The Impact of Early Childhood Access to Community Health Workers: Evidence From China's *Barefoot Doctors*

Ying Liu*

November 9, 2021

Abstract

This paper studies the short-run and long-run effects of China's *Barefoot Doctors* — the world's first Community Health Workers (CHWs) project — that started in 1969. I collect a unique county-level dataset on the *Barefoot Doctor* program from historical gazetteers. Using the geographic variation in the programs intensity across counties after the introduction of the program in 1969, I employ a difference-in-differences model to identify the programs influence. I find that greater exposure to *Barefoot Doctors* significantly reduced neonatal mortality in the short-run, and significantly improved self-reported health, boosted educational attainment, and increased working hours in the long-run. I find that one underlying mechanism may be prolonged breastfeeding. This paper provides the first evidence that CHWs have lasting impacts on health and economic outcomes.

Keywords: community health workers, China's barefoot doctors, health, human capital

*Department of Economics, University of Pittsburgh, Pittsburgh Pennsylvania USA 15213. Email: ying.liu@pitt.edu

1 Introduction

The World Health Organization (WHO) estimates that by 2030, there will be a shortage of 18 million health workers, the bulk of which will be in developing countries.[1] A potential solution to this shortfall is to expand the use of community health workers (CHWs for short), who have played a major role in the provision of primary health care in developing areas for the past two decades (Singh and Sachs, 2013). CHWs are local residents who receive basic medical training in order to provide primary health care to their communities. The usage of CHWs presents a trade-off between coverage and quality: CHWs can reach a wider population at lower costs, but the quality of care may fall below formal standards. While the short-run health benefits of well-implemented CHWs projects have been documented (Bjrkman Nyqvist et al., 2019; Bhutta et al., 2010; Gilmore and McAuliffe, 2013; Perry et al., 2014), little is known about their long-run impacts. Existing evidence suggests that early improvements in health can benefit later life outcomes (Currie and Stabile, 2003, 2006; Case et al., 2005; Miguel and Kremer, 2004).[2] However, substandard quality of care, such as high diagnostic error or over-prescription of antibiotics, may be problematic in the long-run (Llor and Bjerrum, 2014). In addition, CHWs may crowd out formal medical care in rural areas (Deserranno and Qian, 2020). Therefore, the long-run impact of CHWs remains an important empirical question.

The main challenge in studying the long-run effects of CHW projects is that they have existed for less than 20 years in most countries. An important exception is China's *Barefoot Doctor* (BFD) program, which is the world's first CHW project. Started in 1969, the BFD program recruited more than 1.3 million villagers for medical training. Most of

[1] See WHO's report:https://www.who.int/news/item/22-05-2019-momentum-for-community-health-workers-at-the-seventy-second-world-health-assembly

[2] See Currie and Madrian (1999) for a review of studies linking health to educational and labor market outcomes.

them were secondary school graduates, and the average length of training was three months (De Geyndt et al., 1992). This program eventually covered more than 85% of villages in China. After China's economic reform in 1978, however, a large number of BFDs started migrating to urban areas or switching to more profitable jobs. Subsequently the program ended in 1985.

To my knowledge, no existing work has identified a causal relationship between BFDs and patient outcomes. One challenge for identification is that the BFD program was rolled out nationwide at almost the same time. Therefore, one cannot compare places that were treated and untreated to estimate treatment effects. One potential strategy would be to exploit variation in the intensity of the program across regions. However, existing data on the BFD program is only available at the national level.

To overcome this challenge, I assemble a unique county-level dataset on the program by manually collecting and digitizing data from historical gazetteers. I find that the number of BFDs per capita varies across counties, which provides variation in the program's intensity.[3] Combining it with the introduction of the BFD program in 1969, I am able to construct a difference-in-differences model for identification. I include county and year of birth fixed effects, survey round fixed effects, county-specific linear time trends, and individual demographics. I also include mother fixed effects for some regressions. Given that the cohorts born before the program are also treated by the BFDs later in their lives, my estimates should be interpreted as the effect of *additional* years of exposure to BFDs both *in utero* and during early childhood.

I first use data from the China In-Depth Fertility Survey (1985, 1987) and find that in the short-run, exposure to BFDs significantly reduced infant mortality and under-5 child mortality. Next, for the long-term outcomes, I use data from *Chinese General Social Survey*

[3] Counties (*Xian*), or county-level divisions, are in the third level of the administrative hierarchy after Provinces (1st level) and Prefectures(2nd level).

(2010 to 2015). The results show that greater exposure to BFDs *in utero* and during early childhood raises self-reported health, increases years of schooling, and increases working hours about 40 years later. I find null effect on physical health measurements such as height, weight, and BMI, and null effect on unemployment and labor force participation. These effects are heterogeneous between genders: males experience the largest effects on educational attainment and working hours, while females see the greatest impact on self-reported health and physical health. Finally, I test for potential mechanisms underlying the effects of BFDs. I find that BFDs on average increases breastfeeding duration by about four months, and increases the probability of receiving professional health checks during pregnancy by 10.8%. The effects of the BFDs on breastfeeding and neonatal mortality are greater among low-income families, and mothers without any education.

This paper contributes to four strands of literature. First, it provides the first piece of evidence of the long-run impact of CHWs. Existing literature on CHWs has focused on the short-run effects, and find mixed results. For example, Bjrkman Nyqvist et al. (2019) find that the infant and child mortality are significantly reduced three years after the introduction of CHWs in Uganda. Another study finds that CHWs significant increase infant head circumference at birth and infant developmental scores in the US (St James et al., 1999). However, Sloan et al. (2008) find that a community-based training on general childcare had no effect on neonatal and infant mortality. See Bhutta et al. (2010) and Gilmore and McAuliffe (2013)'s papers for systematic review on the impact of CHWs. For reasons mentioned above, short-run effects may fail to persist or reverse over time. If long-run effects are negative, they should be included as costs in program analyses. Even if they are null or positive, a sense of magnitude could help public officials choose between multiple health investments. Moreover, a long time horizon allows one to look at outcomes such as education and labor market performance.

Second, this paper sheds light on mechanisms of how CHWs can improve health. In particular, I find that prolonged breastfeeding may be one mechanism behind. Studies have shown that breastfeeding provides protection against malnutrition and infectious diseases, which leads to a reduction in morbidity and mortality in the short-run, and may be critical for shaping longer-run health trajectories (Victoria et al., 2000; Fitzsimons and Vera-Hernández, 2013). This is especially important for developing countries, since the protective effects are greatest when maternal education is low (Victoria et al., 2000). Studies also find breastfeeding can improve intelligence quotient (IQ), years of schooling, and income later in life (Victora et al., 2015). Empirical studies using historical data find similar result. For example, Botticini et al. (2019) find that Judaism's religious norm on prolonged breastfeeding contributed to the relatively low infant and child mortality rate among the Jewish population in Central and Eastern Europe from 1500 to 1930.

Third, this paper provides the first piece of empirical evidence on the health impact of China's Barefoot Doctors. Most existing literature on BFDs are descriptive and qualitative mainly due to data limitations. For example, Fang (2012) uses historical documents and personal interviews, and find that the BFD program introduced modern Western medicine to rural China. De Geyndt et al. (1992) described the features of the BFDs and present anecdotal and suggestive evidence of mortality reduction after the BFD program. This paper, by digitizing a novel county-level dataset on the BFD program, is able to empirically study the impact of BFDs.

Fourth, this research adds to the literature on the long-term influence of early childhood health interventions.[4] Existing evidence on the long-term impact of early childhood interventions is limited and the results are mixed. First of all, in terms of health outcomes, although many studies find positive long-term health improvements (Hoddinott et al., 2008;

[4] Early childhood is usually defined as starting at birth and ending at age five (Currie and Almond, 2011)

Hjort et al., 2017; Wherry and Meyer, 2016; Brown et al., 2015; Gertler et al., 2021), a study of the home visiting program in the U.S find small effects on health which was not worth the cost (Aos et al., 2004). Second, results are also mixed as for educational and labor market outcomes. For example, Hjort et al. (2017) study the 1937 Danish home visiting program find null effects on education and labor market outcomes. However, other studies find large and significant effects on income and schooling (Bharadwaj et al., 2013; Gertler et al., 2021; Hoehn-Velasco, 2021; Field et al., 2009). This paper provides a new evidence of a health intervention's lasting impacts on health, educational and labor market outcomes. Third, studies on the effects of early childhood interventions usually find heterogeneous effects across genders. Often, they find the health and education improvements are greater for women (Hoddinott et al., 2008), while improvements in labor market outcomes are greater for men Pitt et al. (2012). This paper, on the contrary, find that although health improvements are greater for women than for men, the education and labor market outcomes are much concentrated on men.

The reminder of the paper is structured as follows. Section 2 introduces the historical and institutional background of China's BFD program. Section 3 describes the data. Section 4 presents the empirical design. Section 5 discusses the results. Section 6 studies the mechanism underlying the effects. Section 7 compares the scale of the effects to similar interventions. Section 8 discusses possible threats to identification and robustness checks. Finally, section 9 concludes.

2 Institutional Background

2.1 The Barefoot Doctor Program

At the beginning of the 1950s, China's rural population barely had access to primary medical care. There were 1.4 million certified doctors in China, of which over 80% were located in urban areas, while 80% of the whole population lived in rural areas (De Geyndt et al., 1992). In 1965, under Chairman Mao's demand,[5] local governments began exploiting methods to improve rural health care. In 1968, one method called *Barefoot Doctors* attracted the government's attention. The name comes from the fact that these doctors spent half of the time practicing health care, while the other half working in farmlands barefooted (De Geyndt et al., 1992). On September 1968, the central government start promoting the BFD program to the whole nation.

The BFD program was managed by each brigade's *cooperative medical system* (CMS). Usually, a village is one unit of brigade.[6] Each villager pays a flat annual participation fee of 1 to 3 yuan (about \$0.50 to \$1.50 based on exchange rate in 1970) into the CMS. This was equivalent to 1% to 3% of a family's disposable income (Zhang and Unschuld, 2008). This fund was used to pay for medical drugs, medical facilities and in some regions, to compensate the BFDs.[7]

Except for several pioneer counties, the majority of the counties started the BFD program in 1969. It was not entirely a new thing though. Health workers similar to BFDs

[5] Mao gave a speech on June 26th, 1965 which is later referred to as the *Six Two Six Instruction*. In this speech, he asked to "switch the focus of health and medical works to the rural areas".

[6] Brigade is a special terminology used at that time. People in one village usually formed into one brigade. In each brigade, farmers are organized to work together and shared their gains. Instead of monetary income, the brigade recorded each persons work points for his workload, and the reallocation of resources were based on the earned work points.

[7] The economic returns of being a BFD varies across regions. In some places, BFDs receive payment from health visits and from selling medical drugs, and make above average earnings. In other places, BFDs did not receive extra payment. And since they have to spend time doing health practice instead of farm works, they made the same or even less than average villagers.

appeared as early as in Republican China (Xu and Hu, 2017). After the 1950s, some rural health workers and midwives were trained by township hospitals, and many of them became BFDs after 1969. But the scale of the rural health workers are not comparable to BFDs after 1969. What's more, BFDs after 1969 were supported by the CMS, which made them more effective and sustainable.

The initial expansion of BFD program was drastic and radical. Many counties were aiming at "free medical care for everyone". They required BFDs to offer services free of charge and provided patients with free medical drugs. This caused a serious moral hazard problem. According to gazetteer records, it was not uncommon that the funds in the cooperative medical system for a whole year ran out within 2 months. Consequently, in many counties, the BFD program came to a low point and even was suspended, between 1972 and 1973. After 1974, local governments started to change the way CMS and BFDs worked. Many changed from free medical care to a co-pay or fully charged system and BFDs were compensated better. The BFD program gradually recovered in 1974. The number of BFDs reached its peak in 1975, with more than 1.6 million BFDs across the nation. This program eventually covered over 85% of villages. Figure **??** in the appendix gives a brief view of the timeline of the BFD program.

Still, the BFD program did not have proper economic incentives. The reason many BFDs continued their jobs was that they were not allowed to opt out under the cooperative system. Also, many of them valued this job as a political honor. Therefore, it is not surprising to see that the number of BFDs dropped quickly after the beginning of economic reform in 1978. Many of the BFDs migrated to cities or went back to agriculture works. In 1982, the government required all existing BFDs to take the qualification exams for further medical practice. Those with certificates were then referred to as *Village Doctors*, and they could set up private clinics and charge market price. In 1985, the cooperative medical system was abandoned, and fiscal support was no longer provided from the government. According to Hsiao (1984), the ending of BFD programs is "the unintended consequences of economic reform".

There is some suggestive evidence for health improvements during the time of BFD programs. From 1969 to 1978, China's child mortality rate declined from 117.2 to 69.6 (per 1000 live births), infant mortality rate decreased from 82.9 to 52.6 (per 1000 live births), and life expectancy at birth increased from 57.6 to 65.9 years. The improvement of health can hardly be attributed to increased nutrition, since the average calorie intake for the rural population dropped below adequacy standards shortly after 1958 and remained low for two decades (Piazza, 2019).

2.2 The Jobs of Barefoot Doctors

The jobs of BFDs were both preventive and curative. In general, they were in charge of disease diagnosis, health education, medicine prescription, etc. In some areas the BFDs also grew and made traditional herb drugs. The method adopted by BFDs was a mixture of traditional and western medicine (Young, 1989), and they were encouraged to use the most cost-effective medical practice and drugs.

Since this paper focus on *in utero* and early childhood exposure to BFDs, it is important to understand healthcare provided to pregnant mothers and children by BFDs. According to "*A Manual for Barefoot Doctors (Chijiao yisheng shou ce)*", BFDs were required to provide health checks for pregnant mothers and newborns, birth delivery, vaccine injection, and health education. The health checks for pregnant mothers included disease screening, personal hygiene education, and healthy diet advice. The health education for new mothers included encouraging mothers to breastfeed for the first one to two years and to provide exclusive breastfeeding for at least the first 6 months.

BFDs also took part in other health campaigns around the same time. For example, the "New birth delivery method (*Xinfa jiesheng*)", [8] "Patriotic Health Campaigns (*Aiguo weisheng*)", [9] and "Health for children (*Ertong Baojian*)", etc.[10] In addition, BFDs worked on public health campaigns, for example promoting usage of clean water and garbage disposal. According to gazetteers, most of the health campaigns were conducted with the assistance of barefoot doctors. Therefore, the BFD's program's interaction with health campaigns above is also included in the estimates of BFDs' effects on health.

2.3 Distribution of Barefoot Doctors' Density

Historical sources point to at lease two main factors that introduced the variation in the density of BFDs across counties: the degree of support from the local government, and the number of available BFD candidates. First, the local government, especially the prefecture level government, plays a crucial role in training BFDs. For example, in the gazetteer of *Cangzhou*, it records: "After Mao's command to focus healthcare work on rural areas, the city and county hospital of Cangzhou sent their doctors and other medical workers to the rural villages. They helped control the disease and train local barefoot doctors, and helped build cooperative medical clinics." Therefore, the amount of health resources from the prefecture level government may be correlated with the local intensity of the BFD program. The second determinant is the number of available of BFD candidates. According to gazetteer records, BFDs come from four main sources: local youth with some level of education and clean political background (born in a poor family and loyal to the cultural revolution); sent-down youth, and sent-down doctors from urban areas; previously trained health workers; and existing traditional village doctors. Unfortunately there's

[8] Introducing modern delivery methods to replace the traditional home-delivery.

[9] Which encouraged people to use clean water, build toilets, kill pests and take care of personal hygiene.

[10] Aims to give out vaccines for children and free health examination.

no data on the exact portion of each source as far as I known. In order to avoid those potential confounders, in the empirical study I will control for county-level and prefecture-level characteristics at baseline, including the density of sent-down youths, the education level, and the governments' health investment.

3 Data and Variables

3.1 Gazetteer Data

I construct a unique county-level dataset on the details of the barefoot doctor program.[11] I manually collected it from China's historical gazetteers.[12] As far as I know, this is the first dataset that contains quantitative information on China's barefoot doctor program.

I collect data on three categories of variables: the timeline of BFD program, the number of BFDs at each county, and health investment from local governments. First, I document the begin year, end year, and interrupted years of BFD program in each county. The begin year of BFD program is late 1968 to early 1969 for almost all counties. As for the end year, many counties don't have clear records on when the BFD program ends. More often, the gazetteer would record that the program gradually phased out from 1978 to 1985. Therefore, to be more conservative, I use year 1978 as the ending year for all counties. In addition, it is often recorded that the BFD program was interrupted during 1972-1973, and I will exclude cohorts born during this time period in the robustness check.

Second, I collect the number of BFDs in each county. I use this to calculate the density of BFDs represents the program's intensity in each county. It is calculated using number of BFDs of a county divided by its population size in 1964, times one thousand.[13] There

[11] A county(*Xian*) is the third level of the administrative hierarchy, after provinces and prefectures. There are 2844 counties in total, and 1303 of them are rural counties.

[12] Gazetteers are book-length volumes of local history documenting the country's major events.

[13] The average population size of a village is one thousand during that time, so that I time 1000 to

are two reasons to use the density of doctors as a measure of intensity. First, according to Das et al. (2016), for health workers who work in disadvantaged environments, their time spent with each patient is a better predictor of the health outcomes than quality of diagnoses. Such a relationship also applies to BFDs. Since BFDs' time allocations are not available, the number of BFDs per capita can be used as a proxy. Second, the density of BFDs reflects how much the local government support BFD program, which may be associated with the training quality and medical resources available for the BFDs.

Figure 1 plots the histgram of the county-level BFD densities in the sample. The mean of BFD density is 2.763, and the median of BFD density is 2.6. This is in accordance with other documents that records "there are about 2 to 3 barefoot doctors in each brigade" (Zhang and Unschuld, 2008).

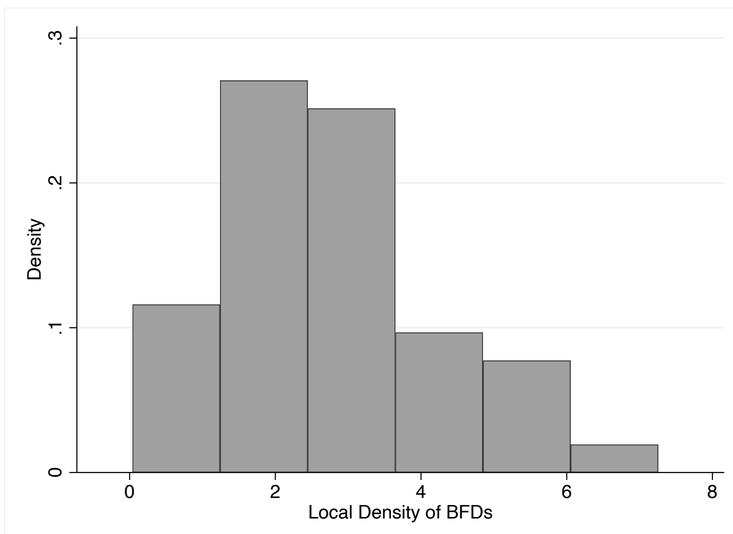


Figure 1: Distribution of local density of BFDs

Third, I collect data on local governments' health investment at baseline. Those variables include the number of hospital beds per capita and number of health workers per

approximate number of BFDs in each village. Usually, BFDs are attached to their villages or production teams, and would only cover patients in their own villages.

capita at the baseline (1960 - 1968). Those data are collected at local prefecture level instead of county level. This is because the support from government came from the prefecture level, usually the cities within the same prefecture. For example, the city sent down their hospital doctors to the rural counties to be BFDs directly, or the hospital doctors went to rural villages to build BFDs' training camps. Counties with different levels of health investments at baseline could trend differently, therefore I use them as controls in the main regressions.

3.2 Census Data

Data from Census of Population Survey in 1964 and 1990 is used for county-level aggregate statistics. Population data from census 1964 is used for calculating the density of BFDs in each county. Data from the 1990 censuses are used as controls for county characteristics at baseline, such as the population ratio of primary school graduates and ratio of junior high school graduates.

3.3 Survey Data

3.3.1 Chinese General Social Survey

Long-term individual outcome data is from the *Chinese General Social Survey*(CGSS) in 2010, 2011, 2012, 2013 and 2015. This is a repeated cross-section dataset. CGSS covers 100 randomly-selected counties in 23 provinces, among which 58 of them are rural counties.[14]

The outcome variables of interest are health, educational attainment and labor market performance. Health outcomes include self-reported health, self-reported mental health, health issues, height and body weight. I use height and body weight to construct variables such as BMI,[15] underweight ($BMI \leq 18.5$) overweight ($24 \leq BMI \leq 28$), and height

[14]In total, there are around 2,850 counties in China, while around 1500 of them are rural counties.

[15]BMI is short for Body Mass Index, which equals to weight (unit:kg) over square of height (unit: cm)

stunting; Education outcomes including highest education level achieved, years of schooling and whether completed middle school. Labor market outcomes including labor force participation, employment status and working hour per week. Demographic features such as gender and ethnicity are used as controls. The descriptive statistics of these variables are in table 10 in the Appendix. Table 10 shows that there's no systematic difference in demographic characteristics between treatment and control group. As for the outcome variables, the exposed group does significantly better in health, education and labor market outcomes. But such difference could be simply due to the fact that the exposed cohorts were born later, and the health and economic conditions have improved overtime.

3.3.2 China In-depth Fertility Survey

Short-term individual outcome data is from the China In-depth Fertility Survey (CIDFS). These are two random surveys that are conducted in 1985 and 1987 by China's National Statistics Bureau. The main subjects are married female under the age of 50. Other household members as well as community officers were also surveyed. Information was collected on the complete reproductive history and households' socioeconomic backgrounds.

The first survey round in 1985 covered Shanghai, Hebei and Shaanxi, and the second round in 1987 covered Guangdong, Shandong, Liaoning, Guizhou, Gansu and Beijing. I exclude Shanghai and Beijing, the two city-governed districts which are more-developed urban areas in my sample. Two other provinces: Gansu and Guizhou are also not included because their raw data is missing in the database. Therefore, the final sample includes five provinces (Hebei, Shaanxi, Guangdong, Shandong and Liaoning). Each province has about 30 counties surveyed, and I only use the rural counties. The final sample contains 95 rural counties. Only a few counties in the CIDFS sample overlapped with counties in CGSS sample. About 6000 households in every province are selected for the survey, with a high completion rate of 98% on average.

For each household, the survey asked the mother for their birth history. I use the mother-birth record as my unit of observation. The outcome variables fall into three categories. The first one is mortality, including prenatal mortality, neonatal mortality, and birth weight. The second category is breastfeeding. I use the months of breastfeeding, as well as a set of dummy variables that indicate the distribution of breastfeed duration as outcomes. The final set of outcomes pertains to health care received. This category includes dummy variables indicating whether the mother received any pregnancy check from professionals, whether the mother was assisted during birth delivery by professionals, whether it was a home delivery, and whether the baby received any vaccination within the first 6 months. More details on the construction of the variables can be found in the Appendix.

3.4 Other Data

Data from previous studies is also used. I use data on the sent-down-youth density in each county shared by Chen et al. (2020) and the intensity of the Cultural Revolution from 1968 to 1971 by Walder (2014, 2019) as controls.

4 Empirical Design

4.1 Difference-in-Differences

A difference-in-differences model with continuous treatment variable is employed for the main regressions. In this model, I compare outcomes for individuals born before and after the introduction of the BFD program (refereed to as the *Unexposed* and *Exposed* accordingly) and then compare the differences between counties with different intensities of the BFD program. The model estimated is shown below:

$$Y_{i,t,c,p,s} = \beta_0 + \beta_1 \%BFD_c \times I(t \in [1970, 1978]) + \mathbf{X}_i + \delta_c + \sigma_t + \gamma_s + \delta_c \times t + \epsilon_{i,t,c,p,s} \quad (1)$$

where i indexes the individual, c the county of birth, t the year of birth, p the province of birth, and s the year of survey. $\%BFD_c$ represents local density of BFDs, calculated using the number of BFDs in a county divided by its population in the 1964 census, times one thousand. $I(t \in [19701978])$ is a dummy variable that equals to 1 if individual i is born between 1970 and 1978, and equals to 0 if born between 1963 and 1969. I use 1970 as the beginning year because they are the first cohort to be exposed to BFD program *in utero*, and I use 1978 as the ending year because it is when the most counties ended the program. My unexposed group here are the cohorts born between 1963 and 1969. I choose this window of control to avoid the influence of China's Great Famine (1959-1961), which had a significant negative effect on health, education, labor market for the survivors in the long term (Meng and Qian, 2009; Chen and Zhou, 2007). I will also use event study later to show that this cutoff is valid.

$\mathbf{X}_i$ is a vector of demographic controls, including individual's gender (male=1) and ethnicity (Han ethnic=1). δ_c stands for county of birth fixed effects, σ_t stands for birth cohort fixed effects, γ_s for survey year fixed effects, and $\delta_c \times t$ stands for county-specific linear time trends. Note that $\%BFD_c$ is captured by county fixed effect and $I(t \in [1970, 1978])$ is captured by cohort fixed effect. Using repeated cross-section data alleviates the confounding age effect, since the surveys took place in different years, and individuals who fall into the same birth cohorts can be at different ages in the sample.

One difficulty in studying the long-term individual outcomes is migration. First, for individuals who emigrate to another county, I cannot observe his or her outcomes. It raises the concern of migration selection, since individuals who are healthier and better educated are more likely to emigrate (Zhao, 1997). To alleviate this concern, I compare places with high/low BFDs densities, and show that the ratios of population who emigrate are not significantly different. Second, 27% of the full samples in CGSS have some migration history (the rest 73% answered "I've been living here since born"). For those who have lived in other places, their long-term outcomes may be significantly affected by their residential counties. To control for this, I restricted the sample to individuals without any migration history.

4.2 Difference-in-Differences with Mother Fixed Effects

For the short-run outcomes, the CIDFS data allows me to use a DID model with mother fixed effects. I compare outcomes for the children born after 1970 to their siblings born before 1970, in counties with different BFD density. Using mother fixed effects can control for unobserved household characteristics such as attitudes towards children's health. The model is presented below:

$$Y_{b,m,t,c,p} = \beta_0 + \beta_1 \% BFD_c \times I(t \in [1970, 1978]) + \phi_m + \sigma_t + \mathbf{X}_b + \delta_c \times t + \epsilon_{b,m,t,c,p} \quad (2)$$

where b indexes the mother-birth record, m the the mother, t the child's year of birth, c the child's county of birth, and p the child's province of birth. $I(b \in [1970, 1978])$, $\delta_c \times t$, σ_t are same to equation (1). ϕ_m represents mother fixed effect. $\mathbf{X}_b$ are demographic controls, which include mother's age category when giving birth to child b (19 to 24 years old; 25 to 34 years old; over 35 years old) and the child's gender, following Currie and Walker (2011). Since the first and second survey rounds of CIDFS covers different counties, the survey round fixed effects are captured by σ_t. I restrict the sample to mothers with pregnancy records in both pre-treatment and post-treatment periods. The one-child policy was introduced after 1978, therefore most households have more than one child during this time period. All standard errors are clustered at county level, and the regressions use the weight given by CIDFS.

4.3 Event Study

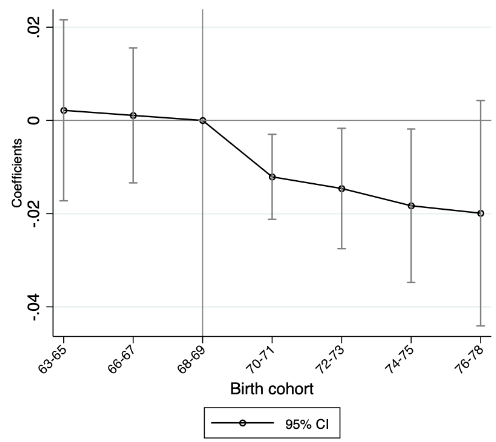
An event model is used for additional specification of the effects of the BFD program for two reasons. First, a key assumption for the DID model is parallel pre-trends. A by-cohort specification would show whether there exists any pre-trends before the introduction of BFD program. Second, the nature of the BFD program is that cohorts in the *Unexposed* group is also affected by the BFDs later in their lives. The event study allows me to explore the timing of BFD exposure and to evaluate the validity of the choice of *Exposed* and *Unexposed* cohorts. In particular, the figures present non-parametrically the relationship between age at the introduction of BFDs and adult outcomes (Hoynes et al., 2016). The event study model is same to equation (1) and (2) except now $I(t \in [1970 - 1978])$ is replaced by a series of dummies for two-year or three-year intervals of birth cohorts (e,g. cohort 1968-69, 1970-71, 1972-73, and so on. I combined the first three years 1963-1965 into one bin and the last three years 1976-1978 into one bin). I combine the cohorts because the number of observations in each single cohort is too small. Cohort 1968-1969 is used as the omitted cohort group. The birth-cohort specific coefficients will be plotted in figures in the next section.

5 Results

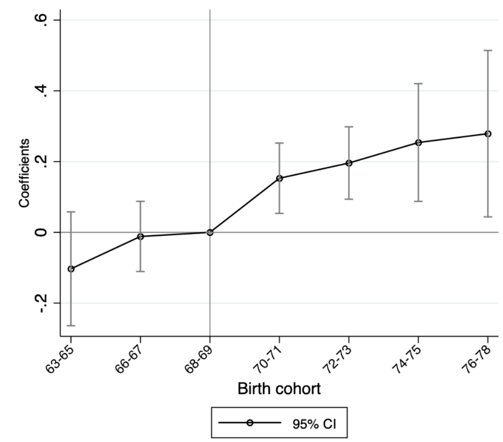
5.1 Event Study Results

Figure 2 presents the event study for selected outcomes, and the event study figures for all the other outcome variables are in the Appendix. In the figures, the vertical axis is the coefficient $\beta_{i,\lambda}$ for every two/three birth cohorts, and horizontal axis is the birth cohort group. The 95% confidence interval for each coefficient point is also plotted in the same graph for reference. Cohorts 1968-1969 is used as the omitted cohort group.

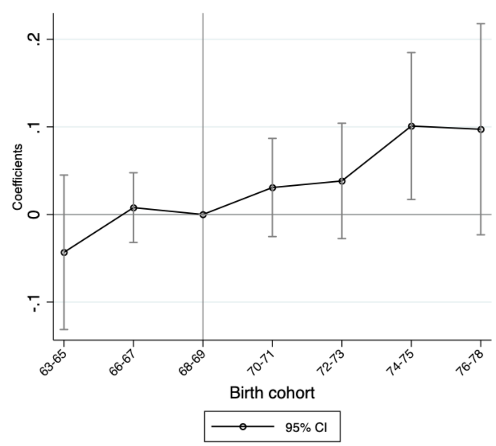
Figure 2 shows that the coefficients for the *Unexposed* cohorts are not significantly different from zero for a 95% confidence interval. This result supports the parallel trend assumption for the difference-in-differences model. In Panel a, b, and d of figure 2, for cohorts who were born right after the introduction of the BFD program, the coefficients for infant mortality, health index, and years of schooling change steadily. In Panel c, there is some upward trend in physical health index after the introduction of BFD program, but not statistically significant. Figure 2 also shows that the largest effects of the BFD treatment are to those who are exposed *in utero* and early childhood. The results suggest that the adult health and economic impacts of the BFDs are not significant if the child is exposed after age three.



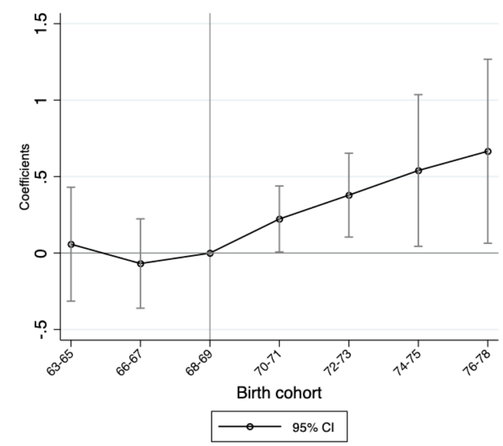
(a) Infant mortality



(b) Self-reported health index



(c) Physical health index



(d) Years of education

Figure 2: Event Study

Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.

5.2 Main Results of DID

This section presents regression results from the difference-in-differences model. The outcome variables fall into five categories: mortality, self-reported health, nutrition, educational attainment and labor market performance.

5.2.1 Birth Outcomes

The first category of results look at mortality and birth weight. China had a very high infant mortality rate during the 1960s and 1970s.[16] According to the World Bank, in 1969 the infant mortality rate was 83.1 per 1000 live births. In my sample, the probability of dying before the age of one is 63 per 1000 live births for the unexposed group. Infant mortality is commonly linked to the health environment during pregnancy (Almond et al., 2011). In contrast, post-neonatal mortality is more determined by post-birth factors such as infectious diseases and accidents, and is more responsive to health care access than neonatal deaths. Results in table 1 show that greater exposure to the BFD program significantly reduced infant and child mortality. Given the mean of BFD density is 2.7 (per 1000 people), being exposed to the BFD program on average reduces infant mortality by 0.029 percentage point ($= 2.7 \times 0.0111$) which is 47.8% of the control mean, and reduces under-5 mortality by about 36%.

I then look at BFDs on probability of low birth weight. Results from table 1 shows statistically insignificant effect on the probability of low birth weight, although the event study figure in the appendix shows a slightly upward sloping trend. The greatly reduced mortality together with slightly increased probability of low birth weight may imply that there might be reverse mortality selection: the marginal child who would have not lived is now surviving because of the BFDs intervention.

[16] Infant mortality rate is defined as the number of infants dying before reaching one year of age, per 1,000 live births in a given year.

Table 1: Exposure to BFDs on birth outcomes

	(1) live $\leq$ 1 yr	(2) live $\leq$ 5 yrs	(3) low birth weight
BFD density	-0.0111**	-0.0106**	-0.000280
* Exposed cohort	(0.00462)	(0.00414)	(0.00493)
Control Mean	0.0606	0.0795	0.174
R-Squared	0.291	0.286	0.569
Obs. Num	14064	14064	14063

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.2.2 Self-reported Health

Now, I am going to look at the long-term outcomes. Results in table 2 show that exposure to BFDs has a positive impact on self-reported health. The effects are both statistically and economically significant. On average, a unit increase in local BFD density will increase the health index by 0.114. To be more specific, being exposed to the BFD program on average raises self-reported general health by 0.28 percentage point ($= 2.7 \times 0.105$). The mean of reported health level for the control group (3.521). This implies there's a 8% increase in the self-reported health level for the exposed cohorts in counties with an average BFD density.

Column 3 in table 2 display the results using the self-reported mental health level as the outcome. Although BFDs did not provide direct mental health treatment to the child, the relationship between early childhood environment and mental health mainly comes from the physical and mental health status of the mother. Zuckerman et al. (1989) show that depressive symptoms among low income pregnant women and mothers have been associated with low nutrition intake, unintended pregnancy, the use of tobacco, alcohol, and illicit drugs. And depressive symptoms can be transmitted to their children (Kahn et al., 2002). The result in column 3 shows that exposure to BFDs on average will improve self-reported mental health level by 8% ($= 2.7 \times 0.114/3.755$).

The outcome in column 4 measures whether the individual has any health issue. This outcome variable can reflect health conditions such as chronic disease history, the condition of the immune system, disability and so on, which can also be greatly influenced by early childhood health (Currie and Stabile, 2006). Results show that greater exposure to BFDs can reduce health issue by about 9% ($= 2.7 \times 0.13/3.827$).

Table 2: Exposure to BFD on Health: Self-reported Health

	(1) Health Index	(2) General Health	(3) Mental Health	(4) Health Issue Free
BFD density	0.114***	0.105***	0.134***	0.130***
* Exposed Cohort	(0.0315)	(0.0335)	(0.0392)	(0.0399)
Control Mean	0.0206	3.521	3.755	3.827
R-Squared	0.123	0.149	0.0833	0.0989
Obs. Num	5110	5675	5124	5115

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

NOTE: For all 4 columns in this table, the higher number means better health. *General Health* is a variable generated from the question "In general, how's your health condition?" ("very good=5" , "good=4", "fair=3", "not good=2" and "very bad=1"). *Mental Health* is from the question "In the past four weeks, how often do you feel depressed or upset? ("never=5" , "very few=4", "sometimes=3", "often=2" , "always=1"). *Health Issue Free* is from the question "In the past four weeks, how often did your health issue affect your working and other daily activities?" ("never=5", "very few=4", "sometimes=3", "often=2" and "always"=1). The health index is calculated using the z-score of all three variables. All standard errors are clustered at county level.

5.2.3 Physical Health

Next, I look at physical health outcomes, including height, the probability of being stunt, BMI, and the probability of having a healthy weight (not underweight or overweight). Results in table 3 show that none of the coefficients are statistically different from zero, although the event study graph shows some upward-sloping trend in physical health index. This result could mean that BFDs have limited effect and do not improve physical health for individuals in the long-run. Physical health measurements are greatly determined by nutrition intake in early years, and nutrition improvement is not amongst the jobs of BFDs. In addition, previous research on early health intervention often finds null results in clinic health measurements when other evidence shows that health conditions are improved (Goodman-Bacon, 2016). However, another possible explanation is the reversed mortality selection. Since BFDs reduce neonatal mortality, it could result in a negative compositional effect from the improved survivability of marginal fetuses (Almond et al., 2011). And this would downward bias the estimates.

Table 3: Exposure to BFD on Health: Physical health

	(1) Physical health index	(2) Height	(3) No stunt	(4) BMI	(5) healthy weight
BFD density	0.00118	-0.0150	-0.00715	0.0147	0.0121
* Exposed Cohort	(0.0218)	(0.236)	(0.0140)	(0.127)	(0.0147)
Control Mean	-0.0000387	164.3	0.904	22.93	0.695
R-Squared	0.237	0.514	0.0681	0.0868	0.0397
Obs. Num	5667	5668	5668	5667	5667

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.2.4 Educational Attainment and Labor Market Outcomes

In this section I'm going to look at the impacts of BFDs on educational attainment and labor market performances in the long-run. Literature shows that early childhood health is strongly correlated with long-term outcomes, such as test scores in school, achieved education level, labor supply, income and welfare receipt in adulthood (Hoynes et al., 2016; Currie et al., 2008; Smith, 2009). Several different mechanisms can explain such relationship. For example, Currie and Stabile (2006); Currie et al. (2010) show that better early childhood physical health would last into adulthood, and better physical health will predict better performance in school and in the labor market. On the other hand, they also find that mental health status in early childhood has a direct impact on human capital through cognitive ability. Other studies also find this relationship can be explained by diversity of nutritional resources (Fogel, 1994), reduced energy and school performance (Bleakley, 2007), and change in parental investment decisions (Pitt et al., 2012).

Table 4 presents the DID results of the effect of the BFD program on educational attainment. The outcome variable in column 1 is a discrete variable that measures the highest education degree achieved (e.g. 0 represents no education, 1 for home education, 2 for primary school, 3 for junior high school, etc). The higher number not only indicates longer years of schooling, but also indicates a more selective degree. For example, a vocational high school degree is 5, while a regular high school degree is 6. Result in table 4 shows that on average, exposure to the BFD program increases the level of education by 7% ($= 2.7 \times 0.101/3.804$). Given the education level is a discrete variable, it is not easy to interpret the result. To better understand the effect of BFDs on education, I also use a continuous variable: years of schooling, which is calculated based on the achieved education level. Results in column 2 show that exposure to BFDs increased the years of schooling by 6.5% ($= 2.7 \times 0.184/7.672$).

As for the labor market performance, I look at both the extensive margin and intensive margin of the impact. For the extensive margin, I find null effects of BFDs on labor force participation and unemployment. This could be explained by the fact that most of the surveyed samples work in their own farm land, and the variation of labor force participation and unemployment is very low. For the intensive margin, on the other hand, result in column 6 of table 4 shows a positive and significant effect on working hours per week. On average, exposure to BFDs will increase the working hours for the exposed group by 4.8 hours per week, which equals to a 13% increase.

Table 4: Exposure to BFD on Education and Labor Market Outcomes

	(1) Educ Level	(2) Years of schooling	(3) Labor force participation	(4) Unemployed	(5) Work hour
BFD density	0.101**	0.184*	0.00485	-0.0111	1.765**
* Exposed Cohort	(0.0474)	(0.0959)	(0.0112)	(0.0208)	(0.846)
Control Mean	3.804	7.672	0.915	0.0872	36.22
R-Squared	0.171	0.194	0.0979	0.0987	0.357
Obs. Num	5677	5677	5678	5204	5382

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.3 Heterogeneity

Previous studies on early childhood health intervention usually find the effects are heterogeneous for different sub-populations, especially between genders. For example, Orr et al. (2003) and Anderson (2008) find larger effects of health intervention for girls than for boys. Furthermore, the pattern of influence can be different between genders. Milligan and Stabile (2009) studied the influence of Canadian child benefits, and find that boys show greater increase in test scores while girls benefit more on mental health and behavioral improvements. In Pitt et al. (2012)'s theoretical model, improved health in childhood lead to more healthy days. In a brawn-based economy, healthy days are allocated differently between genders based on their comparative advantage: boys tend to allocate more on labor, while girls tend to stay in school. Healthier bodies also result in boys dropping school at earlier ages to start working. However, situations might be different in China, where discrimination against girls is severe. With health improvements parents may allocate more resources towards the boys. Therefore, it is unclear how BFDs will affect different genders.

To study that, I split the sample by gender and present the results for the sub-groups in table 5. Panel A shows the results for the female sub-sample and panel B shows the results

Table 5: Heterogeneous Effects of BFDs by Gender

	(1) Reported health Index	(2) Physical health Index	(3) Years of Schooling	(4) Work hour
Panel A: Female Sub-sample				
BFD density	0.145***	0.0270*	0.119	1.110
* Exposed Cohort	(0.0403)	(0.0156)	(0.137)	(1.119)
Control Mean	-0.0750	-0.184	6.766	32.59
Obs. Num	2492	2805	2812	2622
Panel B: Male Sub-sample				
BFD density	0.0928**	-0.0245	0.238**	2.541
* Exposed Cohort	(0.0402)	(0.0367)	(0.110)	(1.629)
Control Mean	0.107	0.172	8.521	39.47
Obs. Num	2618	2862	2865	2760

for the male sub-sample. I find that exposure to the BFD program has a larger effect on the self-reported health and physical health of females, which is in accordance with previous literature. Interestingly, column 3 and column 4 find that the effects are greater on educational attainment and working hours for males. This is different from previous studies, and could be explained by the fact that girls are less likely to be sent to schools for education than boys in rural China. Therefore, when the health status of a boy improves, he might be able to stay in school longer for better education, instead of going to work in the farmland like in other societies. Note that although not statistically significant, the education and labor market improvements for females is still quite large economically. The coefficients are not statistically significant mainly due to higher variance for the females, and could be due to different preferences and family structures of the households.

6 Mechanism

This section is going to look at mechanisms underlying BFDs' impacts. Previous literature have not tested which particular BFD practices improved health. BFDs provided health visits to pregnant mothers, better birth delivery, vaccinations, and breastfeeding advice for new parents. Determining the exact mechanisms through which the BFDs affected health outcomes can help us better understand CHW programs in general.

6.0.1 Breastfeeding and Received health care

One possible mechanism behind the reduction in mortality as well as BFDs' long-term impact is through prolonged breastfeeding. Breastfeeding has been shown to have a significant effect on both short-term and long-term outcomes in developing counties (Victora et al., 2015; Victoria et al., 2000; Botticini et al., 2019). In the short-run, breastfeeding significantly reduces infant mortality from infectious disease (Victoria et al., 2000). MONTEIRO et al. (1990) find that a program which increased duration of breastfeeding from 84 to 146 days led to reductions in deaths caused by diarrhoea of 32%, in respiratory infections of 22% and in deaths due to other infections of 17%. In the long-run, it could affect IQ, educational attainment and income level 30 years later (Victora et al., 2015). There are four main channels that breastfeeding can benefit children's outcomes. First, breastmilk are nutritious. Second, breastfeeding prevents infants from diarrhea-related disease by preventing them from drinking unsanitary water. Third, breastfeeding requires the mother's frequent care of the child. Fourth, prolonged breastfeeding also leads to longer birth intervals, and longer birth spacing can reduce infant and child mortality (Jayachandran and Kuziemko, 2011; Lithell, 1981; Woods et al., 1989; Tu, 1989).

Different from other developing countries, breastfeeding was almost universal and of very long duration in China, especially in the less developed regions (Tu, 1989). On average, mothers in the CIDFS sample breastfed for more than 30 months. The intensity of breastfeeding started to decline in the late 1970s and early 1980s, when formula milk became more accessible the households, and job opportunities opened for females outside household. The breastfeeding rate started to recover after 1992. Even so, breastfeeding duration was especially important in rural China during the sample period, where the population had limited access to modern medical facilities, and infant formulas were not yet introduced. Using a hazard model and data from two regions in CIDFS 1985 (Shanghai and Shaanxi), Tu (1989) find that breastfeeding is significantly associated with infant and child survival in less developed regions in China. He find that exclusive breastfeeding reduces the probability of dying in the first six months by about 70 deaths per 1000 live births.[17] The effects are especially strong in the first five years after birth, but decrease as the child grows older.

Breastfeeding was highly recommended by barefoot doctors, according to *A Manual for Barefoot Doctors*. In the book, it suggests: *"Breastfeeding is the most ideal (way) ......Breastfeeding should start after 24 hours of birth, and feed every four hours. Mothers should try their best to breastfeed at least for the first year of birth."* Therefore, here I use the duration of breastfeeding for each child as outcome.

Results in table 6 show that exposure to BFDs significantly increased duration of breastfeeding. Column 1 finds that on average, exposure to BFDs increased breastfeeding duration by about 4 months (2.7×1.475), which is about 13% of the control mean. I then look at the effect of BFDs on the distribution of breastfeeding duration. First, I create a dummy variable "Ever breastfeed" which equals to 1 if there's any breastfeeding for child b, and

[17] Exclusive breastfeeding is recommended by the World Health Organization and UNICEF for the first six months followed by breastfeeding combined with complementary foods until the age of two. Introduction of complementary foods in the first four months of life is not recommended. Exclusive breastfeeding becomes inadequate after a child reaches 12 months of age, at which point introducing supplemental food into the diet becomes important.

equals to 0 if the mother never breastfeed the child. I find a positive but not statistically significant effect on "Ever breastfeed" in column 2. This could be explained by the fact that 96% of the children in my sample were breastfed, and the variation is very small.

Next, according to *A Manual to Barefoot Doctors*, BFDs should advice mothers to breastfeeding at least for the first 12 months, and encourage exclusive breastfeeding (feeding infants only breast milk) for the first 6 months. In column 3, I create a dummy variable as outcome: whether child b was breastfed more than one year, and I find a significant positive effect on the probability for being breastfed more than one year. In the last column, I look at whether the child received exclusive breastfeeding in his or her first 6 months of life. The result in column 4 shows a significant positive effect for the probability of being exclusively breastfed for the first six months. These results are in accordance with breastfeeding suggestions given by the BFDs.

Table 6: Exposure to BFDs on Breastfeeding

	(1) Breastfeed Duration	(2) Ever breastfeed	(3) Breastfeed $\geq$ 1yr	(4) Exclusive breastfeed 1st 6mo
BFD density	1.475***	0.00809**	0.0152***	0.0149***
* Exposed cohort	(0.320)	(0.00364)	(0.00569)	(0.00460)
Control Mean	30.17	0.960	0.880	0.568
R-Squared	0.452	0.344	0.383	0.737
Obs. Num	14183	14186	14183	14186

As mentioned before, breastfeeding can benefit health through different channels. One testable channel using CIDFS data is the birth interval. I use previous birth intervals (PBI) and interpregnancy interval (IPI) as the outcome variables. The results are in Table 11 in the Appendix. I find BFDs has positive but not statistically significant impact on birth interval. This could be explained by the "Later, Longer, Fewer" family planning policy that was happening at the same time. I will discuss the policy in more details in the

Robustness Checks section. In short, this policy encourages mothers to have longer birth intervals. This could result in longer birth intervals in general, which makes the impacts from BFDs undistinguishable.

Last, I want to look at whether mothers and children receive more health care after the BFD program. I create three dummy variables: "Receive prof preg check" which equals to 1 if the mother received any health check from doctors, nurses, or other health workers; "Home delivery" equals to 1 if the mother gives birth to child b at home; and "Vaccinated 1st 6 months" if the child received vaccination within the first 6 months of birth. The results in Table 7 show that mothers who have greater exposure to BFDs are more likely to receive health checks by about 10.8% ($2.7 \times 0.009/0.224$). But contrary to Hjort et al. (2017) who find that Danish home visiting program significantly reduced home delivery, I find null results on the probability of delivery at home. This could be explained by the fact that many BFDs perform home delivery for mothers instead of sending them to hospitals. Last, I find null result on the probability of getting vaccination within the first six months of birth.

Table 7: Exposure to BFDs on Received Health Care

	(1) Receive prof preg check	(2) Home delivery	(3) Vaccinated 1st 6 months
BFD density	0.00948**	-0.00309	-0.00339
* Exposed Cohorts	(0.00466)	(0.00211)	(0.00686)
Control Mean	0.224	0.901	0.585
R-Squared	0.791	0.761	0.767
Obs. Num	8871	14064	8871

7 Comparing the Effect Size

Comparing the scale of effect to similar early childhood interventions will give us a better understanding of the results of this paper. First, this paper finds similar size of effect on infant and child mortality compared to other CHWs projects. Bjrkman Nyqvist et al. (2019) find that the CHWs program in Uganda reduced the under 5-year child mortality by 27%, infant mortality by 33% and neonatal mortality by 27% after three years. Kidane and Morrow (2000) study the impact of teaching mothers to provide antimalarials to their sick children at home using a community health worker approach, and find a 40% reduction in under 5-year mortality. In India, Baqui et al. (2008) and Kumar et al. (2008) documented a 36% to 54% reduction in neonatal mortality by the introduction of a community-based newborn-care intervention package. In Denmark, Wüst (2012) find the home visiting program increased infant survival rate by 0.5% to 0.8% (5 to 8 lives saved per 1000 live births)

There has been no study on the long-term effect of CHWs yet, so it may be helpful to look at other health interventions at early childhood environment. In Guatemala, a program in the 1970s which gave children access to a protein drink increased educational attainment of female by more than one grade, and increased wage of male by 46% (Hoddinott et al., 2008; Maluccio et al., 2009). In Tanzania, Field et al. (2009) find children born to mother subjected to an iodine supplement program while pregnant attain 0.35 to 0.56 years of additional schooling. In Jamaica, a home visit program provide stunted children not only with nutrition supplements, but also with classes to develop cognitive skill, language, and socioemotional skills. They find that 30 years after the intervention, the treatment group had 43% higher hourly wages and 37% higher earnings than the control group. The Danish home visit program however find very marginal effect on education and labor market outcomes in the long-term (Hjort et al., 2017).

In conclusion, early childhood health environments or interventions could generate long-term effects on a large scale. And the magnitude of effects found in this paper are either similar or more conservative compared to other interventions.

8 Robustness Checks

8.1 Different Specifications

To show that the results are robust to different bundles of fixed effects, tables 8 presents the main estimates from 3 specifications with different combinations of fixed effects. All columns include cohort fixed effects and survey year fixed effects. Column 1 includes province specific linear time trend, and prefecture-level baseline characteristics times a linear time trend. Column 2 includes prefecture specific linear time trends. Column 3 is the main regression model, which includes county linear time trends. I find the estimates are robust to different specifications.

8.2 The "Later, Longer, Fewer" policy

Another contemporaneous event that deserves attention is the family planning policy known as "Later, Longer, Fewer". Started in 1970, this policy aims to encourage people to postpone their first birth, prolong birth intervals, and have fewer children. Previous literature has shown that this policy was effective since 1970 and had a large impact on household's fertility decisions (Chen and Fang, 2021). Suppose places with higher BFD density also did better in promoting this policy, then the observed effects improvements could come from the family planning instead. To see that, I control for the mother's age category when giving birth to the individual (between 19-24, 24-34, and 35 above), and the results are shown in table 9. I find the coefficients are robust compared to the main

Table 8: Robustness - Different Specifications

	(1) prov linear trend +baseline char linear trend	(2) prefect linear trend	(3) county linear trend (main reg)
Self-reported health index	0.0895***	0.0867***	0.114***
	(0.0242)	(0.0264)	(0.0315)
Physical health index	-0.00845	-0.00705	0.00118
	(0.0155)	(0.0183)	(0.0218)
Years of schooling	0.189**	0.258***	0.184*
	(0.0764)	(0.0867)	(0.0959)
Work hours	1.923***	2.118***	1.765**
	(0.670)	(0.781)	(0.846)

regressions.

9 Conclusion

Started in 1969, China's *Barefoot Doctors* is the world's first community health workers program. The wide coverage and massive intensity of the program provides both opportunity as well as empirical difficulties in studying the effects of the program. This paper evaluates the impact of *in utero* and early childhood exposure to the *Barefoot Doctor* program on mortality in the short-run, as well as individual's health, educational and labor market outcomes about 40 years after the program. Using a unique data set collected from China's local gazetteers, I use geographic variation in the intensity of the BFD program across counties and *pre-* / *post-* birth cohorts to construct a difference-in-differences model for identification.

I find that additional exposure to BFDs *in utero* and during early childhood has positive

Table 9: Robustness - Control for impacts of "Later, Longer, Fewer"

	(1) Self-reported health index	(2) Physical health index	(3) Schooling	(4) Work hour
BFD density	0.113***	0.000516	0.167*	1.813**
* Exposed Cohort	(0.0307)	(0.0217)	(0.0993)	(0.842)
age(19-24)	-0.0369	0.0446	0.364	1.032
	(0.0719)	(0.0407)	(0.407)	(2.421)
age(25-34)	-0.121	0.0313	0.310	0.470
	(0.0773)	(0.0425)	(0.362)	(2.405)
age(35+)	-0.132	0.0103	0.0101	2.905
	(0.0953)	(0.0455)	(0.413)	(2.410)
Control Mean	0.0206	-0.0000387	7.672	36.22
R-Squared	0.127	0.238	0.207	0.359
Obs. Num	5110	5667	5677	5382

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

effects in both the short-run and the long-run. On average, greater exposure to BFDs reduced infant mortality and child mortality. In the long-run, exposure to BFDs raised self-reported health, but I find no significant improvement on physical health measurements such as height, body weight and BMI. As for educational outcomes, exposure to BFDs increased achieved education level and increased years of schooling. Last, for the labor market outcomes, BFDs significantly increased working hours per week, while have null effect on unemployment and labor force participation. The size of effects are larger for males in educational attainment and working hours, while larger for females in health outcomes. The scales of effects are similar to other early life health interventions. For the mechanisms behind the results, I find BFDs increased duration of breastfeeding by about 4 months and health care received during pregnancy by 10.8%. In the robustness checks, I rule out other possible explanations for the observed effects, for example the influence of the Cultural Revolution and the family planning policy.

To the best of my knowledge, this paper provides the first empirical evidence on the long-term health and economic benefits of community health workers, and is the first empirical study on China's Barefoot Doctor program. We have to keep in mind that the effects I find are in a context of a time when China's rural population had little access to health care. How the community health workers program compares to other health interventions and the cost-benefit analysis of the program are some topics for future researches.

References

Almond, D., Hoynes, H. W., and Schanzenbach, D. W. (2011). Inside the war on poverty: The impact of food stamps on birth outcomes. The review of economics and statistics, 93(2):387–403.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. Journal of the American statistical Association, 103(484):1481–1495.

Aos, S., Lieb, R., Mayfield, J., Miller, M., and Pennucci, A. (2004). Benefits and costs of prevention and early intervention programs for youth.

Baqui, A. H., El-Arifeen, S., Darmstadt, G. L., Ahmed, S., Williams, E. K., Seraji, H. R., Mannan, I., Rahman, S. M., Shah, R., Saha, S. K., et al. (2008). Effect of community-based newborn-care intervention package implemented through two service-delivery strategies in sylhet district, bangladesh: a cluster-randomised controlled trial. The lancet, 371(9628):1936–1944.

Bharadwaj, P., Løken, K. V., and Neilson, C. (2013). Early life health interventions and academic achievement. American Economic Review, 103(5):1862–91.

Bhutta, Z. A., Lassi, Z. S., Pariyo, G., and Huicho, L. (2010). Global experience of community health workers for delivery of health related millennium development goals: a systematic review, country case studies, and recommendations for integration into national health systems. Global health workforce Alliance, 1(249):61.

Bjrkman Nyqvist, M., Guariso, A., Svensson, J., and Yanagizawa-Drott, D. (2019). Reducing child mortality in the last mile: Experimental evidence on community health promoters in uganda. American Economic Journal: Applied Economics, 11(3):155–92.

Bleakley, H. (2007). Disease and development: evidence from hookworm eradication in the american south. The quarterly journal of economics, 122(1):73–117.

Botticini, M., Eckstein, Z., and Vaturi, A. (2019). Child care and human development: insights from jewish history in central and eastern europe, 1500–1930. The Economic Journal, 129(623):2637–2690.

Brown, D. W., Kowalski, A. E., and Lurie, I. Z. (2015). Medicaid as an investment in children: what is the long-term impact on tax receipts? Technical report, National Bureau of Economic Research.

Case, A., Fertig, A., and Paxson, C. (2005). The lasting impact of childhood health and circumstance. Journal of health economics, 24(2):365–389.

Chen, Y., Fan, Z., Gu, X., and Zhou, L.-A. (2020). Arrival of young talent: The send-down movement and rural education in china. American Economic Review, 110(11):3393–3430.

Chen, Y. and Fang, H. (2021). The long-term consequences of china's later, longer, fewer campaign in old age. Journal of Development Economics, page 102664.

Chen, Y. and Zhou, L.-A. (2007). The long-term health and economic consequences of the 1959–1961 famine in china. Journal of health economics, 26(4):659–681.

Currie, J. and Almond, D. (2011). Human capital development before age five. In Handbook of labor economics, volume 4, pages 1315–1486. Elsevier.

Currie, J., Decker, S., and Lin, W. (2008). Has public health insurance for older children reduced disparities in access to care and health outcomes? Journal of health Economics, 27(6):1567–1581.

Currie, J. and Madrian, B. C. (1999). Health, health insurance and the labor market. Handbook of labor economics, 3:3309–3416.

Currie, J. and Stabile, M. (2003). Socioeconomic status and child health: why is the relationship stronger for older children? American Economic Review, 93(5):1813–1823.

Currie, J. and Stabile, M. (2006). Child mental health and human capital accumulation: the case of adhd. Journal of health economics, 25(6):1094–1118.

Currie, J., Stabile, M., Manivong, P., and Roos, L. L. (2010). Child health and young adult outcomes. Journal of Human resources, 45(3):517–548.

Currie, J. and Walker, R. (2011). Traffic congestion and infant health: Evidence from e-zpass. American Economic Journal: Applied Economics, 3(1):65–90.

Das, J., Holla, A., Mohpal, A., and Muralidharan, K. (2016). Quality and accountability in health care delivery: audit-study evidence from primary care in india. American Economic Review, 106(12):3765–99.

De Geyndt, W., Zhao, X., and Liu, S. (1992). From barefoot doctor to village doctor in rural china.

Deserranno, E. and Qian, N. (2020). Aid crowd-out: The effect of ngos on government-provided public services.

Fang, X. (2012). Barefoot doctors and Western medicine in China. University of Rochester Press.

Field, E., Robles, O., and Torero, M. (2009). Iodine deficiency and schooling attainment in tanzania. American Economic Journal: Applied Economics, 1(4):140–69.

Fitzsimons, E. and Vera-Hernández, M. (2013). Food for thought? breastfeeding and child development. Technical report, IFS Working Papers.

Fogel, R. W. (1994). Economic growth, population theory, and physiology: The bearing of long-term processes on the making of economic policy1. The American Economic Review, 84(3):369–395.

Gertler, P., Heckman, J. J., Pinto, R., Chang, S. M., Grantham-McGregor, S., Vermeersch, C., Walker, S., and Wright, A. (2021). Effect of the jamaica early childhood stimulation intervention on labor market outcomes at age 31. Technical report, National Bureau of Economic Research.

Gilmore, B. and McAuliffe, E. (2013). Effectiveness of community health workers delivering preventive interventions for maternal and child health in low-and middle-income countries: a systematic review. BMC public health, 13(1):847.

Goodman-Bacon, A. (2016). The long-run effects of childhood insurance coverage: Medicaid implementation, adult health, and labor market outcomes. Technical report, National Bureau of Economic Research.

Hjort, J. et al. (2017). Universal investment in infants and long-run health: evidence from denmark's 1937 home visiting program. American Economic Journal: Applied Economics, 9(4):78–104.

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., and Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. The lancet, 371(9610):411–416.

Hoehn-Velasco, L. (2021). The long-term impact of preventative public health programs. The Economic Journal, 131(634):797–826.

Hoynes, H., Schanzenbach, D. W., and Almond, D. (2016). Long-run impacts of childhood access to the safety net. American Economic Review, 106(4):903–34.

Hsiao, W. C. (1984). Transformation of health care in china. New England Journal of Medicine, 310(14):932–936.

Jayachandran, S. and Kuziemko, I. (2011). Why do mothers breastfeed girls less than boys? evidence and implications for child health in india. The Quarterly journal of economics, 126(3):1485–1538.

Kahn, R. S., Zuckerman, B., Bauchner, H., Homer, C. J., and Wise, P. H. (2002). Womens health after pregnancy and child outcomes at age 3 years: a prospective cohort study. American Journal of Public Health, 92(8):1312–1318.

Kidane, G. and Morrow, R. H. (2000). Teaching mothers to provide home treatment of malaria in tigray, ethiopia: a randomised trial. The lancet, 356(9229):550–555.

Kumar, V., Mohanty, S., Kumar, A., Misra, R. P., Santosham, M., Awasthi, S., Baqui, A. H., Singh, P., Singh, V., Ahuja, R. C., et al. (2008). Effect of community-based behaviour change management on neonatal mortality in shivgarh, uttar pradesh, india: a cluster-randomised controlled trial. The Lancet, 372(9644):1151–1162.

Lin, Y. (2020). Unleash the pandora's box: Political turmoil and malaria outbreak during china's cultural revolution. Australian Economic History Review, 60(2):220–249.

Lithell, U.-B. (1981). Breast-feeding habits and their relation to infant mortality and marital fertility. Journal of Family History, 6(2):182–194.

Llor, C. and Bjerrum, L. (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Therapeutic advances in drug safety, 5(6):229–241.

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., and Stein, A. D. (2009). The impact of improving nutrition during early childhood on education among guatemalan adults. The Economic Journal, 119(537):734–763.

Meng, X. and Qian, N. (2009). The long term consequences of famine on survivors: evidence from a unique natural experiment using china's great famine. Technical report, National Bureau of Economic Research.

Miguel, E. and Kremer, M. (2004). Worms: identifying impacts on education and health in the presence of treatment externalities. Econometrica, 72(1):159–217.

Milligan, K. and Stabile, M. (2009). Child benefits, maternal employment, and children's health: Evidence from canadian child benefit expansions. American Economic Review, 99(2):128–32.

MONTEIRO, C., REA, M., and VICTORA, C. (1990). Can infant mortality be reduced by promoting breastfeeding? Evidence from So Paulo city. Health Policy and Planning, 5(1):23–29.

Orr, L., Feins, J., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., Liebman, J. B., and Kling, J. R. (2003). Moving to opportunity: Interim impacts evaluation.

Perry, H. B., Zulliger, R., and Rogers, M. M. (2014). Community health workers in low-, middle-, and high-income countries: an overview of their history, recent evolution, and current effectiveness. Annual review of public health, 35:399–421.

Piazza, A. (2019). Food consumption and nutritional status in the PRC. Routledge.

Pitt, M. M., Rosenzweig, M. R., and Hassan, M. N. (2012). Human capital investment and the gender division of labor in a brawn-based economy. American Economic Review, 102(7):3531–60.

Singh, P. and Sachs, J. D. (2013). 1 million community health workers in sub-saharan africa by 2015. The Lancet, 382(9889):363–365.

Sloan, N. L., Ahmed, S., Mitra, S. N., Choudhury, N., Chowdhury, M., Rob, U., and Winikoff, B. (2008). Community-based kangaroo mother care to prevent neonatal and infant mortality: a randomized, controlled cluster trial. Pediatrics, 121(5):e1047–e1059.

Smith, J. P. (2009). The impact of childhood health on adult labor market outcomes. The review of economics and statistics, 91(3):478–489.

St James, P. S., Shapiro, E., and Waisbren, S. E. (1999). The resource mothers program for maternal phenylketonuria. American journal of public health, 89(5):762–764.

Tu, P. (1989). The effects of breastfeeding and birth spacing on child survival in china. Studies in family planning, 20(6):332–342.

Victora, C. G., Horta, B. L., De Mola, C. L., Quevedo, L., Pinheiro, R. T., Gigante, D. P., Gonçalves, H., and Barros, F. C. (2015). Association between breastfeeding and intelligence, educational attainment, and income at 30 years of age: a prospective birth cohort study from brazil. The lancet global health, 3(4):e199–e205.

Victoria, C. et al. (2000). Effect of breastfeeding on infant and child mortality due to infectious diseases in less developed countries: a pooled analysis. Lancet (British edition), 355(9202):451–455.

Walder, A. G. (2014). Rebellion and repression in china, 1966–1971. Social Science History, 38(3-4):513–539.

Walder, A. G. (2019). Agents of Disorder: Inside Chinas Cultural Revolution. Harvard University Press.

Wherry, L. R. and Meyer, B. D. (2016). Saving teens: using a policy discontinuity to estimate the effects of medicaid eligibility. Journal of Human Resources, 51(3):556–588.

Woods, R. I., Watterson, P. A., and Woodward, J. H. (1989). The causes of rapid infant mortality decline in england and wales, 1861–1921. part ii. Population studies, 43(1):113–132.

Wüst, M. (2012). Early interventions and infant health: Evidence from the danish home visiting program. Labour Economics, 19(4):484–495.

Xu, S. and Hu, D. (2017). Barefoot doctors and the health care revolution in rural china: a study centered on shandong province. Endeavour, 41(3):136–145.

Young, M. E. (1989). Impact of the rural reform on financing rural health services in china. Health Policy, 11(1):27–42.

Zhang, D. and Unschuld, P. U. (2008). China's barefoot doctor: past, present, and future. The Lancet, 372(9653):1865–1867.

Zhao, Y. (1997). Labor migration and returns to rural education in china. American Journal of Agricultural Economics, 79(4):1278–1287.

Zuckerman, B., Amaro, H., Bauchner, H., and Cabral, H. (1989). Depressive symptoms during pregnancy: relationship to poor health behaviors. American journal of obstetrics and gynecology, 160(5):1107–1111.

10 Appendix

10.1 Data Appendix

This section summarizes the data used in the analysis. In this data appendix, additional information on the various data source as well as on the construction of variables is provided.

10.1.1 China General Social Survey

The long-term individual outcome variables are generated from CGSS. The data quality of CGSS is widely recognized. For a long time, the CGSS is the only large-scale survey project with open data in China mainland. The CGSS data has become the most important data resource for studies on China. By December 2014, there are more than 20,000 registered users and there are more than 700 journal papers published based on CGSS data.

I use data from CGSS to generate four categories of variables: self-reported health, physical health, educational attainment, and labor market performance. The self-reported health outcome variables are described in section 5. For physical health, the standards for underweight and overweight come from General Administration of Sport of China at http://www.sport.gov.cn/n16/n1077/n1422/7331093.html. Height stunting is defined as shorter than two standard deviations of the average height of each gender. Height distribution data is from General Administration of Sport of China at http://www.sport.gov.cn/n16/n1077/n1422/7331

As for educational attainment, I calculated the years of schooling using the highest degree achieved. For example, if the individual's highest education achieved is "primary school", then years of schooling is 6 years; if the highest level is "middle school" then years of schooling equals to 9 years. Note that from 1969 towards the end the Cultural Revolution, the 6-3-3 education system (6 years of primary school, 3 years of middle school, and 3 years of high school) was changed to 5-2-2 in many places. It restored to 6-3-3 system gradually after the end of the Cultural Revolution: the duration of junior high school started to resume to 3 years in 1978 in most regions while the length of senior high school was increased to 3 years over the period 1981-1985. This change of education system would inflate the calculated years of schooling for cohorts in the control group more, while cohorts in the exposed group are less likely to be affected. This would potentially underestimate the results on years of schooling.

10.1.2 China In-Depth Fertility Survey (CIDFS)

Data from CIDFS (1985, 1987) is used for short-term outcome variables. CIDFS is China's first household-level sample survey on fertility. The design of the survey is similar to standard fertility surveys used in many other countries, and the survey was conducted with the help from the International Statistics Research Center. The quality of CIDFS was recognized by researchers (Lin, 2020).

Two provinces in the survey: Gansu and Guizhou are not included in the analysis because their raw data is missing in the database. This is unfortunate since they are less developed provinces compared to others in CIDFS, and the BFD program is expected to have a larger impact there. According to the report of CIDFS 1987, Guizhou has the highest infant mortality rate of 70.6 per 1000 live births, and Gansu has the second highest infant mortality rate of 44.6 per 1000 live births. Meanwhile, their mortality rates declined drastically from the 1960s to the 1980s.

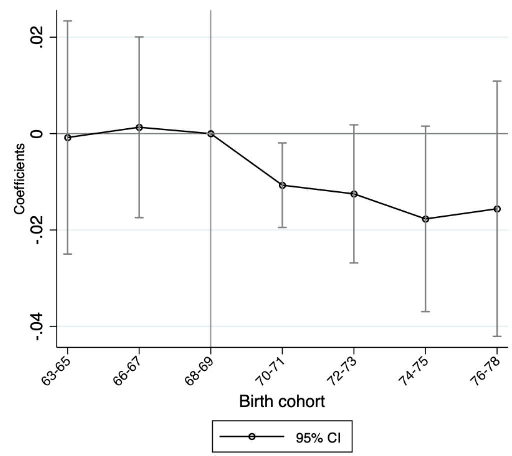
I use CIDFS to create prenatal and neonatal health outcomes. In CIDFS, the survey asks whether each child is still alive; if not, the age (year, month, day) of death. I use these two questions to construct a bunch of dummy variables that indicate whether the infant lived less than one day, less than three months, less than 12 months, less than 3 years, and less than 5 years. As for the prenatal mortality, the survey asks whether the pregnancy outcome is a live birth; if not, how long is the pregnancy before termination. Based on these two questions, I create a dummy variable "non-live birth" indicating whether the pregnancy outcome is a non-live birth, and a dummy variable "miscarriage" if the pregnancy is terminated before 5 months (20 weeks). I also use continuous birth weight, and a dummy indicator of "low birth weight" (birth weight less than 2500 grams, which is about 5.5 pounds) as outcomes. As for breastfeeding, the survey asks how long did the mother breastfeed each baby. I use the months of breastfeeding, as well as a set of dummy variables that indicate whether the baby was breastfed for the first 6 months, for the first year, and for the second year.

10.2 Descriptive Statistics

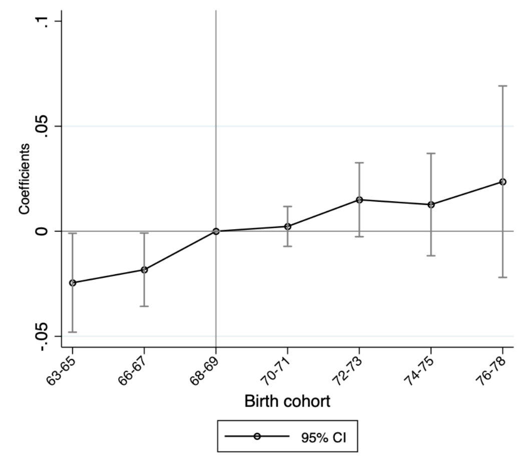
Table 10: Descriptive Statistics

Variables	Unexposed		Exposed		
	Mean	Obs.	Mean	Obs.	Mean diff
male =1	6213	0.511	6239	0.506	0.00500
Han ethnic =1	6213	0.909	6239	0.912	-0.00200
self-reported health	6206	3.571	6234	3.807	-0.235***
self-reported mental health	5600	3.829	5550	3.925	-0.096***
health issue free	5588	3.942	5549	4.161	-0.219***
height	6196	164.3	6222	164.9	-0.578***
no stunt	6196	0.907	6222	0.925	-0.018***
BMI	6193	23.12	6221	22.87	0.250***
underweight	6193	0.0510	6221	0.0570	-0.00600
overweight	6193	0.254	6221	0.231	0.023***
education level	6210	4.413	6238	5.026	-0.614***
years of schooling	6210	8.588	6238	9.460	-0.872***
midschool complete	6213	0.646	6239	0.711	-0.065***
labor force participation	6213	0.907	6239	0.914	-0.00700
unemployed	5636	0.0980	5702	0.0750	0.023***
work hour per week	5819	34.37	5974	35.28	-0.913*

10.3 Additional Event Study Figures



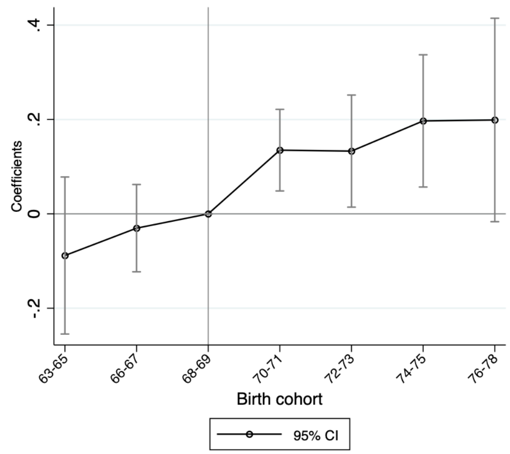
(a) Under-5 mortality



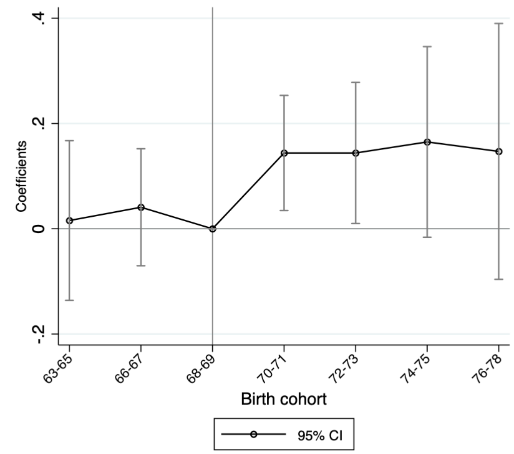
(b) Low birth weight

Figure 3: Event study for each outcome variable(1)

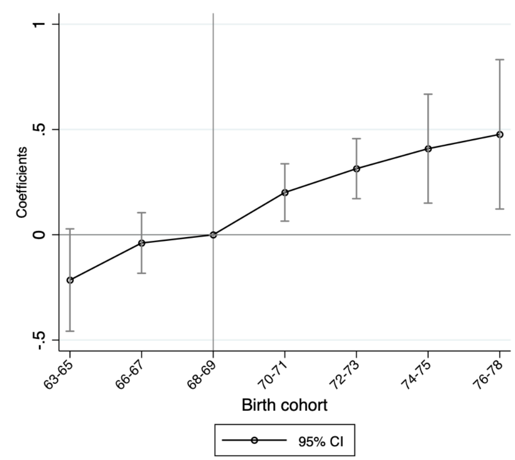
Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.



(a) Reported Health Level



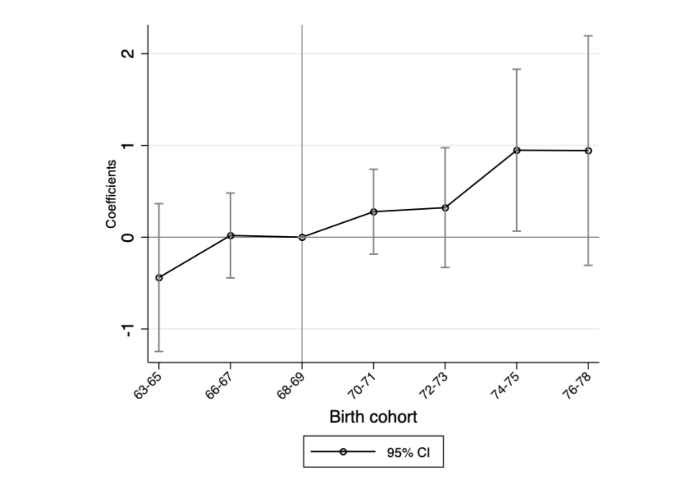
(b) Mental Health Level



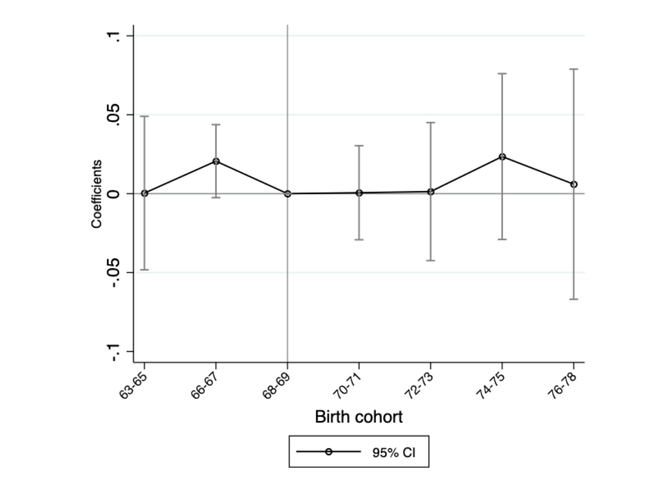
(c) Health Issue Free

Figure 4: Event study for each outcome variable(2)

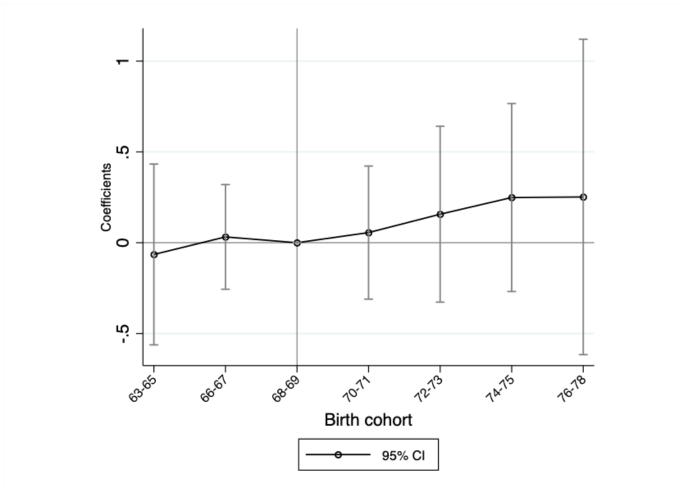
Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.



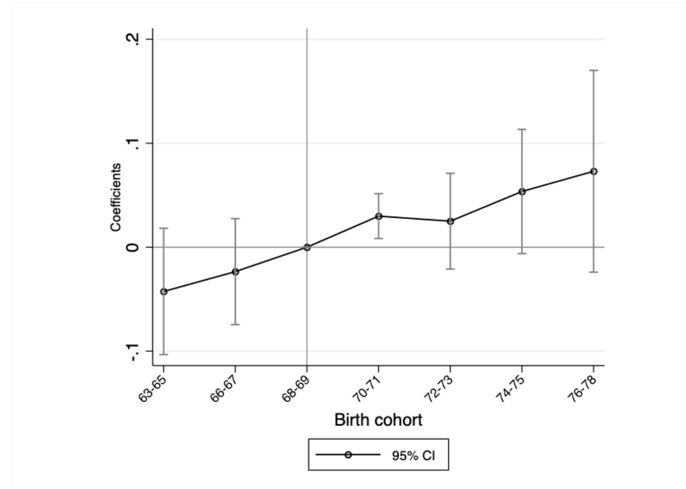
(a) Height



(b) No Stunt



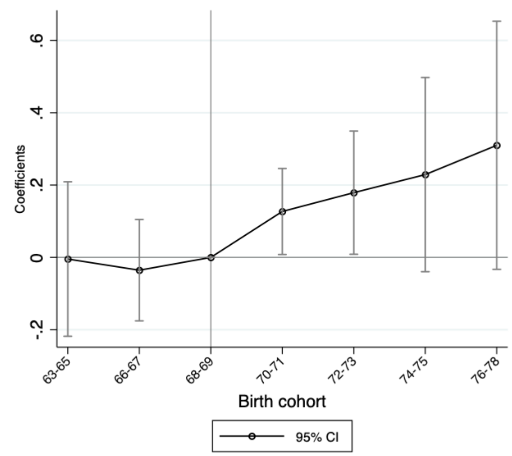
(c) BMI



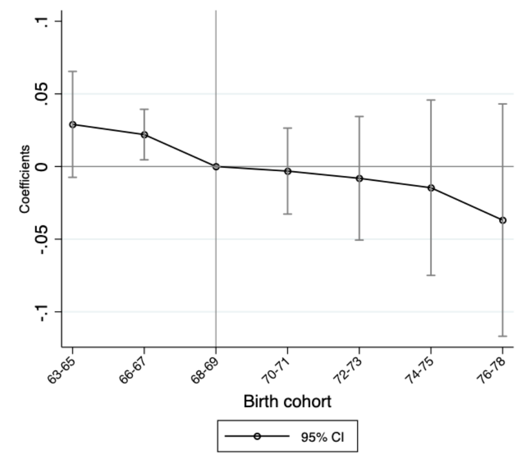
(d) Healthy weight

Figure 5: Event study for each outcome variable(3)

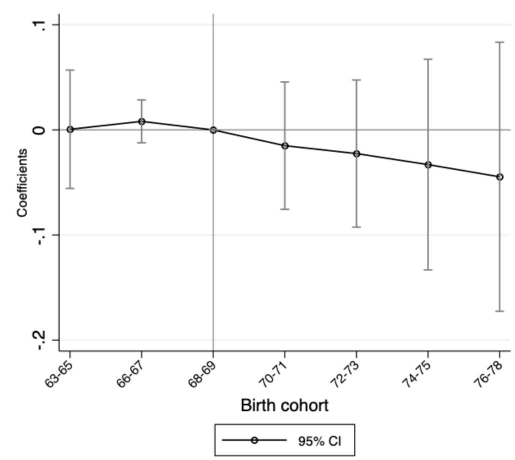
Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.



(a) Education Level



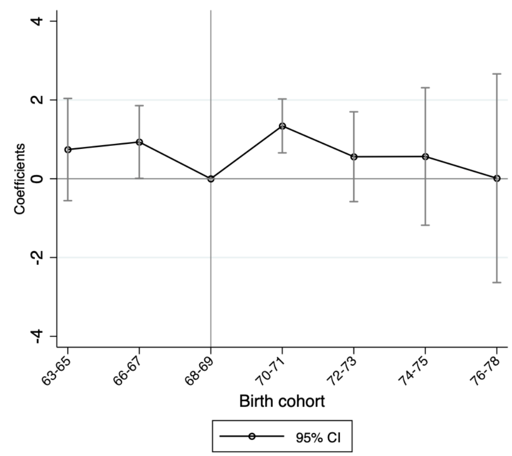
(b) Labor Force Participation



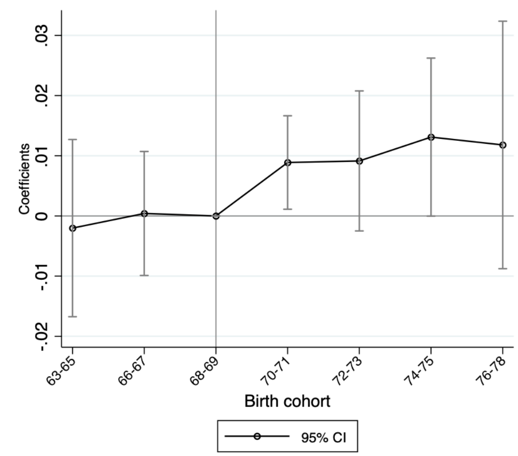
(c) Unemployment

Figure 6: Event study for each outcome variable(4)

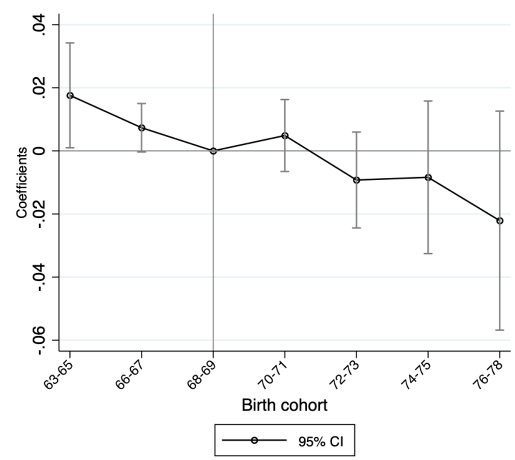
Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.



(a) Breastfeed Duration



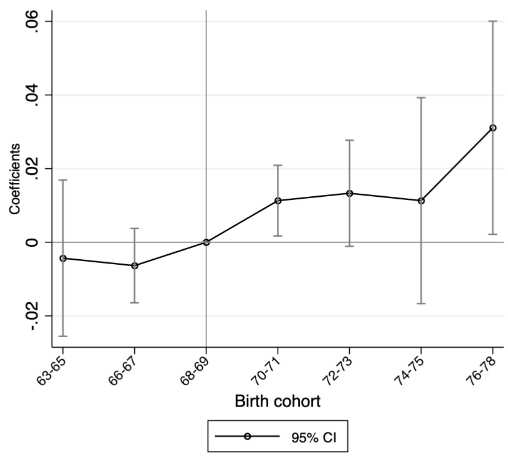
(b) Ever breastfeed



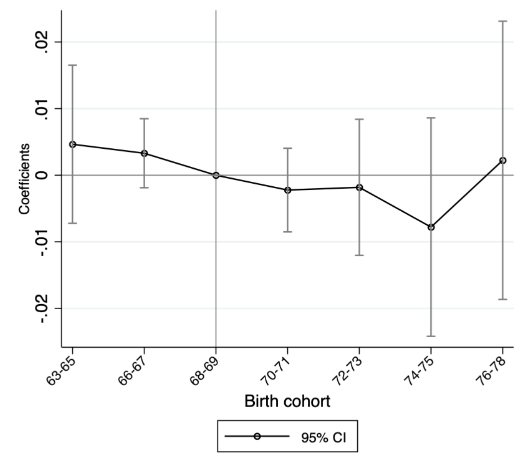
(c) Breastfeed more than 1 year

Figure 7: Event study for each outcome variable(5)

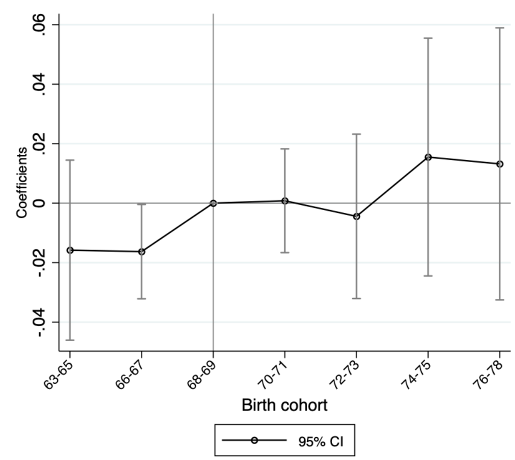
Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.



(a) Received pregnancy check



(b) Home delivery



(c) Received vaccination within 6 months

Figure 8: Event study for each outcome variable(6)

Note: The y axis represents the coefficients from equation (1), which captures the effect of BFDs densities on different cohorts. The x axis represents birth cohorts every two years.

10.4 Figures

10.5 Tables

10.5.1 The impacts of BFDs on Birth Interval

Table 11 presents the results using birth interval as outcome variable. "Birth interval post" is calculated using the year-month of birth of child b and his or her successive sibling, and "Birth interval pre" is calculated using the year-month of birth of child b and his or her previous sibling. Twins are excluded from the sample. The children who are the youngest of all siblings in the sample period are excluded from column 1 and column 2's regressions, and the children who are the firstborns are excluded from column 3 and column 4's regressions.

Table 11: BFDs on Birth Intervals

	(1) PBI	(2) ln(PBI)	(3) IPI	(4) ln(IPI)
BFD density	0.350	0.0102	0.155	0.00929
* Exposed cohort	(0.244)	(0.00665)	(0.389)	(0.0119)
Control Mean	32.81	3.404	29.83	3.318
R-Squared	0.408	0.401	0.493	0.482
Obs. Num	9990	9988	6586	6585

10.6 Additional Robustness Checks

10.6.1 The Interruption of BFD program during 1972-1973

As mentioned in the background section, the BFD program was temporarily interrupted between 1972 to 1973. The Cooperation Medical System was suspended, many village clinics were closed and the training of BFDs was stopped. Therefore, if the health improvements really came from the practice of BFDs, it should predict that cohorts who were in utero

Table 12: Robustness - Interruption during 1972-1973

	(1) Self-reported health index	(2) Physical health index	(3) Schooling	(4) Work hour
BFD density	0.121***	0.0000413	0.188*	1.409
* Exposed Cohort	(0.0324)	(0.0203)	(0.104)	(0.913)
Control Mean	0.0206	-0.0000387	7.672	36.22
R-Squared	0.125	0.244	0.197	0.362
Obs. Num	4513	4992	4999	4731

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

and born between 1972 to 1973 (which are individuals born between 1973 and 1974) have less health improvements compared to other cohorts in the exposed group. We should also expect that after excluding those cohorts the effect sizes will go down. Table 12 reports the results after excluding cohorts who were born in 1973 and 1974 from the exposed group. Results from Table 12 show that compared to previous main results, the coefficients now become larger and more significant. These results provides extra supportive evidence that the BFD program does have effects on the health and economic outcomes.