes individuals born between 1972 and 1989 to a MHSS1 primary household that had migrated out of Matlab between 1977 and 1996 (referred to as pre-1996 migrants).[26] To the extent that a whole household and lineage migrated out of Matlab between the start of the program and 1996, leaving no one in that lineage available for selection into the MHSS1 sample, the MHSS2 sample could still suffer from migration selection. It is rare that whole households and lines of descent migrated out of Matlab prior to 1996 and is estimated to be minimal at 2.6 percent of the study site. In addition, we test the balance of the treatment groups to check that the treatment and comparison group have similar baseline characteristics on average.

With the exception of the phone survey, all tests were collected on the individuals in their home by well-trained testers. Testers were extensively trained to implement protocols in a similar fashion. Retraining took place a couple of times throughout the survey period to restandardize the testers. The testers were generally female, though there were a few male testers who implemented the tests on adult male respondents. For the most part, testers were only responsible for implementing the testing part of the survey (book 6) and did not do other parts of the survey. There were a few enumerators who were trained on the entire survey including book 6 to interview adult migrants who were living on their own.

Census Data.—Periodic censuses were collected for all individuals in the study area (treatment and comparison areas) by iccdr,b. These data typically include household location, household characteristics and composition, employment, education, and assets. We obtain pre-program individual and household data on the analysis sample from the 1974 census (icddr,b 1974) and use these data to test for differences

25. MHSS2 sample also included all panel member descendants, and their co-resident spouses. However, among those members who had migrated out of Matlab, spouses were tracked for interview.

26. The pre-1996 migrants were identified by using the detailed DSS data.

in baseline characteristics between the treatment and comparison areas. We also use the 1974 and 1982 census (icddr,b 1982) to link individuals to the study area (which is the demographic surveillance site) before 1977 to construct an individual's intent-to-treat status (see section C below).

Demographic Surveillance Site (DSS) Data.—Vital registration data provide prospective tracking of every birth, death, marriage, divorce, and in- and out- migration occurring in the study area. As such, we know when someone enters and leaves the study area. Information on migration destination (rural, urban, international) is also available starting in 1982. Data were collected by icddr,b and are high quality in part because they were collected so frequently: every two weeks until 1997, every month between 1998 and 2006, and every two months between 2007 and MHSS2.These data include pre-program data from 1974 onwards, and are used to construct birth dates and an individual's intent-to-treat status. In addition, we use these data to construct pre-program migration network variables for each individual in the analysis sample, as well as, out-migration variables such as whether someone has ever migrated, and out-migration variables for years not covered in the MHSS2 migration history.

B.2 Attrition

Relative to other long-term follow-up surveys, attrition in the MHSS2 sample is small. Table B2 documents attrition rates for our follow-up sample for those who have information on height and the MMSE. This sample includes all individuals born during the experimental period from October 1977 and December 1988 (the 24–30 and 31–34 cohorts) or prior to the intervention from 1947-1969 (the 44-65 cohort) and who were a member of a MHSS1 primary household or a pre-1996 migrant. Including death and any other type of non-response, the attrition rate at the household level is 7 percent. Attrition rates are slightly higher for variables from the individual survey at 10 percent for men and 7 percent for females for height information (Table B2). The low attrition rate is a result of a carefully designed tracking protocol. Migrants were tracked all over Bangladesh, and a rapid response system was developed that allowed trackers in Matlab to connect enumerators placed in different parts of the country with respondents who had left Matlab. Intensive interviewing took place during all the Eid holidays from 2012–2014. Survey teams targeted international migrants, far away domestic migrants, and hard-to-track migrants returning to Matlab for the holiday. Finally, a phone survey was employed to collect information on a subset of questions from the main survey from predominately international migrants who did not return to Bangladesh during the survey period. While there is a limited set of variables available for this group, most anthropometric and educational outcomes used in this study were collected during the phone survey. Self-reported height was collected in the phone survey, but the MMSE was not part of the phone survey, so there is less attrition on height than for the MMSE. For the overall sample, we were able to survey 96.4 percent of living men and 96.8 percent of living females. When including the deceased in these calculations, follow-up rates for those for whom there is data on height falls to 86 percent for men and 91.2 percent for females. Without the phone survey, the attrition rate is higher for men at almost 22.7 percent, but the same for females, because females do not migrate internationally for work. Restricting to non-missing MMSE information produces similar attrition rates as excluding people surveyed by phone since the MMSE was not part of the phone questionnaire.

Even though the attrition rates are low and not statistically different between treatment and comparison area, there could still be differential attrition between the treatment and comparison area, potentially biasing the results. Individuals were considered surveyed in our sample if the respondent had complete information on height. Table B1 presents the means of individual and household characteristics by our attrition and surveyed sample for all the men and females in the sample together. Individual and household characteristics are well balanced between those who were surveyed and those who were lost to attrition. The attrited are more likely to be male and older which is not surprising given men migrate for

work and the attrited group includes the dead. Results are similar if we split them by sex or by the three cohorts of interest, and if we include the phone survey group as being in the attrition group rather than the surveyed group. For the second generation, attrition in similarly low. Of the 677 females aged 24-30 with a firstborn child, 629 were surveyed and had a valid height measurement. Only 47 children were lost (6.92% attrition rate), 4 of which had invalid height measurements, 29 died before being surveyed and 14 were born to migrant females whose children were not followed. The attrition rate did not differ between treatment and comparison areas by any definition or among males and females separately.

TABLE B1 — FIRST GENERATION: ATTRITION BALANCE OF 1982-1988 COHORT'S PRE-PROGRAM CHARACTERISTICS

	Attrited		Surveyed		Difference		
	Mean	SD	Mean	SD	Mean	T-stat	Mean/SD
Individual Characteristics							
Male (=1)	0.54	0.48	0.50	0.46	0.04	0.96	0.06
Birth year	1984	2.20	1984	1.84	-0.12	-0.68	-0.04
Islamic (=1)	0.93	0.28	0.90	0.59	0.02	1.03	0.03
Mother's height (cm)	150.27	5.66	149.36	10.29	0.91	1.63	0.08
Father's height (cm)	160.88	7.03	160.79	8.41	0.10	0.13	0.01
Mother's years of education	1.45	2.31	1.70	3.61	-0.25	-1.14	-0.06
Father's years of education	3.13	4.04	3.48	5.23	-0.35	-0.93	-0.05
Household Characteristics							
HH Bari size	9.27	7.37	8.27	10.21	1.00	2.10	0.08
HH Family size	6.84	3.07	6.65	3.43	0.18	0.76	0.04
Latrine (=1)	0.80	0.40	0.85	0.72	-0.05	-1.55	-0.06
Owns a lamp (=1)	0.57	0.53	0.61	0.71	-0.04	-1.02	-0.04
Owns a watch (=1)	0.13	0.32	0.15	0.45	-0.02	-0.89	-0.04
Owns a radio (=1)	0.04	0.18	0.07	0.30	-0.03	-2.11	-0.09
Number of cows	1.33	1.98	1.37	2.12	-0.04	-0.28	-0.01
Number of boats	0.61	0.66	0.65	0.98	-0.04	-0.81	-0.03
Wall tin or tinmix (=1)	0.26	0.40	0.29	0.59	-0.03	-0.98	-0.04
Tin roof (=1)	0.77	0.41	0.83	0.53	-0.06	-1.71	-0.08
Number of rooms per capita	0.21	0.09	0.22	0.12	-0.01	-0.75	-0.04
Drinking water, tubewell (=1)	0.22	0.51	0.22	0.88	0.00	0.07	0.00
Drinking water, tank (=1)	0.36	0.56	0.36	1.24	0.00	0.05	0.00
HH age	45.79	14.36	46.87	17.58	-1.08	-0.96	-0.05
HH <2 years of education (=1)	0.65	0.49	0.60	0.69	0.05	1.44	0.06
HH works in agriculture (=1)	0.61	0.54	0.58	0.72	0.03	0.65	0.03
HH works in fishing (=1)	0.12	0.44	0.07	0.39	0.05	1.60	0.09
HH spouse's age	36.33	10.90	36.39	15.70	-0.06	-0.07	0.00
HH spouse <2 years of education (=1)	0.89	0.31	0.87	0.46	0.02	0.83	0.03
1982 Land size	9.97	15.54	10.58	19.98	-0.62	-0.56	-0.02
Obs	170		1630				

Notes: The sample includes male and female respondents in the 1982-1988 cohort. Attrition is defined as missing height information prior to dropping outliers. Standard deviations are clustered at the village level. Attrition is also balanced for the 1947-1969 cohort (results not shown).

TABLE B2 — FIRST GENERATION: 2012 ATTRITION RATES OF 1982-1988 COHORT

	Pooled			Males			Females		
		T-C Difference			T-C Difference			T-C Difference	
	Attrition Rate	Mean	(SE)	Attrition Rate	Mean	(SE)	Attrition Rate	Mean	(SE)
Not found or refused	5.61%	-0.007	(0.013)	5.11%	0.008	(0.018)	6.13%	-0.023	(0.016)
Not found, refused, or dead	7.22%	-0.008	(0.014)	6.96%	0.005	(0.019)	7.49%	-0.020	(0.017)
Non-missing height information	9.44%	-0.003	(0.015)	10.01%	0.011	(0.020)	8.85%	-0.017	(0.019)
Non-missing MMSE information	17.00%	-0.012	(0.017)	24.59%	-0.013	(0.029)	9.08%	-0.012	(0.019)
Non-missing height, no phone survey	16.78%	-0.006	(0.017)	24.37%	0.004	(0.028)	8.85%	-0.017	(0.019)

Notes: Sample includes 1982-1988 cohort. The Attrition Rate column displays the percent of the sample who attrited, while mean and SE show the difference and standard error of the difference in attrition rates by treatment and comparison area. The standard error of the difference is clustered at the pre-program village level. There are 919 men and 881 females across the two cohorts in the sample frame. Missing height information indicates the respondent was not able to be measured, including not found, refused, or dead, but outliers are not dropped here as they are in the final analysis.

TABLE B3 — SECOND GENERATION: 2012 ATTRITION RATES

	Pooled				Males				Females			
	Number Attrited	Attrition Rate	T-C Difference Mean	(SE)	Number Attrited	Attrition Rate	T-C Difference Mean	(SE)	Number Attrited	Attrition Rate	T-C Difference Mean	(SE)
Child missing height	47	6.92%	-0.037	(0.020)	29	8.48%	-0.060	(0.030)	18	5.34%	-0.014	(0.025)
Child not found	43	6.33%	-0.032	(0.019)	28	8.19%	-0.054	(0.030)	15	4.45%	-0.010	(0.023)
Child died	29	4.27%	-0.024	(0.016)	22	6.43%	-0.034	(0.027)	7	2.08%	-0.014	(0.016)
Mother migrated	14	2.06%	-0.008	(0.011)	6	1.75%	-0.020	(0.014)	8	2.37%	0.004	(0.017)

Notes: Sample includes live firstborn children of 1982-1988 females. Number Attrited is the number of observations in the sample who attrited, while Attrition Rate displays the percent of the sample who attrited, and mean and SE show the difference and standard error of the difference in attrition rates by treatment and comparison area. The standard error of the difference is clustered at the pre-program village level. There are 342 males and 337 females in the sample frame. Missing height information indicates the respondent was not able to be measured, including not found, refused, or dead, but outliers are not dropped here as they are in the final analysis.

TABLE B4 — SECOND GENERATION: SKILLED DELIVERY LOCATIONS AND TRAINED DELIVERY WORKER DEFINITIONS

Panel A: Skilled Delivery Locations
Public hospital/District hospital/Maternal and Child Welfare Center/Public medical college hospital
Thana health center
Family welfare center
Satellite or community clinic (Government)
icddr,b hospital, sub-center, or fixed site clinic/CHRW's house
NGO clinic
Private hospital/private clinic/private medical college hospital
Private doctor's chamber/clinic (MBBS)
Panel B: Trained Delivery Workers
MBBS physician (government, icddr,b, NGO, or Private)
Family Welfare Assistant/Family Welfare Visitor/Health Assistant
Nurse
icddr,b or NGO health worker
Trained midwife

Notes: MBBS=Bachelor of Medicine, Bachelor of Surgery. Thana=local district. CHRW=community health research worker.

B.3 Construction of Selected Variables

B.3.1 First Generation

Height & Short Stature.—Adult respondents were measured standing using a Shorr height board. If the respondent was interviewed by phone, their height was self-reported. We set height values below the bottom half percentile or above the top half percentile to missing, separately for males and females. We defined an individuals short stature if their height was less than 155cm for males or 145cm for females.

Cognition Index.—This index was created by taking the average of standardized MMSE, Digit Spans Forward and Backward, and Ravens scores. All three components were internally standardized using the comparison group mean and standard deviation by sex.

Metabolic Syndrome.—This index was created by taking the average of standardized underweight, overweight, and stage 1 hypertension. Respondents were considered underweight if their BMI was below 18.5 kg/m^2 and overweight if their BMI was above 23 kg/m^2, following the WHO recommendation that for Asian populations a cutoff of 23 is more appropriate than the 25 used for the US population. If a respondent's BMI was above the 99.5 percentile or below the 0.5 percentile for her age, they were dropped from the sample. Height was measured using the Shorr height board. Weight was measured using the SECA881 U digital scale (150kg maximum and 0.01kg increments). Three measurements of blood pressure were taken and then averaged. Stage 1 hypertension is defined as a systolic blood pressure above 130 mm Hg or a diastolic blood pressure above 85 mm Hg. For outliers, if systolic blood pressure was less than 60 or more than 250 it was set to missing. Similarly, if diastolic blood pressure was less than 40 or more than 150 it was set to missing. Blood pressure was measured using the Lifesource 767-PV automatic blood pressure device. All three components were internally standardized using the comparison group mean and standard deviation by sex.

Skilled Delivery Location and Trained Delivery.—Mothers of second generation children in the analytic sample were asked where the child was delivered. Panel A of Table B4 displays the locations that were considered adequate for delivering a birth in a safe manner. Mothers were also asked who assisted in the delivery. Panel B of Table B4 displays the types of workers that were believed to have proper training to safely deliver the baby.

Mothers' Height Terciles.—We calculated terciles of mothers' MHSS2 height (using their measurement from MHSS1 if their MHSS2 height was missing) across the entire sample. First generation respondents were in the lowest tercile if their mother's height was below 147.9cm, in the middle tercile if between 147.9 and 152.7cm and in the highest tercile if above 152.7cm. 55 percent of first generation respondents in our analytic sample had a mother of short stature (height less than 145cm). We used the same definition of terciles in the second generation to maintain consistency.

B.3.2 Second Generation

Height-for-age and Stunting.—Second generation children were measured using the Shorr height board lying down if the child was less than two years old or shorter than 83cm and standing up if older than two or taller than 83cm. We calculated standardized height-for-age using WHO growth standards. Any values more than five standard deviations from the mean were set to missing. Stunting was defined as a height-for-age of two standard deviations or more below the mean.[27]

Child Development Index.—For children aged 0-6, we construct a Child Development Index by taking the average of standardized Denver Language, Fine Motor and Gross Motor skills. All components were internally standardized using the comparison group mean and standard deviation by sex and age in years.

Cognition Index.—For children aged 7-14, we construct a Cognition Index by taking the average of standardized MMSE, Memory score, Digit Spans Forward and Backward, Ravens, and Matching score. All components were internally standardized using the comparison group mean and standard deviation by sex and age in years.

Vaccinations.—BCG, DPT/Penta, Polio and MMR vaccine histories for children were recorded by questioning the mother or caretaker. We calculate the number of vaccination types (out of four) the child received.

Interviewer Observation of Positive Parenting.—For all ages, the enumerator observed how the parent or caretaker treated the child at hand. Eight binary responses were recorded to questions such as "Caretaker uses some term of endearment [...] for [child's name]..." and "Did caretaker hit child?". All responses were recoded so that an affirmative represented a qualitatively positive outcome. Scores were summed (for a possible maximum of 8) and internally standardized.

Mother's Empowerment Index.—We construct an index measuring the mother's relative empowerment from four subindicies: decision making, gender equivalence in social issues, attitudes towards husband violence, and mobility. All measures are scored or recoded so that a larger value represents more

27. Guideline: Assessing and Managing Children at Primary Health-Care Facilities to Prevent Overweight and Obesity in the Context of the Double Burden of Malnutrition: Updates for the Integrated Management of Childhood Illness (IMCI). Geneva: World Health Organization; 2017. Table 1, World Health Organization (WHO) classification of nutritional status of infants and children. Available from: https://www.ncbi.nlm.nih.gov/books/NBK487900/table/fm.s1.t1/

empowerment of the woman. Each subindex and its components are described below.

Decision Making.—Married or previously married women were asked who had the final say in four categories: major household purchases, decisions about spousal health and treatment, decisions about their own health and treatment, and visits to family or relatives. We coded each response with 1 if the woman made decisions independently or jointly about the matter and 0 if someone else had the final say. Finally, we created an index by averaging the four responses.

Gender Equivalence in Social Issues.—Women were asked to give an individual opinion on six claims: (1) women should practice purdah (female seclusion), (2) girls should be educated as much as boys, (3) women should be allowed to go out alone, (4) women should have the right to initiate divorce, (5) daughters should support elderly parents and (6) husband and wife should jointly agree to have children. We coded responses to (1) as 0=agree, 0.5=partially agree and 1=do not agree while responses to (2)-(6) were coded in reverse so that a larger value represents a higher opinion of gender equivalence. We then took the average of all six responses to form an index of gender equivalence in social issues.

Attitudes Toward Husband Violence.—Women were asked whether a husband is justified in hurting his wife if he is ever angered or annoyed by his wife's behavior in four situations: (1) going out without telling her him, (2) disobeying family elders, (3) neglecting their children and (4) refusing to have sex. We rcoded responses so that yes=0 and no=1 to represent a more egalitarian viewpoint and averaged her four responses to create an index of attitudes toward husband violence.

Women's Mobility.—Women were asked if in the past twelve months they had gone out alone or with someone else for five activities: (1) visiting women in other baris, (2) visiting someone outside their village, (3) visiting any place on public transport, (4) visiting the local store for small purchases (oil, rice, salt, etc.) and (5) visiting a store or market to make larger purchases such as clothing. We coded each response as 1 if they went alone, 0.5 if they went out with someone and 0 if they never went out. Then we averaged responses to create an index that represents more mobility.

Appendix C Potential Confounders

We accounted for several other important and well-documented changes that occurred in Matlab over the 35-year period since program inception that could confound results.

The Bangladesh Female Secondary Education Stipend Program.—One government program that is pertinent for MCH-FP effects on education, and difficult to control for, is the Bangladesh Female Secondary Education Stipend Program. This was a national program that became available in Matlab in 1984 for females attending grades 6-10 who were unmarried, had 75 percent attendance and scored 45 percent on school exams. The program targeted individual females, not schools, and provided a stipend and covered many school costs. This program was available to all the females in the first generation sample in this paper for the entirety of their secondary schooling in both the treatment and comparison areas, providing no variation to test for heterogeneity treatment effects of MCH-FP based on differential access to this program. In addition, we do not have data on who received the stipend program, nor sufficient data to determine who may have been eligible. Shamsuddin (2015) estimates that five years of exposure to the program led to one year gain in education. As a result, any program effects on educational outcomes of the MCH-FP program on females needs to outweigh the already large effects on education from the female stipend program.

Erosion and Flood Risk.—One potential confounder is the Meghna Dhonnogoda Irrigation Project. In 1987

the government of Bangladesh completed this project, which involved constructing a river embankment along the northern bank of the major Meghna River where it meets the west bank of the smaller Dhonnogoda River, which runs through Matlab (see Figure 1). The villages near this project were all located in the comparison area, and the embankment had two important consequences for these villages. First, seven villages in this area lining the river were partially or fully inundated between 1984 and 1986 as part of the project. All households in these villages were displaced, with most initially relocating to adjoining villages within the comparison area. To control for potential differences in the Meghna area in general, we include a variable indicating if the person's treatment village was submerged as a result of the project.

1993 BRAC Microcredit Experiment.—In 1993, BRAC introduced an experiment that provided landless females with access to microcredit. The rollout was designed to include villages in both treatment and comparison areas, but the presence of the program could still bias our results. We include an indicator for whether the village ever participated in BRAC during its experimental period.

Primary School Access.—There was a large expansion of education during the 1990s, including construction of primary schools. We control for differential primary school access by including an indicator variable that is one if a primary school was present in the individual's treatment village during the years that they were age-eligible.

Health Facility Access.—Data on access to healthcare come from the MHSS1 Village Survey. MHSS1 surveyed village leaders about health facilities used by people from their village. We construct indicators for the presence of a Family Welfare Center (a government clinic, FWC), a Family Welfare Assistant (a government health worker that travels to villages, FWA), a non-MBSS allopathic doctor, and a Trained Traditional Birth Attendant (a midwife) in the individual's birth year to account for differential access to pre- and post-natal care. We use the MHSS1 Village Survey rather than its MHSS2 equivalent to avoid selection from the timing of facility opening and closings.

Arsenic Exposure.—The control for arsenic exposure is created using 2003 measures of arsenic in tube well water. These data were collected by icddr,b. Wells are linked to MHSS1 households using the ID of the person who takes care of the well. For household who do not take care of a well, we take the average arsenic level in the 3 closest wells. For households that reported not using a tubewell in MHSS1 (which was prior to knowing about arsenic in the well), the value of arsenic is set to zero. We use the 2003 measure of arsenic rather than the one collected in 2010 because it was measured prior to knowledge of arsenic in the well, so before families engaged in well switching which could be correlated with treatment status, and since it was measured at a time closer to when the sample of interest were young children. The included control is an indicator of an arsenic level above 100 parts per billion (ppb, micrograms per liter). Note a majority of the children in the sample were born after the wells were established, so the age fixed-effects control for the length of time exposed to the well water.

2001-2003 Maternal and Infant Nutritional Intervention - MINIMat.—Between November 2001 and October 2003 pregnant females in the HDSS area were randomized into a nutritional intervention program. The randomization was done at the individual level and was conducted in both the MCH-FP treatment and comparison areas (El Arifeen et al., 2018). The children impacted by the MINIMat intervention were largely born before the second generation cohort that is considered in this study.

2007 Maternal, Neonatal and Child Health Project (MNCH).— Beginning in 2007 this program aimed to increase facility-based childbirth and neonatal health. Pregnant women enrolled in the program were visited by a community health worker at 12-14 weeks and again between weeks 32 - 34. Women were

encouraged to deliver at a facility with skilled obstetric care rather than at home. Postnatal visits were made by a community health worker on days 3, 7, and 28 to provide counseling on newborn and maternal well-being. See Rahman et al. (2011) for a detailed description of the program. The targeting of this program overlapped with the treatment area of the MCH-FP program, however children born before 2007 did not receive any benefits. Furthermore, this program was aimed at changing childbirth delivery location and it is unclear how that would impact human capital attainment. For robustness, we control for mother's exposure to the MNCH program based on her residence in 2005 interacted with a dummy indicating whether the child was born on or after 2007. Results are similar when including this control and available upon request.

Appendix D Robustness Analysis

We perform a number of robustness checks in Tables D1-D7 to examine the validity of the results for key outcome variables. Results remain similar unless noted. Below we provide details of each check. In sum, the findings reported in the paper are similar across a variety of robustness checks including no weights and for attrition bounds.

Pre-Trends in Human Capital.—We test whether human capital was improving in the treatment area prior to the onset of MCH-FP by estimating Equation 2 for a sample of individuals who were aged 24-30 or 44-65 in 1996, the same ages of the 1982-1988 and 1947-1969 cohorts in 2012. Using data from MHSS1, Table D1 shows we find no impacts on the height, cognition or educational attainment at the same point in the lifecycle for this sample of individuals who did not have access to the program during their formative years. This implies the effects estimated in Table 2 are not driven by pre-trends in the treatment area.

Extended Controls.—To account for potential confounders that may be determinants of anthropometric and cognitive development, we include a set of controls in addition to the baseline covariates listed in Table 1. This extended set of controls includes erosion and flodo risk, access to the 1993 BRAC microcredit experiment, primary school access, access to various health facilities in their year of birth and arsenic exposure above 100 parts per billion, outlined in further detail in Appendix C. Table D2 Panel B show that our results are not sensitive to the inclusion of the additional controls.

Unbalanced Controls.—Our main specification controls for a large number of baseline household characteristics with the dual purpose of controlling for potential imbalances across treatment and comparison households and improving statistical precision by controlling for characteristics related to the outcome variable. However, it is possible in doing so we are over-controlling in a context where treatment was assigned quasi-randomly. To test this possibility, we reduce the control set to only those that were not balanced between experimental areas. These controls include religion, whether the household draws water from a tubewell, the size of the respondent's family in 1974, the household head's age in 1974, an indicator for household head having less than two years of education in 1974, the household head's spouse's age in 1974 and the household head's spouse having less than two years of education in 1974 (see Table 1). The results of estimating with this limited set of controls are presented in Table D2 Panel C.

Muslims Only.—Table 1 revealed imbalance in the treatment and comparison area by religion. To determine if this imbalance affects the results, we restrict the sample to only those who report their religion as Muslim. There is insufficient sample size on those who reported Hindu as their religion to run results separately for this religious group. Table D2 Panel D shows the main results remain the same, occasionally becoming insignificant as the sample size decreases.

North and West Comparison Areas.—We take advantage of the two, geographically distinc comparison areas to the north and west of the treatment area (Figure 2) to test for confounding shocks to comparison units. Panels F and G of Table D2 restrict the comparison group sample to those living in the north and west areas, respectively. Results are qualitatively similar with some loss of significance due to the reduced sample size.

Village Fixed Effects.—We include pre-program village fixed effects to account for any time-invariant, village-specific unobservable factors that may bias our double difference estimates. Table D2 Panel H shows the double difference estimate for the 24-30 cohort is essentially unchanged with the inclusion of village fixed effects.

Spatially Correlated Errors.—Because the treatment and comparison areas are contiguous, it is possible that errors are spatially correlated in either the treatment or the comparison area. This could arise, for example, if there was a health shock such as a disease outbreak or a flood that led to migration in a given year in one of the experimental areas but not the other. Clustering at the village level is not sufficient to correct for the resulting lack of independence. To examine the possibility, we test whether the error terms from the regressions on height, education and cognition are spatially correlated, using Moran's I test with the Euclidean distance between village centroids as a weight. We perform the test at the village level, and create village level error terms by predicting the errors from our main model, and averaging the errors at the village level separately for each cohort. We find no evidence of spatial correlation in the error terms (results not reported).

Kling-Leibman and Worst Case Attrition Bounding.—We attempt to bound potential bias from survey attrition by imputing attritors' outcomes. For binary outcomes, we assign treatment (comparison) individuals a 0 (1). For continuous measures, we follow Kling, Liebman and Katz (2007) by assigning the outcome mean +/- 0.10 and 0.25 SD among the treatmnet and comparison groups as applicable. Tables D3 and D7 show significant results are bounded away from zero.

Phone Survey Sensitivity.—For the men who responded via phone, their height was self-reported rather than measured by the enumerator. Table D4 tests if the effects of the program on height are driven by this self-reporting. Column 1 shows the results for the full sample (including those surveyed by phone) without accounting for the survey method. In Column 2, we include a control that indicates if the respondent was surveyed by phone. Column 3 excludes anyone who answered the phone survey. Results are similar across specification.

TABLE D1 — FIRST GENERATION: ITT EFFECTS ON HUMAN CAPITAL OF ADULTS AGED 24-30 IN 1996

	Pooled		Males		Females	
	Height (cm) (1)	MMSE (z-score) (2)	Height (cm) (3)	MMSE (z-score) (4)	Height (cm) (5)	MMSE (z-score) (6)
Treat*(Age 24-30 in 1996)	-0.589	-0.107	-0.999	0.074	-0.225	-0.254
	(0.707)	(0.099)	(1.254)	(0.159)	(0.983)	(0.143)$^{+}$
Treat	0.298	0.021	0.864	-0.160	-0.263	0.175
	(0.635)	(0.086)	(0.916)	(0.122)	(0.788)	(0.128)
Mean	154.616		161.593		150.117	
Obs	1219	1229	465	448	754	781

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table tests for differential pre-trends in the treatment area by estimating double difference ITT effects on the height, internally standardized MMSE and educational attainment of people aged 24-30 in 1996 using people aged 44-65 in 1996 as a comparison group. Standard errors are clustered at the pre-program village level and reported in parentheses. Age 24-30 means are for the comparison group. All regressions include individual and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 1996 MHSS1 survey from which the measures of human capital were gathered. Individual characteristics include year of birth fixed effects, age cohort fixed effects, and religion. Pre-intervention characteristics include all characteristics in Table 1.

TABLE D2 — FIRST GENERATION: ROBUSTNESS CHECKS

	Pooled				Males				Females			
	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Base Results												
Treat*(1982-1988)	0.968	-0.041	-0.011	0.419	1.047	-0.037	0.061	0.960	0.999	-0.042	-0.073	-0.098
	(0.342)**	(0.022)+	(0.049)	(0.206)*	(0.629)+	(0.029)	(0.072)	(0.377)*	(0.460)*	(0.034)	(0.066)	(0.239)
Treat	0.034	0.021	0.018	0.029	-0.022	0.023	-0.026	-0.049	-0.011	0.019	0.063	0.084
	(0.280)	(0.018)	(0.043)	(0.159)	(0.421)	(0.022)	(0.057)	(0.262)	(0.337)	(0.025)	(0.046)	(0.164)
Obs	4213	4213	4039	4340	1893	1893	1723	1952	2320	2320	2316	2388
Panel B: Add Extended Controls												
Treat*(1982-1988)	0.930	-0.037	0.001	0.424	1.038	-0.037	0.086	1.043	0.958	-0.038	-0.074	-0.096
	(0.336)**	(0.022)+	(0.051)	(0.204)*	(0.630)	(0.030)	(0.073)	(0.377)**	(0.452)*	(0.034)	(0.068)	(0.238)
Treat	-0.025	0.019	-0.005	0.005	-0.075	0.025	-0.053	-0.021	-0.071	0.015	0.051	0.032
	(0.264)	(0.019)	(0.042)	(0.157)	(0.409)	(0.023)	(0.056)	(0.260)	(0.337)	(0.027)	(0.047)	(0.167)
Obs	4210	4210	4036	4337	1891	1891	1721	1950	2319	2319	2315	2387
Panel C: Unbalanced Controls												
Treat*(1982-1988)	1.085	-0.032	-0.026	0.257	0.840	-0.025	0.031	0.760	0.886	-0.035	-0.068	-0.221
	(0.446)*	(0.021)	(0.047)	(0.221)	(0.578)	(0.028)	(0.073)	(0.375)*	(0.446)*	(0.033)	(0.061)	(0.246)
Treat	-0.074	0.014	0.026	0.080	0.170	0.014	-0.006	0.020	0.055	0.009	0.061	0.128
	(0.362)	(0.019)	(0.041)	(0.163)	(0.403)	(0.021)	(0.057)	(0.252)	(0.320)	(0.026)	(0.043)	(0.174)
Obs	4213	4213	4039	4340	1893	1893	1723	1952	2320	2320	2316	2388
Panel D: Only Muslims												
Treat*(1982-1988)	0.887	-0.035	-0.020	0.447	1.058	-0.049	0.026	1.095	0.815	-0.019	-0.060	-0.181
	(0.389)*	(0.024)	(0.054)	(0.224)*	(0.687)	(0.030)+	(0.077)	(0.409)**	(0.519)	(0.036)	(0.071)	(0.264)
Treat	0.164	0.015	0.023	0.022	-0.008	0.030	-0.028	-0.148	0.218	0.002	0.070	0.148
	(0.304)	(0.019)	(0.044)	(0.170)	(0.466)	(0.024)	(0.059)	(0.290)	(0.346)	(0.026)	(0.050)	(0.167)
Obs	3783	3783	3621	3896	1695	1695	1537	1748	2088	2088	2084	2148
	⋮											

Notes: + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Standard errors are clustered at the pre-program village level and reported in parentheses. Panel A replicates the results of Table 2. Panel B includes year of birth fixed effects, cohort fixed effects and controls for religion but limits the pre-intervention characteristics to only those that were imbalanced (family size, tubewell access, household head individual and spouse characteristics [ages in 1974 and an indicators for less than two years of education]). Panel C adds controls for arsenic exposure above 100ppb, living in a village that was fully or partially eroded, access to the BRAC microcredit experiment, access to primary school and village access to healthcare [family welfare clinic, family welfare assistant, midwife, and alopathic] as of the individual's year of birth. Panel D estimates the original models using the subsample of Muslim respondents.

TABLE D2 — FIRST GENERATION: ROBUSTNESS CHECKS (CONT.)

	Pooled				Males				Females			
	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)	Height (cm)	Short Stature (=1)	Cognition Index (z-score)	Grades Attained (years)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
⋮												
Panel E: Unweighted												
Treat*(1982-1988)	0.890	-0.034	0.007	0.366	0.758	-0.030	0.062	0.864	1.046	-0.036	-0.033	-0.116
	(0.336)**	(0.020)	(0.048)	(0.199)+	(0.622)	(0.028)	(0.070)	(0.359)*	(0.456)*	(0.031)	(0.063)	(0.221)
Treat	0.101	0.012	0.003	0.090	0.209	0.019	-0.031	0.020	-0.027	0.009	0.034	0.144
	(0.266)	(0.016)	(0.039)	(0.141)	(0.411)	(0.022)	(0.051)	(0.234)	(0.326)	(0.022)	(0.044)	(0.140)
Obs	4213	4213	4039	4340	1893	1893	1723	1952	2320	2320	2316	2388
Panel F: North Comparison Area												
Treat*(1982-1988)	0.858	-0.043	0.013	0.597	1.020	-0.031	0.108	1.237	0.775	-0.051	-0.097	-0.012
	(0.401)*	(0.026)	(0.064)	(0.302)+	(0.743)	(0.035)	(0.084)	(0.435)**	(0.557)	(0.040)	(0.086)	(0.311)
Treat	0.211	0.012	0.011	-0.211	0.013	0.010	-0.039	-0.459	0.270	0.016	0.068	0.002
	(0.351)	(0.023)	(0.050)	(0.186)	(0.523)	(0.028)	(0.067)	(0.291)	(0.423)	(0.031)	(0.053)	(0.169)
Obs	3111	3111	2989	3208	1396	1396	1278	1440	1715	1715	1711	1768
Panel G: West Comparison Area												
Treat*(1982-1988)	0.952	-0.033	-0.029	0.163	0.890	-0.040	0.043	0.534	1.103	-0.027	-0.062	-0.164
	(0.413)*	(0.027)	(0.054)	(0.208)	(0.744)	(0.031)	(0.094)	(0.474)	(0.570)+	(0.042)	(0.076)	(0.275)
Treat	-0.115	0.034	0.019	0.333	-0.056	0.041	-0.035	0.434	-0.233	0.024	0.062	0.209
	(0.305)	(0.020)+	(0.048)	(0.204)	(0.457)	(0.023)+	(0.068)	(0.318)	(0.383)	(0.031)	(0.057)	(0.248)
Obs	3074	3074	2948	3174	1381	1381	1258	1430	1693	1693	1690	1744
Panel H: Village Fixed Effects												
Treat*(1982-1988)	1.013	-0.047	-0.013	0.412	0.955	-0.047	-0.010	0.907	1.019	-0.032	-0.025	-0.075
	(0.338)**	(0.023)*	(0.049)	(0.206)*	(0.673)	(0.032)	(0.072)	(0.386)*	(0.485)*	(0.035)	(0.070)	(0.242)
Obs	4213	4213	4052	4340	1893	1893	1729	1952	2320	2320	2323	2388

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Standard errors are clustered at the pre-program village level and reported in parentheses. Panel E does not reweight for attrition between birth and the 2012 MHSS2 survey. Panels F and G restrict the sample of comparison units to those living in the North and West comparison areas, respectively (see Figure 2). Panel H includes fixed effects for pre-program villages which are collinear with *Treat*.

TABLE D3 — FIRST GENERATION: KLING-LEIBMAN ATTRITION BOUNDING

	Height (cm)					Cognition Index (z-score)					Grades Attained (years)				
	Base	0.10 SD		0.25 SD		Base	0.10 SD		0.25 SD		Base	0.10 SD		0.25 SD	
		+	-	+	-		+	-	+	-		+	-	+	-
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Panel A: Pooled															
Treat*(1982-1988)	0.968	0.820	0.913	0.750	0.982	-0.016	-0.010	0.001	-0.017	0.008	0.419	0.410	0.420	0.402	0.428
	(0.342)**	(0.310)**	(0.309)**	(0.312)*	(0.310)**	(0.048)	(0.044)	(0.044)	(0.045)	(0.044)	(0.206)*	(0.202)*	(0.200)*	(0.204)+	(0.199)*
Mean	155.766	155.861	156.003	155.755	156.109						4.258	4.236	4.312	4.179	4.368
Obs	4213	4764	4764	4764	4764	4048	4584	4584	4584	4584	4340	4752	4752	4752	4752
Panel B: Males															
Treat*(1982-1988)	1.047	0.802	0.971	0.676	1.097	0.049	0.037	0.055	0.024	0.069	0.960	0.793	0.855	0.747	0.901
	(0.629)+	(0.554)	(0.547)+	(0.562)	(0.544)*	(0.069)	(0.062)	(0.062)	(0.063)	(0.063)	(0.377)*	(0.352)*	(0.350)*	(0.356)*	(0.349)*
Mean	162.404	162.315	162.492	162.183	162.625						4.985	4.934	5.035	4.859	5.110
Obs	1893	2197	2197	2197	2197	1728	2021	2021	2021	2021	1952	2188	2188	2188	2188
Panel C: Females															
Treat*(1982-1988)	0.999	0.880	0.923	0.848	0.955	-0.072	-0.048	-0.043	-0.052	-0.039	-0.098	0.020	-0.004	0.039	-0.022
	(0.460)*	(0.418)*	(0.417)*	(0.420)*	(0.419)*	(0.067)	(0.062)	(0.061)	(0.062)	(0.061)	(0.239)	(0.233)	(0.234)	(0.233)	(0.235)
Mean	150.330	150.275	150.386	150.191	150.469						3.663	3.636	3.690	3.595	3.731
Obs	2320	2567	2567	2567	2567	2320	2563	2563	2563	2563	2388	2564	2564	2564	2564

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows double difference ITT effects across different attrition bounding schemes. Base models replicate the effects shown in Table 2 while subsequent columns impute attritors' human capital measures with the within-sample mean ± 0.10 and 0.25 standard deviations (SD). Treatment effect estimates for the 44-65 year-old comparison group are omitted for brevity. Age 24-30 means are for the comparison group. Standard errors are clustered at the pre-program village level and reported in parentheses.

TABLE D4 — FIRST GENERATION: SENSITIVITY OF MALE HEIGHT TO SURVEY METHOD

	Base Results (1)	Self-Report Control (2)	Drop Self-Report Sample (3)
Treat*(1982-1988)	1.047	1.030	1.251
	(0.629)$^{+}$	(0.627)	(0.662)$^{+}$
Treat	-0.022	-0.015	0.022
	(0.421)	(0.420)	(0.425)
Obs	1893	1893	1729

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Column (1) replicates the male height results of Table 2. Column (2) controls for whether the respondent self-reported their height in the phone survey. Column (3) limits the sample to those whose height was measured by an enumerator directly. Standard errors are clustered at the pre-program village level and reported in parentheses.

TABLE D5 — SECOND GENERATION: DOUBLE DIFFERENCE ITT EFFECTS ON HUMAN CAPITAL

	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)
	Ages 0-14		Ages 0-6	Ages 7-14
	(1)	(2)	(3)	(4)
Panel A: Pooled				
Mom Treat*(1982-1988 Mom)	0.076	-0.030	-0.020	0.136
	(0.112)	(0.042)	(0.167)	(0.124)
Mom Treat	-0.012	-0.004	-0.021	-0.150
	(0.066)	(0.027)	(0.120)	(0.088)$^{+}$
Mean	-1.262	0.258		
Obs	2796	2796	639	740
Panel B: Males				
Mom Treat*(1982-1988 Mom)	-0.140	0.061	-0.071	0.004
	(0.148)	(0.064)	(0.216)	(0.168)
Mom Treat	0.026	-0.036	-0.055	-0.164
	(0.078)	(0.034)	(0.145)	(0.102)
Mean	-1.221	0.242		
Obs	1433	1433	320	377
Panel C: Females				
Mom Treat*(1982-1988 Mom)	0.388	-0.166	0.044	0.386
	(0.161)*	(0.062)**	(0.240)	(0.163)*
Mom Treat	-0.069	0.039	0.020	-0.166
	(0.084)	(0.036)	(0.187)	(0.107)
Mean	-1.302	0.273		
Obs	1363	1363	319	363

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table reports estimates from an ITT double-difference specification on human capital outcomes for the second generation using similarly aged children of mothers born prior to 1982 as the comparison group. Standard errors in parentheses are clustered at the treatment village level. Means are for children of 24-30 year-old comparison group mothers. All regressions include individual characteristics and pre-intervention characteristics interacted with birth cohort and are weighted to correct for attrition between birth and the 2012 MHSS2 survey. See Table 5 for a description of second generation human capital measures. All models include child's six-month age fixed effects, mother's year of birth fixed effects, mother's religion and mother's pre-intervention characteristics listed in Table A3.

TABLE D6 — SECOND GENERATION: ROBUSTNESS CHECKS

	Pooled				Males				Females			
	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)	Height-for-Age (WHO z-score)	Stunted (=1)	Child Development Index (z-score)	Cognition Index (z-score)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Base Results												
Mom Treat (=1)	0.106	-0.044	-0.027	0.004	-0.022	-0.002	-0.099	-0.198	0.335	-0.139	0.059	0.268
	(0.103)	(0.035)	(0.109)	(0.098)	(0.157)	(0.062)	(0.146)	(0.144)	(0.146)*	(0.048)**	(0.137)	(0.145)+
Obs	629	629	331	262	311	311	158	135	318	318	173	127
Panel B: Add Extended Controls												
Mom Treat (=1)	0.156	-0.039	-0.080	0.026	0.142	-0.045	-0.158	-0.263	0.250	-0.089	-0.061	0.315
	(0.115)	(0.040)	(0.118)	(0.111)	(0.182)	(0.073)	(0.155)	(0.155)+	(0.158)	(0.054)+	(0.163)	(0.150)*
Obs	626	626	328	262	309	309	156	135	317	317	172	127
Panel C: Control for MNCH Eligibility												
Mom Treat (=1)	0.026	-0.020	0.040	0.004	-0.173	0.072	0.003	-0.198	0.277	-0.131	0.038	0.268
	(0.100)	(0.039)	(0.131)	(0.098)	(0.162)	(0.069)	(0.194)	(0.144)	(0.163)+	(0.058)*	(0.176)	(0.145)+
Obs	629	629	331	262	311	311	158	135	318	318	173	127
Panel D: Mother's Unbalanced Controls												
Mom Treat (=1)	0.091	-0.037	-0.008	0.026	-0.025	0.006	-0.108	-0.167	0.289	-0.116	0.128	0.351
	(0.102)	(0.035)	(0.094)	(0.094)	(0.159)	(0.059)	(0.140)	(0.132)	(0.145)*	(0.046)*	(0.108)	(0.148)*
Obs	629	629	331	262	311	311	158	135	318	318	173	127
Panel E: Muslims Only												
Mom Treat (=1)	0.016	-0.012	-0.077	0.046	-0.217	0.081	-0.255	-0.118	0.278	-0.138	0.087	0.269
	(0.106)	(0.032)	(0.116)	(0.091)	(0.177)	(0.060)	(0.166)	(0.154)	(0.151)+	(0.049)**	(0.142)	(0.147)+
Obs	573	573	298	243	276	276	140	121	297	297	158	122
Panel F: Unweighted												
Mom Treat (=1)	0.103	-0.041	-0.031	-0.013	-0.029	0.002	-0.140	-0.209	0.314	-0.133	0.077	0.254
	(0.101)	(0.032)	(0.107)	(0.100)	(0.148)	(0.059)	(0.146)	(0.149)	(0.149)*	(0.047)**	(0.133)	(0.150)+
Obs	629	629	331	262	311	311	158	135	318	318	173	127

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. Panel A replicates the results of Table 5. Panel B includes six month age fixed effects but limits the mother's attributes and preintervention characteristics to only those that were imbalanced (age, religion, family size, tubewell access, household head's age and household head's spouse's age). Panel C includes controls for available healthcare supply in the child's year of birth (family welfare clinic, family welfare assistant, midwife and alopathic) and distance to education (government, private and informal primary school, government secondary school). Panel D limits the sample children of Muslim mothers. Panel E does not reweight for mother's attrition. Panel F controls for mother's eligibility for MNFP based on her 2005 residence interacted with a dummy indicating if the child was born on or after 2007.

TABLE D7 — SECOND GENERATION: KLING-LEIBMAN AND WORST CASE ATTRITION BOUNDING

	Height-for-Age (z-score)					Stunting (=1)			Cognition Index 0-6 (z-score)					Cognition Index 7-14 (z-score)				
	Base	0.10 SD		0.25 SD		Base	Worst case		Base	0.10 SD		0.25 SD		Base	0.10 SD		0.25 SD	
		+	-	+	-		+	-		+	-	+	-		+	-	+	-
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Panel A: Pooled																		
Mom Treat	0.106	0.130	0.100	0.152	0.077	-0.044	0.014	-0.118**	-0.027	-0.014	-0.049	0.013	-0.075	0.004	0.005	-0.023	0.027	-0.044
	(0.103)	(0.096)	(0.096)	(0.097)	(0.096)	(0.035)	(0.034)	(0.034)	(0.109)	(0.094)	(0.093)	(0.094)	(0.092)	(0.098)	(0.087)	(0.087)	(0.087)	(0.087)
Observations	629	672	672	672	672	629	672	672	331	374	374	374	374	262	305	305	305	305
Panel B: Males																		
Mom Treat	-0.022	-0.015	-0.049	0.011	-0.074	-0.002	0.071	-0.083	-0.099	-0.092	-0.135	-0.059	-0.168	-0.198	-0.174	-0.209+	-0.147	-0.235+
	(0.157)	(0.146)	(0.146)	(0.146)	(0.146)	(0.062)	(0.058)	(0.060)	(0.146)	(0.127)	(0.126)	(0.128)	(0.126)	(0.144)	(0.125)	(0.124)	(0.126)	(0.124)
Observations	311	339	339	339	339	311	339	339	158	186	186	186	186	135	163	163	163	163
Panel C: Females																		
Mom Treat	0.335*	0.340*	0.312*	0.362*	0.291*	-0.139**	-0.074	-0.192**	0.059	0.073	0.049	0.091	0.030	0.268+	0.245+	0.223+	0.262*	0.207
	(0.146)	(0.138)	(0.139)	(0.138)	(0.139)	(0.048)	(0.047)	(0.049)	(0.137)	(0.125)	(0.125)	(0.125)	(0.125)	(0.145)	(0.131)	(0.132)	(0.131)	(0.132)
Observations	318	333	333	333	333	318	333	333	173	188	188	188	188	127	142	142	142	142

Notes: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$. This table shows single difference ITT effects across different attrition bounding schemes. Base models replicate the effects shown in Table 5 while subsequent columns impute attritors' human capital measures with the within-sample mean ± 0.10 and 0.25 standard deviations (SD). For the binary outcome stunting, we impute the attritors with ones (+) or zeros (-). Standard errors are clustered at the pre-program village level and reported in parentheses.

Appendix E Weights

The main results are weighted for attrition between birth and MHSS2 using inverse propensity weights. As described in Section III of the main paper, the analysis sample includes respondents from MHSS1 and individuals from MHSS1 households that had migrated out of the DSS area prior to the survey conducted in 1996. The main reasons for non-response are migration in early adulthood and death primarily during infancy. Weights are constructed in two steps. First, we estimate weights to account for selection into the MHSS1 sample frame between birth and MHSS1, which is mainly a result of mortality. Second, we estimate weights to account for attrition of MHSS1 respondents in the MHSS2 survey. We estimate these two probabilities separately and then multiple them to obtain a weight to account for attrition between birth and MHSS2.

The weight the account for attrition between birth and MHSS1 estimate the conditional probability that an individual born in the study site was present to be surveyed in MHSS1 using demographic surveillance data. To do this, we assign treatment status to the universe of individuals born in the study site between 1977 and 1988. Separately by cohort and sex, we use a probit model to predict the probability an individual is present in the study site on January 1, 1996 using the set of baseline household and household head characteristics (which includes pre-program migration networks for the household compound), their interactions with the treatment variable, month of birth and year of birth fixed effects, and indicators for whether an individual was from a village that experienced erosion or was exposed to the Meghna Dhonnogoda Irrigation Project.

The weight to account for attrition between MHSS1 and MHSS2 is constructed in a similar manner. We estimate the probability of non-attrition between the two survey waves for each cohort-sex group using a probit model and the same set of covariates. The resulting attrition weight is the inverse of the product of the two probabilities.