# The Long-Term Impact of Community Health Workers on Health and Economic Outcomes: Evidence From China's *Barefoot Doctor* Program *

Ying LIU†

May 20, 2020

**Abstract**

Community Health Worker projects (CHWs) in developing regions train local residents with basic medical skills to provide primary health care for their neighborhoods. While short-term benefits of CHWs have been documented, little is known about the long-term effects because of data limitations. As the world's first CHW project, China's *Barefoot Doctor* program (1969-1985) provides an opportunity to study long-run effects. In this paper, I manually collect a unique county-level dataset on the *Barefoot Doctor* program from China's historical gazetteers. From the data, I exploit geographic variation in the program's intensity across counties. Combined with the introduction of the program in 1969, I employ a fuzzy difference-in-difference model to identify the program's influence. I use Chinese General Social Survey (CGSS) data (2010-2015) on individuals health, education and labor market outcomes. Results show that exposure to the *Barefoot Doctor* program *in utero* and during early childhood has positive effects on adult health and economic outcomes. On average, a 50% increase in the number of BFDs raises reported health by 9.4%, and reduces the probability of depression and serious health problems by 8.5% and 8% respectively. It also reduces unhealthy behaviors like smoking and heavy drinking, but I find no significant improvement on nutrition. As for economic outcomes, a 50% increase in number of BFDs would increase years of schooling by 2.4% and middle school completion rate by 7.4%, and increase labor supply by 4.1%. The size of effects are larger for males in educational attainment and labor supply, while larger for females in health.

**Keywords:** Community Health Workers (CHWs), health, human capital, China

*This paper is still preliminary and incomplete. Please do not cite without the author's permission

†Department of Economics, University of Pittsburgh. Email: ying.liu@pitt.edu

# 1 Introduction

A shortage of health workers is a prevalent problem worldwide. It was estimated by the World Health Organization (WHO) that there will be a gap of 18 million health workers by 2030, the bulk of which will be in developing countries.[1] A potential solution is to use community health workers (CHWs), who have played a major role in expanding 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 length of training can be as short as one week and cost as low as 3 dollars per person per year. CHWs represent a trade-off between coverage and quality: they reach a wider population with lower costs, but the quality of care may fall below formal standards. While the short-run health benefit of CHWs has been documented (Bjrkman Nyqvist et al., 2019; Postma et al., 2009; Roman et al., 2009; Spencer et al., 2011), little is known about their long-run impacts. Early improvement in health can benefit later outcomes (Currie and Stabile, 2003; Case et al., 2005; Miguel and Kremer, 2004).[2] However, substandard quality of care, for example in the form of high diagnostic error and over-prescription of antibiotics, may be problematic in the long-run (Llor and Bjerrum, 2014).[3] In addition, CHWs may crowd out formal medical treatment in rural areas. Therefore, it is important to empirically estimate the long-run impacts of CHWs.

The main challenge in studying long-run effects is data limitations. In most countries CHWs have only existed for less than 20 years. One exception is China's *Barefoot Doctor* (BFD) program, which was the world's first CHW program. Started in 1969, the BFD program recruited more than 1.3 million villagers for medical training, most of whom were secondary school graduates. The average length of training was 3 months, during which they learned how to diagnose common diseases, deliver birth, provide first-aids, etc. This program covered more than 85% of villages in China and was recognized as a great success by the WHO. From 1969 to 1978, China's children mortality rate declined from 117.2 to 69.6 (per 1000 live birth), infant mortality rate dropped from 82.9 to 52.6 (per 1000 live births), and life expectancy at birth increased from 57.6 to 65.9 years.[4] However, after China's economic reform in 1978, a large number of BFDs started migrating to

[1] See WHO report: https://www.who.int/hrh/news/2019/health-worker-momentum-wha72/en/

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

[3] See Fang (2012)'s interviews with barefoot doctors in *Zhejiang* province in China.

[4] Data from: World Bank, https://data.worldbank.org

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 the long-run outcomes of their patients. One challenge for identification is that the BFD program was nationwide: all provinces were affected Therefore, one cannot compare places that were "treated" and "not treated" to estimate the treatment effect of the BFD program. One potential strategy would exploit variation in the intensity of the program across regions. However, existing data on the BFD program is only available at the national level, and is unable to reveal such variation.

To overcome this challenge, I will compile a unique county-level dataset on the program. County-level BFD program data can be found in China's historical gazetteers, which I manually collected and digitized. [5] The resulting dataset contains information on the number of BFDs in each county in different years. From the data, I find that the density of BFDs varies across counties, which can be used as a measure of variation of program's intensity. Combining it with the introduction of BFD program in 1969, I am able to construct a difference-in-difference model for identification. I control for individual demographics, family backgrounds, cohort fixed effect, county fixed effect and provincial specific time trends in regressions. My outcome variables including health, educational attainment and labor market outcomes, such as self-reported health, nutrition, years of schooling, unemployment and work hours. These outcomes data come from a series of randomly sampled national surveys: *Chinese General Social Survey* from 2010 to 2015. Data from *Census of Population Survey* (CPS) in 1982 and 1990 is also used for county characteristics. I follow Kling et al. (2007) to create summary standardized indices for each category of outcome and estimate DID model on them. Note that since the cohorts born before the program were also treated by the BFDs later in their lives, my estimates will be interpreted as the effect of *additional* years of exposure to BFDs both *in utero* and in early childhood.

A key assumption of my identification is that the allocation of BFDs should not be endogenous to outcomes. I will justify this assumption using both historical facts and empirical evidence from event studies. First, From gazetteers' records, BFDs were recruited from two kinds of people: youth with some education, which is the majority of them; and existing village doctors. Among the youths that were picked, most of them are so-called *Sent-Down Youths* (SDYs). More historical background of SDYs can been seen in section 6. Generally, they are urban youth, typically junior and senior

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

high school graduates, who were sent to rural villages under the command of central government during 1962 to 1976. Because they are more educated than other villagers, they are more likely to get selected into BFD training. Therefore, the number of BFDs is partly determined by number of SDYs in each county. The allocation of SDYs is provincial-level governments' decision, and when the allocation decisions were made, health and education are not their consideration (Chen et al., 2018). It is determined to some extent by the county level time-invariant characteristics such as distance to city (You, 2019). Therefore in the event studies, after controlling for county fixed effect and provincial specific time trends, I show that there's no pre-trend difference across counties with different BFD densities.

Results show that additional exposure to BFDs *in utero* and during early childhood has positive effects in the long-run. On average, a 50% increase in the number of BFDs raises reported health by 9.4%, and reduces the probability of depression and serious health problems by 8.5% and 8% respectively. It also reduces unhealthy behaviors like smoking and heavy drinking, but I find no significant improvement on nutrition. As for economic outcomes, a 50% increase in number of BFDs would increase years of schooling by 2.4% and middle school completion rate by 7.4%, and increase labor supply by 4.1%. The size of effects are larger for males in educational attainment and labor supply, while larger for females in health. After excluding cohorts born between 1973-1974, when the BFD program was interrupted, the sizes of effects are even larger. A placebo test later shows that those effects cannot be explained by other overlapping events like cultural revolution.

An important contribution of this research is to provide the first evidence of the long-run impact of CHWs. Existing literature on CHWs has focused on the short-run effects. For example, Bjrkman Nyqvist et al. (2019) find that three years after introduction of community health providers in Uganda, children mortality rate was reduced by 27%, infant mortality rate decreased by 33%, and neonatal mortality rate by 28%. CHWs in the U.S has been documented to have reduced chronic and mental disease such as depression (Roman et al., 2009), asthma (Postma et al., 2009), diabetes (Spencer et al., 2011), hypertension (Brownstein et al., 2007) and cardiovascular disease (Balcazar et al., 2006). See Bhutta et al. (2010) and Gilmore and McAuliffe (2013)'s papers for systematic review on CHWs case in developing countries. 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 me to look at other outcomes like education and labor market performance.

This research also evaluate the effects of "low-tech" innovations in developing countries. Low-tech innovations are affordable technologies and methods that are substitutes for their high-tech alternatives. One example is the *Kangaroo Mother Care* (KMC) for preterm birth infants. In the U.S, preterm babies will be put into incubators, which costs $51,600 per person. Much cheaper is the KMC method, in which mothers can use a piece of cloth to keep the skin-to-skin contact of the baby, and it reduces preterm infant mortality rate by 36% (Lawn et al., 2010). A similar example is the clean delivery kit in South Asia, which provides mothers with hygienic tools and instructions for home delivery. Seward et al. (2012) show the usage of such kits is associated with a 48% reduction in neonatal morality, while each kit only costs 3 dollars. Besides health care, one similar case is *Balsakhis* in India, who are young women providing helps for students' education. They are much less educated than the teachers in private or public schools, many of them had only 10 years of schooling and one week of Balsakhis training. And they raised students' test scores by almost one fold (Banerjee et al., 2007).

Last but not least, this research adds to the literature on the long-term influence of early childhood interventions. [6] Studies on early childhood interventions often find mixed results. For example, Hoynes et al. (2016) show *Food Stamps* bring large improvements on health and economic outcomes decades after exposure. Likewise, Wherry and Meyer (2016); Brown et al. (2015) find U.S.'s *Medicaid* program has positive effects on long-term health. However, a study on a home visiting program in the U.S. shows on average it had small effects on health and was not worth the cost (Aos et al., 2004). Note that studies on effects of early childhood interventions usually find heterogeneous effects across genders. For example, Anderson (2008) studied the Head Start program in the U.S and find the effects are largely concentrated on girls. This paper will supplement the above papers by providing a new evidence of positive health intervention's long-lasting impact in a developing economy setting.

The reminder of the paper is structured as follows. Section 2 will introduce the historical and institutional background of BFD program. Section 3 describes the data and variables used for empirical analysis. Section 4 shows empirical design. Section 5 presents the results of regressions.

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

Section 6 discusses possible treat to identifications. Section 7 is the conclusion.

# 2 Historical Background

## 2.1 Barefoot Doctors Program

**Innovation by local government** At the beginning of the 1950s, China's rural population barely had access to primary medical care. There were just 1.4 million certified doctors in China, of which over 80% were located in urban areas. Meanwhile, 80% of the whole population lived in rural areas (De Geyndt et al., 1992). In 1965, after a speech by Mao Zedong, [7] local governments began experimenting with different methods to improve rural healthcare. In 1968, a method called *barefoot doctors*, which was pioneered by some villages in Hubei province, attracted the central government's attention. The name comes from the fact that these doctors also participated in farm work, which they tended to complete barefoot. In fact, barefoot doctors spent about half of their time on medical care and half on farming (De Geyndt et al., 1992). On September 1968, Mao noticed this new method and asked local governments to adopt similar interventions. Soon after, the BFD program was introduced to the whole nation.

**Timeline of BFD program** Figure 1 shows the density of BFDs in selected counties between 1968 and 1985. Only counties with at least 4 separate observations are included. Except for several pioneer counties in Hubei province, the majority of the counties started the BFD program in 1969. The BFD program was managed by each village's *cooperative medical system*. Usually, a village is one unit of brigade. [8] Every villager paid an annual participation fee of 1 to 3 yuan (about $0.50 to $1.50 based on exchange rate in 1970) into the cooperative medical system. This was equivalent to 1% to 3% of a family's disposable income (Zhang and Unschuld, 2008). This fund was then used to pay for medical drugs, facilities and to compensate BFDs.

The initial expansion of BFD programs was drastic and radical. For example, many counties

[7] His speech is later referred to as *6.26 Instruction*

[8] Brigade is a special terminology used in Maos time. People in one village usually form into one brigade. In each brigade, farmers are organized to work together, and sharing their gains as well as all other resources. Instead of monetary income, the brigade records each persons work points for his workload, and the reallocation of resources are based on the amount of points earned

[9] Data source: China's gazetteers. Each line represents a different county. Only counties with a balanced panel data on the number of barefoot is included in this figure

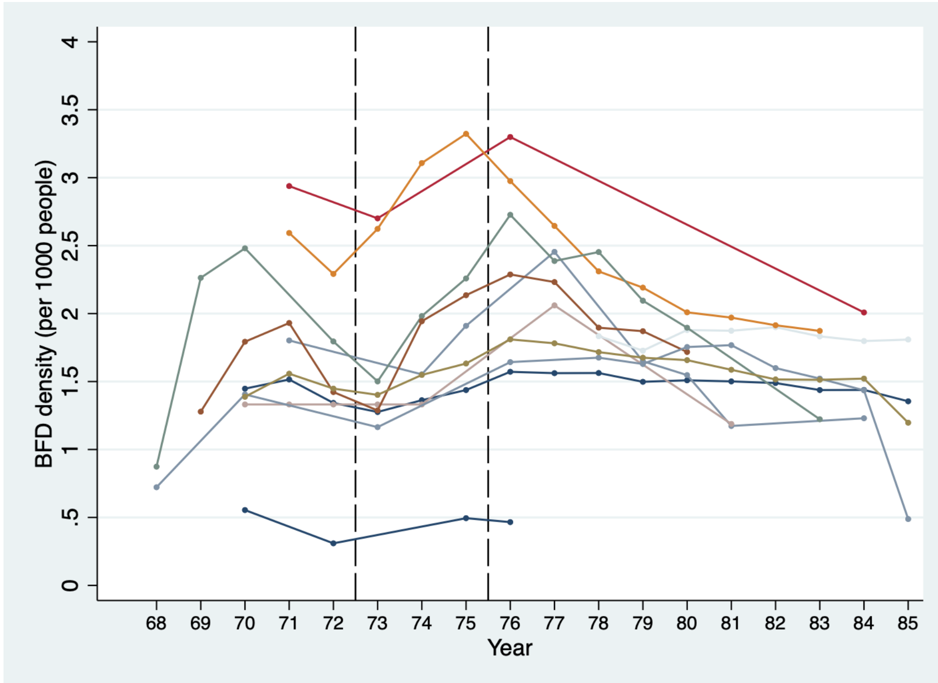
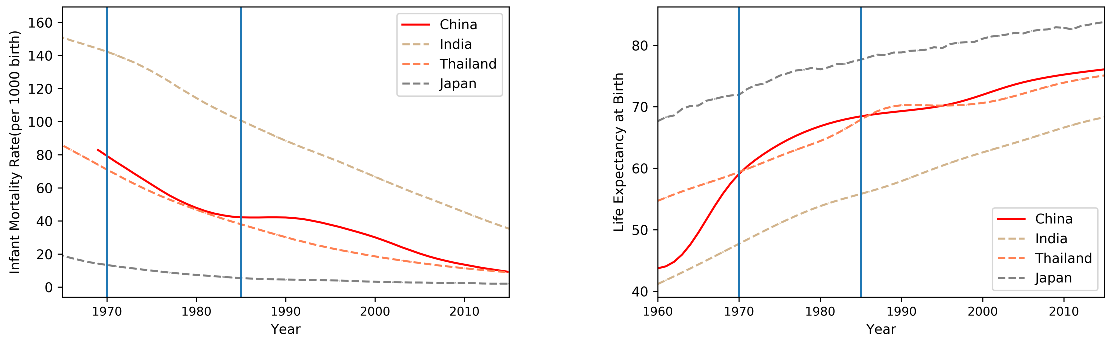



Figure 1: BFD density overtime[9]

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, with villagers demanding treatments and drugs when unnecessary. 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. Because of the reason above, between 1972 to 1973, the BFD program was suspended in most counties.

After 1974, local governments started to change the way cooperative medical systems and BFDs worked. Many changed from free medical care to a co-pay or fully charged system and BFDs were compensated better. The BFD programs 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.

Still, most of the BFDs did not make more than an average farmer's income. Under the cooperative system, the BFDs were not allowed to opt out. This is not to say that BFDs were forced into their positions. In fact, many saw their job as a political honor. In short, the BFD program functioned without many economic incentives. 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




(a) Infant Mortality Rate(per 1000 birth)

(b) Life Expectancy at Birth

Figure 2: Health Conditions Overtime(1960-2017)[10]

BFDs migrated to cities or went back to farming. In 1982, the government required all existing BFDs to take qualification exams for further medical practice. In 1985, the cooperative medical system was abandoned. Fiscal support was no longer provided from the government. Those with certificates were then referred to as *Village Doctors*, and they could set up private clinics and charge market price. Therefore, although parts of BFDs still exist today as *Village Doctors*, they are no longer CHWs. The ending of BFD programs and cooperative medical systems "are the unintended consequences of economic reform" (Hsiao, 1984).

**Short-run Health Improvements** There are some suggestive evidence for health improvements during the time of BFD programs. From 1969 to 1978, China's children mortality rate declined from 117.2 to 69.6 (per 1000 live birth); 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. According to Piazza (2019), average calorie intake for the rural population dropped below adequacy standards shortly after 1958 and remained there for two decades. Figure 2(a) and Figure 2(b) illustrate the comparative magnitude of this improvement in health indices. Those two figures also show that after the end of BFD program in 1985, the curves become flatter for several years. Note that these figures only suggest a possible correlation here, and no study has successfully unidentified causality.

[10]Data Source: World Bank. https://data.worldbank.org/

### 2.2 The Jobs of Barefoot Doctors

The jobs of BFDs were both preventive and curative. Specifically, the services most related to infants and children under age of five included birth delivery, home visits, vaccine injection, epidemic disease ( tuberculosis and malaria) diagnosis, health education, prescriptions, and referrals. They also cooperated with the local government on campaigns for sanitation, such as teaching villagers to use clean water and collect garbage (Young, 1989).

The method adopted by BFDs is a mixture of traditional and western medicine. On the one hand, they were encouraged to use traditional medicines like herbs to reduce the cost. Although some of them worked well, many of the traditional medicines have been proven useless. On the other hand, for many villages, the BFDs introduced western medicines for the first time in history, and significantly changed the attitudes toward western medicine (Fang, 2012).

During the time of the BFD project, there were other campaigns that aimed at improving health. For example, the "new deliver method" introduced modern delivery methods to replace the home-delivery. "Patriot hygiene movement" encourages people to use clean water, build toilets, killing pests and take care of personal hygiene. "Health for children" is to give out vaccines for children and free health examination. All those movements may have benefited the health of children other than barefoot doctor project itself. However, according to gazetteer records, all the movements above were conducted with the help of barefoot doctors: they were the ones who gave out vaccine shots and physical examinations. Without barefoot doctors, those health campaigns may not have been effective. Therefore, the estimates of BFDs' effects contain their interactions with those health campaigns.

## 3 Data and Variables

### 3.1 Gazetteer Data

For this paper, I construct a unique county level dataset on the details of barefoot doctor program. It is manually collected from China's historical gazetteers. As far as I know, this is the first data set that contains quantitative information on China's barefoot doctor program. The variables included are:

**Timeline of BFD program**. That includes begin year, end year, and interrupted years. Begin year for almost all counties is 1969. Many counties don't have clear records on when the BFD program ends. More often, the gazetteer would record:

"*...After the economic reform in 1978, the Cooperative Medical System as well as barefoot doctors began to collapse. 50% of them are still operating in 1980, and in 1982 there were only 20% left....And in 1985 the all barefoot doctors were transformed to village doctors.*"

From those records, it seems that in many counties the program gradually phased out from 1978 to 1985. Therefore, to be more conservative, in the main regressions I will use year 1978 as the ending year for all counties. In addition, many records tell that the BFD program was interrupted during 1972-1973, for the reasons explained in section 2.1. Therefore after main regressions I will excludes cohorts born during this period for robustness checks.

**Number of BFDs**. Density of BFDs is calculated using number of BFDs of county c divided by its population size in 1981, times one thousand. [11] [12] The density of BFDs is used to represent program's intensity in each county.

There are two reasons to use density of doctors to measure intensity. First, according to Das et al. (2016)'s study in India, for health workers who work in disadvantaged environments, their time spent with each patient is an important predictor of results than quality of diagnoses. Such relationship may also applies to BFDs. Since I can't observe their time allocation, one thing that could be correlated is the density of BFDs. Second, for each village, it usually requires at least two BFD: one male and one female, to cover the population. This is because many females would only to go female doctors, especially for birth delivering. Therefore, density of BFD is important for the program to be effective. Descriptive statistics of the distribution of BFDs are shown in 1.

Some concerns may be raised regarding the authenticity of gazetteers. Indeed, there are incentives for the local leader to over-report number of BFDs. If every county had the same concern, then by comparing relative density of BFD, it won't cause estimation bias. However, if counties that

[11] I use 1981 population because the third national census in 1982 is the nearest data on county-level population size that's available.

[12] The average population size of a village is one thousand during that time, so that I time 1000 to approximate number of BFDs in each village. Usually, BFDs are attached their village, or production team, and would only cover patients in their own village.

Table 1: Distribution of BFD density

| | Num of Counties | Mean | SD | P25 | P50 | P75 |
|---|---|---|---|---|---|---|
| BFD density (max) | 130 | 1.845 | 0.887 | 1.300 | 1.710 | 2.266 |
| BFD density (mean) | 130 | 1.642 | 0.805 | 1.181 | 1.502 | 2.017 |

Data Source: China's historical gazetteers and the population census in 1982.

perform worse in BFD program tend to over-report more than the ones that perform better, then the coefficients would be under-estimated. Another concern is missing data. For many counties data is missing on number of BFDs in some years. The reason for missing data is not known. The potential problem behind is that counties might intentionally hide data in those years when they perform badly. Again, in this case, counties with lower BFD density are now considered with higher density, and the positive results shown in my regression are actually under-estimated. Therefore, the estimates could be biased by the selection of data, but won't harm the main conclusions.

### 3.2 Census Data

I use Census of Population Survey (CPS) in 1982 and 1990 for county-level aggregate statistics. Variables including infant mortality rate (1982 census only), mortality rate of children under age of five (1990 census only), agriculture employment rate, age structure and education level of households. I use population data from census 1982 for calculating the density of BFDs.

### 3.3 Survey Data

Individual outcomes data is from two sources: *Chinese General Social Survey*(CGSS)[13] in 2010, 2011, 2012, 2013 and 2015. CGSS is a repeated cross-section data, covering 100 random-selected counties in all 23 provinces, among which 58 of them are rural counties[14] The sample size after compiling five years' data is 51,574. I drop counties in minority ethnic group provinces (Xinjiang, Tibet and Inner Mongolia) because policies are usually very different in those regions. On average, each province has 3 counties in the sample.

[13]For a long time, the CGSS is the only large-scale survey projects that open data in China mainland. The CGSS data has become the most important data resource for China study. By December 2014, the registered CGSS data users have more than 2,0000. More than 700 journal papers published based on CGSS data.

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

Among them, I only use non-migrated sub-sample for analysis. The reason for using non-migrated sample is mainly technical: since this survey only covers 100 counties across the nation, if the individual moves out to a county that is not surveyed, I cannot observe his or her outcomes. This is also a reasonable choice since for those who migrant, their health outcomes may be largely determined by medical conditions in their destinations (usually cities). One major concern using this sub-sample is migration selection: individuals who are healthier are more likely to migrant out, which means the left pool can be on average less healthy. To address this concern, I argue that: First, the main empirical strategy I use is DID, which compares pre/post treatment groups in areas with different intensity. With the assumption that the relative health condition of those who migrant are consistent across counties over time, this migrant selection won't change my estimations. Second, in 2, I compare the descriptive statistics between full sample and no-migrant sample. It shows that for all variables of interest there is no significant difference between full sample and non-migrant sample.

The outcome variables of interest are health, educational attainment and labor market outcomes. Health outcomes including height, body weight, self-reported health, self-reported depression, and self-reported serious health issues, chronic disease, healthy behaviors such as smoking and drinking. I use height and body weight to construct BMI [15], underweight ($BMI \leq 18.5$), overweight ($24 \leq BMI \leq 28$ ), obese ($BMI \geq 28$). [16] and height stunting [17]; Education outcomes including highest education level, years of schooling and probability of middle school completion. Labor market outcomes including employment status and working hours. I also use demographic features, including gender, ethnic, year of birth, and family background as controls. I construct indices for health, nutrition and education outcomes using unweighted z-scores by subtracting the control group mean and dividing by control group standard deviation.

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

[16] Standards for underweight, overweight and obesity come from General Administration of Sport of China. http://www.sport.gov.cn/n16/n1077/n1422/7331093.html

[17] 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

Table 2: Descriptive Statistics

| VARIABLES | Full Sample | | | Non-Migrant Sample | | |
|---|---|---|---|---|---|---|
| | N | Mean | SD | N | Mean | SD |
| **Health Outcomes** | | | | | | |
| Good Health = 1 | 8,772 | 0.614 | 0.487 | 7,298 | 0.618 | 0.486 |
| No Depression = 1 | 7,773 | 0.670 | 0.470 | 6,505 | 0.676 | 0.468 |
| Problem Free= 1 | 7,762 | 0.728 | 0.445 | 6,495 | 0.731 | 0.443 |
| Height(cm) | 8,752 | 163.8 | 7.374 | 7,283 | 164.1 | 7.368 |
| Stunt = 1 | 8,777 | 0.0883 | 0.284 | 7,303 | 0.0909 | 0.288 |
| BMI | 8,748 | 22.86 | 3.291 | 7,281 | 22.86 | 3.286 |
| Overweight = 1 | 8,748 | 0.233 | 0.423 | 7,281 | 0.233 | 0.423 |
| Underweight = 1 | 8,748 | 0.0604 | 0.238 | 7,281 | 0.0588 | 0.235 |
| Obese = 1 | 8,748 | 0.0636 | 0.244 | 7,281 | 0.0619 | 0.241 |
| No Smoking = 1 | 1,936 | 0.808 | 0.394 | 1,603 | 0.800 | 0.400 |
| Over-drinking = 1 | 979 | 0.0889 | 0.285 | 779 | 0.100 | 0.300 |
| **Educational Outcomes** | | | | | | |
| Educ Level | 8,776 | 4.169 | 2.175 | 7,303 | 4.132 | 2.103 |
| Years of Schooling | 7,997 | 8.893 | 2.707 | 6,690 | 8.833 | 2.647 |
| Mid School Completion = 1 | 8,777 | 0.598 | 0.490 | 7,303 | 0.600 | 0.490 |
| **Labor Market Outcomes** | | | | | | |
| Labor Force Participation =1 | 8,777 | 0.897 | 0.304 | 7,303 | 0.899 | 0.301 |
| Unemployment = 1 | 7,871 | 0.0808 | 0.273 | 6,569 | 0.0760 | 0.265 |
| Work hour per week | 8,258 | 35.33 | 30.25 | 6,908 | 35.69 | 30.19 |
| **Demographics** | | | | | | |
| Gender(male=1) | 8,777 | 0.468 | 0.499 | 7,303 | 0.499 | 0.500 |
| Age | 8,777 | 43.45 | 4.768 | 7,303 | 43.52 | 4.751 |
| Married = 1 | 8,777 | 0.984 | 0.124 | 7,303 | 0.983 | 0.130 |
| **Family Background** | | | | | | |
| Illiterate Father = 1 | 8,777 | 0.371 | 0.483 | 7,303 | 0.375 | 0.484 |
| *Hokou* (Agriculture=1) | 8,777 | 0.730 | 0.444 | 7,303 | 0.752 | 0.432 |

# 4 Empirical Design

## 4.1 Difference in Difference

A difference-in-difference model is employed for the main regressions. In this model, I compare outcomes for individuals born before/after the introduction of BFD program as the *Treated* and *Control* accordingly; and then compare the differences between counties with high/low intensity of BFD program. The model estimated is shown below:

$$(1)Y_{i,c,t,s} = \beta_1 Expo_t \times BFD\ Density_c + X_i + \delta_c + \sigma_t$$
$$+ \gamma_s + \theta_p \times t + \alpha + \epsilon_{i,c,t,s}$$

where i represents individual, c for county of birth, t for year of birth, s for round of survey, and p for province of birth. $Y_{i,c,t,s}$ represents outcomes of individual i born in county c, year b, including health outcomes: self-reported health, height, overweight, underweight, self-reported mental health, self-reported health issues; educational outcomes: level of education and years of schooling; labor market outcomes: unemployment and work hours per week.

$BFD\ Density_c$ represents density of BFDs, calculated by number of BFDs in a county divided by total population in 1982 census, times one thousand. $Expo_t$ is a dummy variable that equals to 1 if individual i is born between 1970 to 1978. I use 1978 as the ending threshold because it is when the most county ends the program.

I use 1970 as the start of exposure for two reasons. First, the majority of counties start the BFD program in 1969. There are a few counties started one or two years earlier or later for some unobserved reasons. Therefore I use 1969 as an instrument variable for starting year to avoid endogeneity. Second, I lag one year because previous studies have shown that *in utero* treatment has a large and lasting impact on health (Currie and Almond, 2011). And birth cohorts of 1970 are the first to be exposed to BFD program *in utero*. In addition, training of BFDs takes 3 months on average, and there may be some lags of program implement.

My control group here is cohorts born between 1963 to 1969. Therefore individuals in my sample ages from 33 to 53. I choose this window of control for two reasons. First, China's Great

Famine, which took place from 1959 to 1961, was proven to have a significant negative effect for health, education, labor market for the survivors in the long term (Meng and Qian, 2009; Chen and Zhou, 2007). Those studies show that even being exposed to Great Famine in uterus-which is the birth cohorts born between 1960 to 1962- would predict worse outcomes. Therefore, using cohorts born before 1962 might over-estimate the results. Second, as shown in the event study in table**??** later, birth cohorts before 1963 have unparalleled pre-trends with cohorts born between 1963 to 1968, which will contaminate my DID results. Therefore, I only use 1963-1968 cohorts as control.

$X_i$ is a vector of demographic controls, including gender, ethnic, age, and family backgrounds like whether the father is illiterate or not.

I include a bunch of fixed effects in this model: $\delta_c$ for county fixed effect, $\sigma_t$ for birth cohort fixed effect, $\gamma_s$ for survey round fixed effect, and $\theta_p \times t$ is province-specific time trends. Note that $BFD\ Density_c$ is captured by county fixed effect and $Expo_t$ is captured by province linear time trends. All regressions are clustered at the county level, and use the given weight by CGSS.

### 4.2 Testing Parallel Trends Assumption

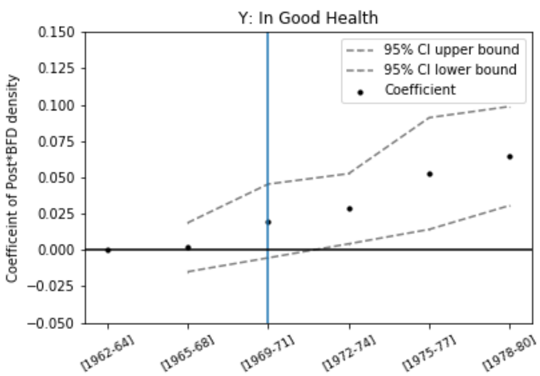
The key assumption for DID model is parallel pre-trends. In equation (1), I have ruled out heterogeneous trends at province level by controlling for province specific linear trends. However, there might still exist nonparallel county-specific trends, for example the endogenous allocation of BFDs in each county. Therefore, I employ an event study to test parallel trend assumption using equation (1), except not variable $Expo_b$ is split into a series of dummies, each represents a birth cohort at three years' interval. I omitted cohorts born in 1962-1964 to avoid multicollinearity. Everything else is same to equation (1). Max of BFD number is used here for calculating BFD density. Coefficient of interest is $\beta_b$ for different birth cohorts, especially for cohorts born before the program: if they are significantly different from zero, then there might be some county-specific trends that are different for counties with different BFD density, which would fail the assumption of DID regression.

I plot the birth-cohort specific coefficients in Figures before each table of regression accordingly. The Y axis stands for the magnitude of coefficients of the interaction terms, and X axis is birth cohort. The dots represent each coefficient and the dashed lines for 95% confidence interval upper/lower bounds. For all figures, it shows that before the introduction of BFD program, coefficients are scattering around zero. There's no evidence showing pre-trends for all outcomes.

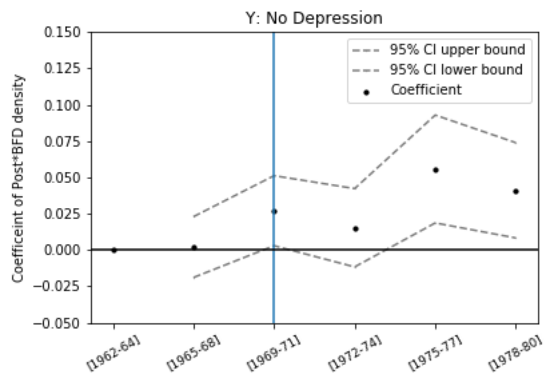
# 5 Results

## 5.1 Main Results of DID

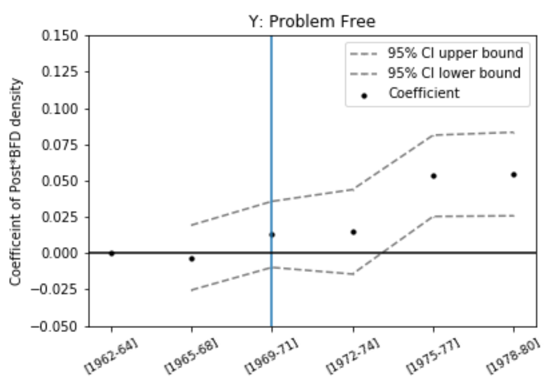
Table 3 contains DID results using reported general health outcomes as dependent variables. The coefficient of interest is $\beta_1$ in equation (1), which is shown in the first rows of each panel. Panel A uses the maximum number of BFDs during 1969 to 1985 to calculate BFD density, and Panel B uses the mean number of BFDs during 1969 to 1985. Comparing coefficients in two panels shows that there are not much difference in scale and significance level, which means the results are robust to different methods of BFD density calculation. Therefore, I will only use the $BFD\_density\_max$ for density in later regressions to include more observations. Comparing results with and without demographic and family background controls also shows that these controls would only reduce the coefficients by a very small scale and won't affect the significance of coefficients.


Y: In Good Health
95% CI upper bound
95% CI lower bound
Coefficient
Coefficeint of Post*BFD density
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
[1962-64]
[1965-68]
[1969-71]
[1972-74]
[1975-77]
[1978-80]

(a)


Y: No Depression
95% CI upper bound
95% CI lower bound
Coefficient
Coefficeint of Post*BFD density
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
[1962-64]
[1965-68]
[1969-71]
[1972-74]
[1975-77]
[1978-80]

(b)


Y: Problem Free
95% CI upper bound
95% CI lower bound
Coefficient
Coefficeint of Post*BFD density
0.150
0.125
0.100
0.075
0.050
0.025
0.000
−0.025
−0.050
[1962-64]
[1965-68]
[1969-71]
[1972-74]
[1975-77]
[1978-80]

(c)

Table 3: Introduction of BFD on Health - General

| | (1) Health Index | (2) Good Health | (3) Good Health | (4) No Depression | (5) No Depression | (6) Problem Free | (7) Problem Free |
|---|---|---|---|---|---|---|---|
| | Panel A: Using Max(Number of BFD) to Calculate Density | | | | | | |
| $BFD\ Density_c \times \text{Expo}_t$ | 0.082** | 0.045** | 0.045** | 0.043** | 0.041** | 0.029 | 0.028 |
| | (0.036) | (0.018) | (0.017) | (0.019) | (0.018) | (0.020) | (0.019) |
| Dep. Mean | 0.132 | 0.630 | 0.630 | 0.676 | 0.676 | 0.737 | 0.737 |
| Effect Size (Absolute) | | 0.08 | 0.08 | 0.08 | 0.08 | 0.074 | 0.072 |
| Effect Size (%) | | 12.8% | 12.8% | 12.2% | 11.9% | 10.2% | 10.0% |
| | Panel B: Using Mean(Number of BFD) to Calculate Density | | | | | | |
| $BFD\ Density_c \times \text{Expo}_t$ | 0.082** | 0.048*** | 0.048*** | 0.037* | 0.036* | 0.028 | 0.027 |
| | (0.039) | (0.018) | (0.017) | (0.020) | (0.019) | (0.022) | (0.021) |
| Dep. Mean | 0.132 | 0.630 | 0.630 | 0.676 | 0.676 | 0.737 | 0.737 |
| Effect Size(Absolute) | | 0.087 | 0.085 | 0.085 | 0.084 | 0.083 | 0.083 |
| Effect Size(%) | | 13.8% | 13.4% | 12.6% | 12.4% | 11.2% | 11.2% |
| Demographic Controls | | | X | | X | | X |
| Year of Birth FE | X | X | X | X | X | X | X |
| Survey Round FE | X | X | X | X | X | X | X |
| County FE | X | X | X | X | X | X | X |
| t*Province of Birth FE | X | X | X | X | X | X | X |
| Obs. Num | 5202 | 5824 | 5819 | 5217 | 5213 | 5209 | 5205 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

NOTE: For the first 3 columns, the higher number means better health. *In good Health* is a dummy that equals to 1 if in the question ”In general, how’s your health condition?” the answer is ”very good” and ”good”. It equals to 0 if the answer is ”fair”, ”not good” and ”very bad”. *No Depression* is a dummy that equals to 1 if in the question ”In the past four weeks, how often do you feel depressed or upset?” the answer is ”never” and ”very few”: it equals to 0 if the answer is ”sometimes”, ”often” and ”always”. *Problem Free* is a dummy that equals to 1 if in the question ”In the past four weeks, how often did your health issue affect your working and other daily activities?” the answer is ”never” and ”very few”: it equals to 0 if the answer is ”sometimes”, ”often” and ”always”.

Results in Table 3 show that the interaction of *Expo* and *BFD Density* has a positive and significant impact on long-term reported health. In column 1 and 2 of panel A, the coefficients are both 0.033 and significant at 5% level. The mean of BFD Density is 1.8 (per 1000 people), which means on average, being exposed to BFD program raise probability of reported good health by 0.06. Given that the mean of reported good health is 0.631, this means a 9.4% increase in probability of being in good health. Exposure to BFDs will also decrease the probability of having a serious health problem in the past four weeks by 7.8% to 9.2%.

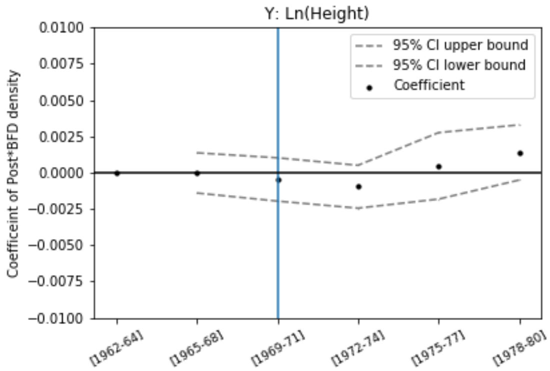
Although BFDs were not providing any mental health treatment directly to the children, the mechanisms behind early childhood environment and depression are mainly through the physical and mental health status of the mother. Zuckerman et al. (1989) show that depressive symptoms among low income pregnant women and mothers have been associated with low nutrition intake, unintended pregnancy, use of tobacco, alcohol, and illicit drugs. And depressive symptoms can be transmitted to their children (Kahn et al., 2002). Results in column 3 and 4 in Table 3 show that exposure to BFD program will increase the probability of not being in depression by 8.3% to 8.9%.

Table 4 uses health outcomes related to nutrition intake, including height, stunting, body weight, BMI, probability of underweight and probability of overweight. Results in Table 4 shows that none of the coefficients are significant. Some of the coefficients are even negative. This result is not unexpected. From the previous discussion in the background section, nutrition intake improvement is not amongest the jobs of BFDs.

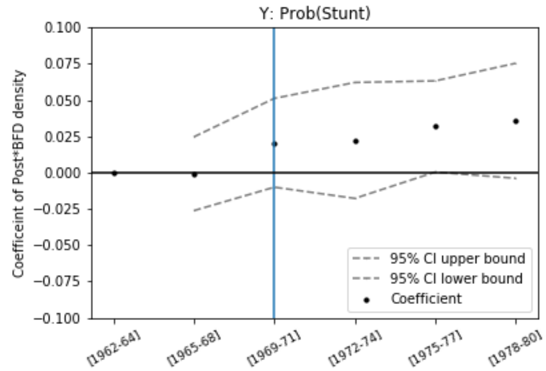
In addition, these null results could help rule out other possible channels in explaining health improvement during this historical time period. One concern on the regression is that many other events were happening at the same time with BFD program, for example cooperative agriculture movement [18]. If the local governments that are better at promoting BFD program are also better at implementing other programs, and these other programs result in higher agricultural productivity, then the health improvements observed can be mainly caused by nutrition increase instead of BFDs. Results in Table 4 show that this is not the case.

Table 5 presents the exposure of BFD program on health behaviors. The event study results are shown in figure 1 in appendix. Notice that the number of observations decrease by a large amount

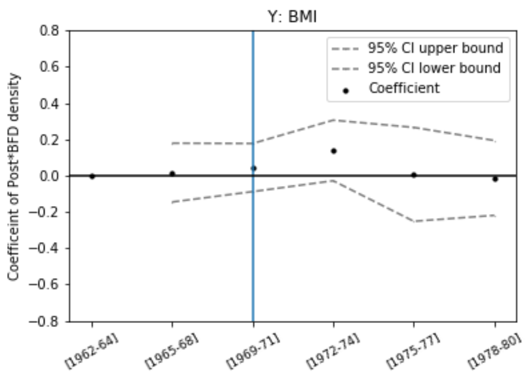
[18] The movement to unionize farmers in the same village, sharing capitals, working as production teams and share the outcomes.



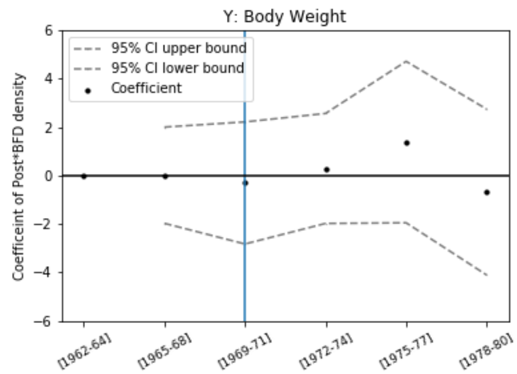

(d)




(e)




(f)




(g)

Table 4: Introduction on BFD on Health - Nutrition

| | (1)<br>Nutrition<br>Index | (2)<br>ln(height) | (3)<br>Stunt | (4)<br>Body<br>Weight | (5)<br>BMI | (6)<br>Under-<br>weight | (7)<br>Over-<br>weight |
|---|---|---|---|---|---|---|---|
| $BFD\ Density_c$ $\times$Expo$_t$ | -0.008 | -0.001 | 0.017 | -0.332 | -0.018 | -0.009 | -0.019 |
| | (0.025) | (0.001) | (0.014) | (1.028) | (0.149) | (0.009) | (0.025) |
| Dep. Mean | 0.042 | 5.101 | 0.088 | 123.995 | 22.884 | 0.059 | 0.236 |
| Obs. Num | 5812 | 5813 | 5822 | 5818 | 5812 | 5812 | 5812 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

NOTE: BMI is short for ”Body Mass Index”, which equals to weight (unit:kg) over square of height (unit: cm). Underweight is a dummy variable equals to 1 if $BMI \leq 18.5$. Overweight =1 if $24 \leq BMI \leq 28$. These standards above come from General Administration of Sport of China (http://www.sport.gov.cn/n16/n1077/n1422/7331093.html). ”Stunting” is defined as shorter than two standard deviations of the average height of each gender. Height distribution data comes from General Administration of Sport of China.

Table 5: Exposure to BFD on Health - Health Behaviors

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | No Smoking | | Over-drinking | | Exercise Often | | Physical Exam | |
| | OLS | Logit | OLS | Logit | OLS | Logit | OLS | Logit |
| $BFD\ Density_c$ | 0.086** | 0.474** | -0.020 | -1.293** | -0.011 | -0.249 | -0.027 | -0.202 |
| $\times Expo_t$ | (0.032) | (0.240) | (0.025) | (0.620) | (0.048) | (0.266) | (0.074) | (0.408) |
| Dep. Mean | 0.720 | 0.720 | 0.086 | 0.110 | 0.211 | 0.223 | 0.455 | 0.454 |
| Obs. Num | 597 | 597 | 595 | 462 | 592 | 561 | 517 | 515 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

NOTE: All four dependent variables are dummy variables. *No Smoking* equals to 1 if in the question "How often do you smoke?" the answer is "Never" or "Used to but not anymore". *Over-drinking* equals to 1 if in the question "How often do you drink more than 4 units of alcohol per day?" the answer is "Every day" and "more than 3 days a week". *Exercises Often* equals to 1 if in the question "How often do you work out more than 20 minutes a day?" the answer is "Very often" and "Often". *Physical Exam* equals to 1 if in question "How often do you do physical examinations" the answer is "More than once in a year" or "Once in a year".

because these questions are only asked in one survey round of CGSS (2011). I use OLS as well as logit regression since the outcome variables are all binary dummies. Results in Table 5 show that exposure to BFD program significantly decreases the probability of smoking and heavy drinking behaviors. It has a negative but not significant impact on exercise, which is not surprising since daily work-outs seem unnecessary for people who work in agriculture. It is also not unexpected to see no significant effects on physical exam behavior, since annual physical examination is not common even in urban areas in China, let alone rural population. The mechanism behind this relationship could come from higher education and more years of schooling, but it is hard to imagine a direct impact of childhood exposure of BFD program on smoking and drinking.

Next, I want to see if the early childhood exposure to BFD program can affect educational and labor market outcomes. Previous literature show that the relationship can be strong and economically significant (Hoynes et al., 2016; Meng and Gregory, 2002). As for the mechanism behind this relationship, Currie et al. (2010) show that early childhood health conditions are predictive of future outcomes, but after controlling for adult physical health conditions, the effect largely goes away. This suggests that physical health in early childhood would affect adult outcomes through affecting adult physical health, rather than through cognitive ability. However, they find that the

Table 6: Introduction of BFD on Educational and Labor Market Outcomes

| | (1) Level of Educ | (2) Years of Schooling | (3) Mid-School Complete | (4) Unemploy-ment | (5) Work Hour Condition on Employed | (6) Work Hour Per Week |
|---|---|---|---|---|---|---|
| $BFD\ Density_c \times \text{Expo}_t$ | 0.145* | 0.217* | 0.045* | -0.009 | 1.401* | 1.497* |
| | (0.085) | (0.128) | (0.024) | (0.021) | (0.804) | (0.881) |
| Dep. Mean | 4.178 | 8.919 | 0.612 | 0.069 | 40.643 | 36.337 |
| Obs. Num | 5331 | 5822 | 5822 | 5291 | 4818 | 5522 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Note: *Level of Education* is a discrete variable, with each number represents a different degree level. The higher the number, the more valuable the degree. *Years of Schooling* is a continuous variable that calculated from level of education.

mental health condition has a direct impact on economic outcomes. Table 6 presents the effect of BFD program on long-term educational attainment and labor market outcomes. In the first three columns, coefficients of the interaction term are all positive. It shows that being treated in high intensity of the program increases level of education, years of schooling and middle school completion. The last three columns presents results on labor market. I find negative but insignificant effects on employment. This is reasonable since most of them work in their own farm land. In column 4 and 5, I find positive effect on labor supply: exposure would on average increase work hour per week by 1.5 hours in general, and 1.4 hours conditioning on being employed.

## 5.2 Heterogeneity of Impact

In the previous studies on early childhood intervention, the influences are usually heterogeneous among different population. The first heterogeneity comes from gender. Many studies find that the effect size of intervention is larger on girls (Orr et al., 2003; Anderson, 2008). Milligan and Stabile (2009) studied the influence of Canadian child benefits and find that the pattern of influence is different between genders, with boys show greater increase in test scores and girls on mental health and behavioral scores. In order to test that, I split sample by gender and run regression of equation (1) on sub-samples. Results are shown in Table 7 and Table 8. The pattern presented is that exposure to BFD program has a larger effect for female on health, and has a larger effect for male on education. This pattern is in accordance with the situation in rural China. Girls are less likely

Table 7: Heterogeneity of Effects Between Genders

| | (1) Health Index | (2) In Good Health | (3) No Depression | (4) Problem Free |
|---|---|---|---|---|
| | **Panel A: Male Sub-sample** | | | |
| BFD Density Expo | 0.011 | 0.003 | 0.032 | -0.008 |
| | (0.039) | (0.024) | (0.021) | (0.021) |
| Dep. Mean | 0.194 | 0.670 | 0.712 | 0.775 |
| R-Squared | 0.099 | 0.096 | 0.090 | 0.076 |
| Obs. Num | 2660 | 2933 | 2670 | 2663 |
| | **Panel A: Female Sub-sample** | | | |
| BFD Density Expo | 0.152*** | 0.084** | 0.052* | 0.071** |
| | (0.052) | (0.032) | (0.028) | (0.029) |
| Dep. Mean | 0.038 | 0.590 | 0.639 | 0.697 |
| R-Squared | 0.151 | 0.149 | 0.100 | 0.130 |
| Obs. Num | 2538 | 2886 | 2543 | 2542 |

Standard errors in parentheses

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

to be sent to schools for education than boys. Therefore, when the health status of a boy improves, he might be able to stay in school longer for better education, while girls are stopped by their family.

### 5.3 Comparing the Effect Size

Comparing the scale of effect to other program or natural experiment on health will give us a better understanding of the results. In the U.S, Hoynes et al. (2016) find that exposure to *Food Stamp* through age of five would increase reported good health by 30 percentage point, with a mean of 0.679, which means a 44% increase. A program in Guatemala in the 1970s where children are given access to a protein drink increased educational attainment of women by more than one grade, and increased wage of male by 46% (Hoddinott et al., 2008; Maluccio et al., 2009). In China, Meng and Qian (2009) estimate that exposure to the Great Famine (1959 - 1961) at one-year's old reduced height by 2.08%(3.34 cm), weight by 6.03% (3.38 kg), BMI by 4%, work hour each week by 6.93% (3.28 hrs), and years of schooling by 3% (0.19 years) on average.

Table 8: Heterogeneity of Effects between Genders

| | (1) Level of Educ | (2) Schooling | (3) MidSchool Complete | (4) Unemployment | (5) Work Hour Per Week if Employed | (6) Work Hour Per Week |
|---|---|---|---|---|---|---|
| | **Panel A: Male Sub-sample** | | | | | |
| *BFD Density* ×Expo | 0.181** | 0.221* | 0.060* | 0.002 | 1.444 | 1.882 |
| | (0.084) | (0.127) | (0.030) | (0.027) | (1.350) | (1.536) |
| Dep. Mean | 4.513 | 9.244 | 0.700 | 0.051 | 42.297 | 40.240 |
| Obs. Num | 2935 | 2821 | 2935 | 2829 | 2637 | 2833 |
| | **Panel B: Female Sub-sample** | | | | | |
| *BFD Density* ×Expo | 0.066 | 0.158 | 0.022 | -0.019 | 1.326 | 0.930 |
| | (0.106) | (0.172) | (0.025) | (0.018) | (1.570) | (1.230) |
| Dep. Mean | 3.838 | 8.553 | 0.523 | 0.089 | 38.644 | 32.225 |
| Obs. Num | 2887 | 2510 | 2887 | 2462 | 2181 | 2689 |

Standard errors in parentheses

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

In conclusion, early childhood environment intervention can generate large scale long-term effects, and the magnitude of effects found in this paper are either similar or smaller than previous literature.

# 6 Possible Threats to Internal Validity

## 6.1 Mortality Selection from the Great Famine (1959-1961)

China's Great Famine from 1959 to 1961 may confound the results above by mortality selection: the healthier people are more likely to survive disadvantaged environments, which means their children might be healthier than other cohorts. Meng and Qian (2009) found that taller individuals were more likely to survive the famine. Although I have deleted birth cohorts before 1962 to avoid influence from that, the *Unexposed* group may still be getting effects from the famine. However, even if mortality selection exists in my study time period, it would only make my result an underestimation of BFD's impact, since the health outcomes of *Unexposed* group is higher than

Table 9: Excluding 1973-1974 Cohort Sample

| | (1) In Good Health | (2) No Depression | (3) Problem Free | (4) Educ Level | (5) Years of Schooling | (6) Work hour If Employed | (7) Work Hour |
|---|---|---|---|---|---|---|---|
| BFD Density $\times Expo$ | 0.047** | 0.044** | 0.033* | 0.147* | 0.214 | 1.514 | 1.978* |
| | (0.019) | (0.021) | (0.019) | (0.084) | (0.128) | (1.018) | (0.990) |
| Dep. Mean | 0.627 | 0.676 | 0.735 | 4.151 | 8.896 | 40.299 | 36.079 |
| Obs. Num | 5136 | 4608 | 4601 | 5138 | 4691 | 4242 | 4865 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

Note: All regressions include year of birth fixed effect, county fixed effect, province-specific time trends, survey round fixed effect and demographic and family background controls.

normal.

### 6.2 Interruption of BFD Program

The BFD was interrupted by the cultural revolution in 1972-1973. According to records in gazetteers, many counties suspended their BFD programs in these two years, by not providing medical training for the doctors and cut fiscal support for the doctors' clinics. Although it is likely that barefoot doctors who finished their training before 1972 were still willing to provide some medical help, the treatment effect are most likely to drop. To take into account this interruption, I dropped cohorts born between 1973 to 1974 and rerun the main regressions. I would expect the treatment effects more significant. The results are shown in table 9. Results show that after dropping samples from those years, the magnitude of coefficients increases, as well as significant level, which is in accordance with the prediction.

### 6.3 Impact of Cultural Revolution

There were many national movements going on at that time period, and the estimation can be coming from them rather than BFD program. Among them, the most salient one is the *Cultural Revolution* from 1966 to 1976. Existing literature shows that exposure to *Cultural Revolution* has negative effects on health, education and labor market outcomes. Since *Cultural Revolution* is overlapping the time period of this study, my estimates can be catching effects from it.

However, this is unlikely the case. First, my study only focus on rural population, while *Cultural*

Table 10: Placebo Test: Using Ending of Cultural Revolution as Treatment

| | (1) In Good Health | (2) No Depression | (3) Problem Free | (4) Educ Level | (5) Years of Schooling | (6) Work Hour if Employed | (7) Work Hour |
|---|---|---|---|---|---|---|---|
| BFD_density $\times CR$ | -0.006 | 0.002 | -0.018 | 0.034 | 0.093 | -1.817$^{***}$ | -1.787$^{**}$ |
| | (0.015) | (0.016) | (0.012) | (0.063) | (0.096) | (0.613) | (0.810) |
| Dep. Mean | 0.630 | 0.661 | 0.660 | 4.184 | 8.924 | 40.647 | 36.353 |
| Obs. Num | 5846 | 5234 | 5226 | 5849 | 5355 | 4842 | 5547 |

Standard errors in parentheses

$^{*}$ $p < 0.1$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$

*Revolution* is mainly an urban disturbance (Giles et al., 2008; Meng and Gregory, 2002). Second, I employ a placebo test in the following Table 10. I use cohorts born between 1966 to 1976 as my "unexposed" group (since being born during cultural revolution is negative for health and economic outcomes), and cohorts born between 1963-1965 and 1977-1978 as my "exposed" group. I run the same regression as equation (1), and results are shown in Table 10. First five columns of Table 10 shows that the coefficients are not significant now. Column 6 and 7 are significantly negative now. This would not affect my positive estimations on conditional and unconditional working hours.

# 7 Conclusion

This paper evaluate the impact of childhood and *in utero* exposure to China's *Barefoot Doctor* program on individual's health, educational and labor market outcomes 20 to 40 years later. Using a unique dataset collected from China's local gazetteers, I use the geographic variation in the intensity of BFD program across counties and *Pre-* / *Post-* birth cohorts to construct a difference-in-difference model for identification. Results from DID show that exposure to BFD program *in utero* and early childhood has a significant and positive impact on health. On average, a 50% increase in the number of BFDs raises reported health by 9.4%, and reduces the probability of depression and serious health problems by 8.5% and 8% respectively. It also reduces unhealthy behaviors like smoking and heavy drinking, but I find no significant improvement on nutrition. As for economic outcomes, a 50% increase in number of BFDs would increase years of schooling by 2.4% and middle school completion rate by 7.4%, and increase labor supply by 4.1%. The size of effects are larger for males in educational attainment and labor supply, while larger for females in health. I ruled out other possible explanations for the observed effects, for example the influence of Cultural revolution (1966-1976), mortality selection from China's Great Famine (1959-1961) and the endogeneity of BFD intensity. Compared to the scale of effects found in other programs, the effect size of BFD program found in this paper is more conservative. This paper is the first piece of empirical evidence on the long-term benefits of CHWs projects.

## References


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.

Balcazar, H., Alvarado, M., Hollen, M. L., Gonzalez-Cruz, Y., Hughes, O., Vazquez, E., and Lykens, K. (2006). Salud para su corazon-nclr: a comprehensive promotora outreach program to promote heart-healthy behaviors among hispanics. *Health promotion practice*, 7(1):68–77.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. (2007). Remedying education: Evidence from two randomized experiments in india. *The Quarterly Journal of Economics*, 122(3):1235–1264.

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.

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.

Brownstein, J. N., Chowdhury, F. M., Norris, S. L., Horsley, T., Jack Jr, L., Zhang, X., and Satterfield, D. (2007). Effectiveness of community health workers in the care of people with hypertension. *American journal of preventive medicine*, 32(5):435–447.

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. (2018). Arrival of young talents: The send-down movement and rural education in china. *Available at SSRN 3102716*.

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

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.

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

Giles, J., Park, A., and Wang, M. (2008). *The great proletarian Cultural Revolution, disruptions to education, and returns to schooling in urban China*. The World Bank.

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.

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.

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.

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.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Lawn, J. E., Mwansa-Kambafwile, J., Horta, B. L., Barros, F. C., and Cousens, S. (2010). kangaroo mother careto prevent neonatal deaths due to preterm birth complications. *International journal of epidemiology*, 39(suppl_1):i144–i154.

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 Gregory, R. G. (2002). The impact of interrupted education on subsequent educational attainment: A cost of the chinese cultural revolution. *Economic Development and Cultural Change*, 50(4):935–959.

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.

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.

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

Postma, J., Karr, C., and Kieckhefer, G. (2009). Community health workers and environmental interventions for children with asthma: a systematic review. *Journal of Asthma*, 46(6):564–576.

Roman, L. A., Gardiner, J. C., Lindsay, J. K., Moore, J. S., Luo, Z., Baer, L. J., Goddeeris, J. H., Shoemaker, A. L., Barton, L. R., Fitzgerald, H. E., et al. (2009). Alleviating perinatal depressive symptoms and stress: a nurse-community health worker randomized trial. *Archives of women's mental health*, 12(6):379.

Seward, N., Osrin, D., Li, L., Costello, A., Pulkki-Brännström, A.-M., Houweling, T. A., Morrison, J., Nair, N., Tripathy, P., Azad, K., et al. (2012). Association between clean delivery kit use, clean delivery practices, and neonatal survival: pooled analysis of data from three sites in south asia. *PLoS medicine*, 9(2):e1001180.

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

Spencer, M. S., Rosland, A.-M., Kieffer, E. C., Sinco, B. R., Valerio, M., Palmisano, G., Anderson, M., Guzman, J. R., and Heisler, M. (2011). Effectiveness of a community health worker intervention among african american and latino adults with type 2 diabetes: a randomized controlled trial. *American journal of public health*, 101(12):2253–2260.

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.

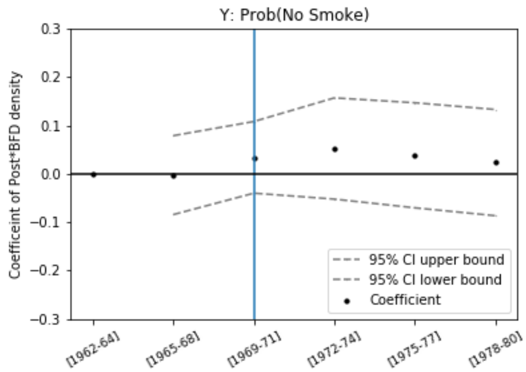
You, Y. (2019). Educated youth should go to the rural areas: A tale of education, employment and social values.

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

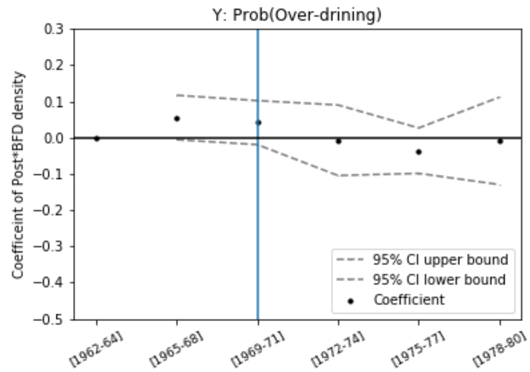
Zhang, D. and Unschuld, P. U. (2008). China's barefoot doctor: past, present, and future. *The Lancet*, 372(9653):1865–1867.

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.

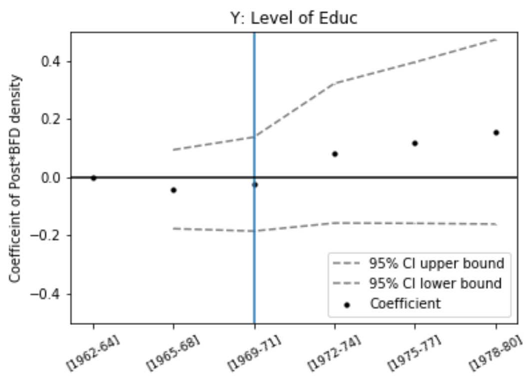
# Appendices



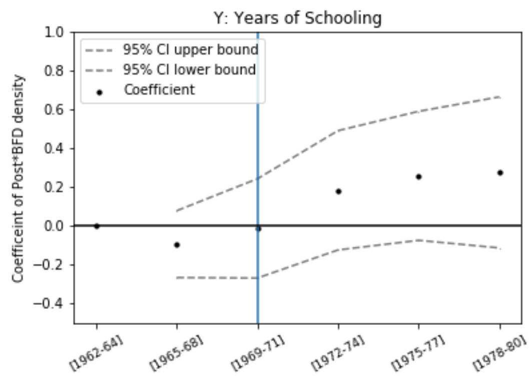

(h)



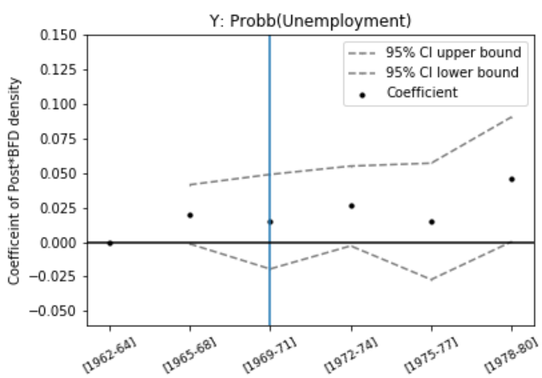

(i)



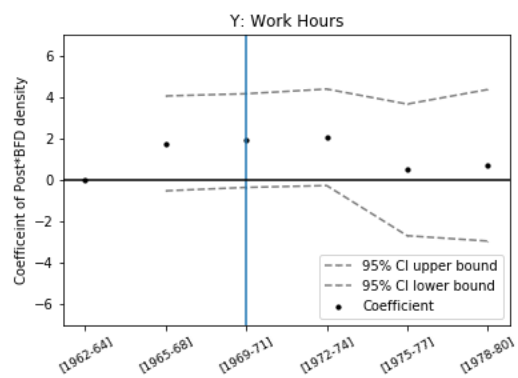

(j)




(k)




(l)




(m)

Figure 3: Event Study Results