ABSTRACT

Title of Document: THE IMPACT OF WATER QUALITY ON HEALTH IN RURAL CHINA

Jing Zhang,
Doctor of Philosophy, 2011

Directed By: Professor Judith Hellerstein
Department of Economics

Providing people with safe drinking water is one of the most important health-related infrastructure programs in the world. This dissertation investigates the effects of a major water improvement program in rural China on the health of adults and children. Using panel data covering about 4,500 households from 1989 to 2006, I estimate the impact of introducing village-level access to water from water plants on various measures of health. Ordinary least squares (OLS) estimation of the impact suggests a weak positive influence of the program on people's health status, but these results may be contaminated by endogenous timing and placement of the water quality interventions across China. To address potential endogeneity problems, I use topographic characteristics of communities as instruments for program placement, as these characteristics affect the costs of the construction of water plants and pipelines into villages. My instrumental variables (IV) results show that the introduction of treated plant water into villages has had a stronger impact on the health status of both adults and children. However, the IV strategy may result into overestimation due to some omitted variables. Combining both OLS and IV estimates, I find that the illness incidence of adults decreased by 11 to 50 percent and their weight-for-height increased by 0.835 to 2.580 kg/m following the program implementation. There was also an improvement in self-reports of health. Children's weight-for-height and height itself both rose, by 0.446 to 0.754 kg/m and 0.962 to 2.489 cm respectively, as a result of the program. Using a variety of robustness checks, I show that the results are not driven by measurement errors, omitted variable bias, or attrition bias, and that the mechanism by which the program was effective was via improved water quality rather than simply via increased access to water. No obvious heterogeneous treatment effects are found across income and educational groups.

THE IMPACT OF WATER QUALITY ON HEALTH IN RURAL CHINA

By

Jing Zhang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Professor Judith Hellerstein, Chair
Professor Maureen Cropper
Assistant Professor Soohyung Lee
Assistant Professor Raymond Guiteras
Associate Professor Anna Alberini

Dedication

To my parents for their never ending support and encouragement throughout my whole study process.

Acknowledgements

Six years of Ph.D. study is coming to an end and this is very challenging and rewarding experience in my life. Here, I wish to express my gratitude to people who have helped me, one way or another, during these years.

First and foremost, I would like to thank my primary advisor, Judith Hellerstein, for her invaluable guidance and support. She is very kind and generous about spending time advising me in many ways, from research philosophy to even English writing. Deepest gratitude is also due to Maureen Cropper and Soohyung Lee without whose knowledge and suggestions this study would not have been successful. Thank you to Raymond Guiteras and Anna Alberini for serving in my dissertation committee.

Comments and suggestions on this work from Mark Duggan, Ginger Zhe Jin, Christopher McKelvey, Peter Murrell, and Alix Peterson Zwane are especially appreciated. I am also grateful to seminar participants at University of Maryland, the World Bank and National Graduate Institute for Policy Studies.

I thank Shufa Du for providing the confidential data and helping me understand the CHNS data, Rachel Rosenberg for her expert opinions on water-related diseases and Maryland Population Research Center for research support.

Special thanks to many professors in the department of economics at University of Maryland for teaching me and helping me get into the academic research. I would like to show my gratitude to Vickie Fletcher, Terry Davis, Elizabeth Martinez and other staff members for helping me in logistic work.

Finally, I want to send my sincere thanks to my family and friends. I could not have gone through this process without the support and encouragement from all of you.

Table of Contents

Dedication ii
Acknowledgements iii
Table of Contents v
List of Tables vi
List of Figures viii
Chapter 1 Introduction and Motivation 1
  1.1 Introduction 1
  1.2 Literature Review 6
  1.3 Background: the Water Improvement Program in Rural China 7
Chapter 2 Estimation on the Water Improvement Program in Rural China 12
  2.1 Data and Variables 12
  2.2 Estimation Strategy 15
  2.3 Summary Statistics 21
  2.4 Program Implementation 23
Chapter 3 Results for Adults 25
  3.1 OLS and IV Regression Results 25
    3.1.1 OLS regression results with fixed effects 25
    3.1.2 IV regression results 29
  3.2 Robustness Check 34
    3.2.1 Justification of the definition of the treatment variable 34
    3.2.2 Sensitivity analysis on the cutoffs used to define the treatment variable .. 35
    3.2.3 Omitted Variable Bias—Other Infrastructure Construction 36
    3.2.4 Channels Clarified—Safe Drinking Water versus Water Accessibility (Quality versus Quantity) 41
    3.2.5 Attrition Bias—Migration and the Data Attrition Problem 43
    3.2.6 Two Placebo Tests 45
  3.3 Heterogeneous Treatment Effects 47
Chapter 4 Results for Children 50
  4.1 OLS and IV Regression Results 50
  4.2 Robustness Check 53
    4.2.1 Sensitivity analysis with regard to the definition of the treatment variable 53
    4.2.2 Omitted Variable Bias—Other Infrastructure Construction 54
    4.2.3 Channels Clarified—Safe Drinking Water versus Water Accessibility (Quality versus Quantity) 56
  4.3 Heterogeneous Treatment Effects 58
Chapter 5 Conclusion 60
Bibliography 102

List of Tables

Table 1 the Cost of the Construction of Water and Sewage Systems

Table 2 Descriptive Statistics of Adults

Table 3 Descriptive Statistics of Children

Table 4 Descriptive Statistics of Infrastructure Variables

Table 5 Mean Differences between Characteristics of Treated and Untreated Villages

Table 6 Treatment Effects on Adults' Health Status

Table 7 Treatment Effects on Adults' Health Status in the Restricted Sample

Table 8 Assignment to Treatment for Adult Sample—OLS (First Stage)

Table 9 Treatment Effects of Water Program on Adults' Health Status

Table 10 Treatment Effects on Adult Health with and without Controlling for Nonlinear Terms of Age and Income

Table 11 Regression Results with Treatment Variables at Different Levels

Table 12 Treatment Effects on Adult Health across Different Cutoffs

Table 13 Treatment Effects on Adult Health Controlling for Households' Sanitation Facilities and Environment

Table 14 Treatment Effects on Adult Health Controlling for Other Infrastructure

Table 15 Treatment Effects Controlling for Water Accessibility

Table 16 Regression Results of Adults' Probability of Leaving the Sample

Table 17 Treatment Effects on Adult Health with and without Correcting for Data Attrition

Table 18 Treatment Effects on Adults' Illness Incidence

Table 19 Treatment Effects on Adults' Height

Table 20 Heterogeneous Treatment Effects on Adult Health across Income and Educational Groups

Table 21 Treatment Effects on Child Health

Table 22 Treatment Effects on Child Health in the Restricted Sample

Table 23 Assignment to Treatment for Child Sample—OLS (First Stage)

Table 24 Treatment Effects of Water Program on Children' Health Status

Table 25 Treatment Effects on Child Health with and without Controlling for Nonlinear Terms of Age and Income

Table 26 Regression Results with Treatment Variables at Different Levels

Table 27 Treatment Effects on Child Health across Different Cutoffs

Table 28 Treatment Effects on Child Health Controlling for Households' Sanitation Facilities and Environment

Table 29 Treatment Effects Controlling for Other Infrastructure

Table 30 Treatment Effects on Child Health with Controlling for Water Accessibility

Table 31 Treatment Effects on Child Health for Mangyo (2008)'s Sample

Table 32 Heterogeneous Treatment Effects on across Income and Parents' Educational Groups

List of Figures

Figure 1 Coverage of Water Plant versus Water Access from 1989 to 2006 (CHNS)

Figure 2 Proportions of Households Violating the Standards of Drinking Water in China in 2006

Figure 3 Map of Population Drinking High Fluoride Water in China in 2006

Figure 4 Fraction of Households reporting plants as water source in a village

Figure 5 Coefficients of the Treatment's Leads and Lags

Figure 6 Kernel Densities of the Treatment Variables with Different Cutoffs

Figure 7 Age Distribution of Missing Observations (CHNS)

Chapter 1

Introduction and Motivation

1.1 Introduction

Helping people gain access to safe drinking water is one of the most important health-related infrastructure programs in the world. As of 2007, around 1.1 billion people were still using unsafe water (WHO World Health Report, 2007). Pathogenic microorganisms in drinking water, the leading causes of diarrhea, have drawn a lot of attention in public health and other related fields. In addition, chemical impurities are growing threats in many developing countries, especially in ones experiencing rapid industrialization, such as China.

Viruses, bacteria, and parasites in water cause world epidemics of diarrheal illnesses (such as gastroenteritis, cholera, and typhoid), and in unsanitary environments without clean water, these diseases may result in severe dehydration and become life-threatening. This is the situation that confronts most developing countries today. About 4500 children under the age of five die from diarrhea in low-income countries every day.[1] For older children and adults, even when diarrheal diseases are not fatal, long-term suffering can lead to malnutrition and diminished productivity. Thus a variety of water-related interventions and trials, including water source treatments, point-of-use

---

[1] "Diarrheal Disease messaging" at http://rehydrate.org/diarrhoea/pdf/diarrheal-disease-messaging.pdf (accessed on May 23, 2011).

disinfection and adoption of improved hygiene, have been performed to tackle the problems of diarrhea and child mortality in developing countries. Although it is the most widespread result of poor drinking water quality, diarrhea is less prevalent in China than in other developing nations because of cultural norms of eating cooked food and drinking boiled water (for example, to make tea). Braudel states that Chinese "…were also concerned about the dangers of pollution and recommended boiling any suspect water" (230) around four thousand years ago. A recent Chinese study by Chen (2009) has shown that the diarrhea mortality rate in rural areas of seven Chinese provinces is 0.51 per thousand, which is much lower than the average (6.5 per thousand) in other developing countries. That study also notes that in China the diarrhea incidence rate is around 836 million per year, one third of which occur in children under the age of five.

Recently, chemical impurities—toxic metals, inorganic and organic compounds—are becoming new threats to drinking water quality in many developing countries. In addition to local soil constituents, human activities are increasingly contributing to the high concentration of chemical elements in water in developing countries. Vast discharges of industrial waste and excessive use of fertilizer and pesticides, along with relatively weak awareness or enforcement of government regulations, result in severe water pollution and, therefore, various diseases. For example, fluorosis is endemic in 25 countries worldwide (Erkin, 2009)[2] and arsenicosis in more than 70 countries (Ravenscroft, 2007). However, these kinds of water pollutants have heretofore been the subject of less public attention. The reason is that in reality the contents of chemical elements in water are minor and hard to detect. In addition, it usually takes long time

---

[2] "Facts and figures about water-related diseases", www.bpwnl.nl/water/arc/0610ff-diseases.doc (accessed on May 23, 2011).

(sometimes more than twenty years) for the caused medical symptoms to show up. Today in China, chemical impurities are the major threat to drinking water. A total of 1,115 counties and about 81.6 million people are at risk of fluorosis via drinking water, and 35 counties along with 385,000 people are at risk of arsenicosis (Chinese National Health Statistics, 2007). The Chinese government uses 0.05mg/L as the cutoff to define high-arsenic drinking water, whereas by employing the current WHO's guideline (0.01mg/L), the number of people exposed to high-arsenic drinking water may be over 15 million (Sun, 2003).

Many chronic diseases, including respiratory problems, skin lesions, spontaneous abortion, and even digestive cancers, can be induced by long-term exposure to poisonous drinking water. The World Bank's report (2007) estimates that about 66,000 people die from water pollution in rural China every year. The existence of harmful chemicals in drinking water has been an important impetus for the water improvement program in rural China since the 1980s and also influences the design and implementation of the program.

Health is of great concern because it not only affects people's wellness and perceived happiness, but also has substantial economic consequences. The influence of health is even greater and more direct in developing countries due to the fact that their health insurance systems are not well-established and that the majority of health expenditures are out-of-pocket. In low income countries out-of-pocket health expenditures account for over 60 percent of the total amount, as compared to 20 percent in high income countries (Schieber et al., 2006). In China, the report of the third National Health Service Survey (2003) shows that 79.1 percent of the rural population does not have any kind of health insurance. This number has decreased to 7.5 percent based on the preliminary results of the fourth National Health Service Survey (2008), as a result of the introduction of the New Rural Cooperative Medical Insurance scheme in 2003. Regardless, the benefits remain limited due to its low premiums.

Thus, medical treatments, especially those of chronic diseases, may lead to a large reduction in households' financial resources and drive them into poverty. It is estimated that every year, about 100 million people are driven into poverty due to unaffordable medical services (WHO World Health Report, 2005), and in rural China this number is about 10 million[3]. In addition, poverty may also make people more vulnerable to illness and trap them in a vicious cycle of diseases and poverty.

Considering the huge negative impact of diseases related to unsafe drinking water, a great number of governments and international organizations have launched water-related programs and interventions all over the world as an effective way to improve people's health and welfare. The United Nations seeks to "halve, by 2015, the proportion of people without sustainable access to safe drinking water" as one of the Millennium Development Goals.[4] The World Bank also places improvement of water and sanitation at the core of its efforts towards poverty reduction. The Chinese government began its nationwide water improvement program in rural areas in the 1980s. Since harmful chemicals in drinking water endanger hundreds of millions of people, the ultimate goal of this program is to provide widespread access to water from water purification plants. These plants can effectively eliminate both microorganisms and

---

[3] National Development and Reform Commission (NDRC), "Guidelines for pharmaceuticals industry development in the eleventh Five-Year Plan", http://www.sdpc.gov.cn/fzgh/ghwb/115zxgh/P020070927315215459276.pdf (accessed on May 25, 2011).

[4] The detailed information about United Nations' Millennium Development Goals is on the website: http://www.unesco.org/water/wwap/facts_figures/mdgs.shtml (accessed on May 25, 2011).

chemical impurities by employing appropriate clean water technology and equipments. In addition, the pipeline systems are combined to deliver plant water to households directly.

Based on the implementation process of the water program, this study estimates its causal effect on the health status of adults and children in rural China by employing data from China Health and Nutrition Survey (CHNS). This survey includes approximately 4,500 rural households in 152 villages, from 1989 to 2006 (7 waves). The treatment—improvement of water quality—employed in this study is defined by water sources (whether water comes from water plants) rather than just access to water. Ordinary least squares estimates show that the water quality improvement resulting from this program only has a moderate effect on health.

One of the significant challenges in estimating causal effects of government programs is the endogeneity problem generated by program placement. Since water facilities are usually constructed and financed by local governments, the underlying placement rule varies greatly across regions. Unobservables affect when and where water plants and pipelines are built and, therefore, lead to positive or negative bias of OLS estimates. To address this omitted variable (endogeneity) issue, I instrument for the treatment using the topographic characteristics of the villages, which are assumed to influence the costs of the construction of water plants and the introduction of pipeline systems. With this instrumental variable, the estimated impacts on the health status of adults and children become stronger.

1.2 Literature Review

Four kinds of interventions that have been used to fight against water-related diseases are: improved hygiene practices, improved sanitation, improved accessibility to water, and improved water quality. Many studies have been performed to evaluate these interventions and compare their effects in reducing the incidences of diarrhea and other diseases, especially for children. In the case of hand washing, the literature includes several dozen randomized trials (Curtis and Cairncross, 2003; Rabie and Curtis, 2006). There are, however, questions about whether compliance with disinfection protocols continues after studies end[5] and whether outcome variables (such as self-reported diarrhea) can be biased if experiments are not double blinded (Schmidt and Carincross, 2008). In the case of community water supply (e.g., piped water connections) and sanitation, controlled experiments are more difficult to conduct and the literature contains many observational studies as well as a few studies that evaluate impacts using quasi-experimental methods. The impact of having improved sanitary facilities and access to water is less clear (Merrick, 1985; Esrey, 1996; Jalan and Ravallion, 2003; Jacoby and Wang, 2004; Mangyo, 2008; Gamper-Rabindrani et al., 2008; Galiani et al., 2009).

In contrast, improved water quality has been shown to play a substantial role in reducing diarrhea and mortality in different countries (Cutler and Miller, 2005; Clasen et al., 2007; Arnold and Colford, 2007; Kremer et al., 2009). The quality of drinking water is a serious issue in China, especially when it comes to chemical pollutants. Ebenstein (2010) describes the strong link between river pollution and digestive cancers in China.

---

[5] The benefits of source treatments of drinking water are small compared to point-of-use treatments (Zwane and Kremer, 2007). Water treated at the source can easily become contaminated during transportation and storage.

In the 1980s, the Chinese government launched its water program, using water plants as a tool to help solve drinking water quality problems. During the implementation of the program, water access and quality have both increased, but along different growth paths. The CHNS data show that, in 1989, over 67 percent of rural households had water facilities (tap or wells) on their premises or inside their houses, while fewer than 21 percent of them had water from water plants. In 2006, these two numbers had risen to 98 and 42 percent, respectively. Figure 1 shows these trends in detail. In contrast to Mangyo (2008) who examines the impact of water access on child health in the early 1990s, in this study I focus on water quality improvement—the goal of the drinking water infrastructure program in China, and use the CHNS longitudinal data to estimate the impact of the water quality improvement program on the health status of adults and children in rural China, respectively. In Section 3.2.4 and 4.2.3, I discuss these two issues (water quality versus water access) and conclude that the program was effective through improved water quality rather than via increased access to water.

1.3 Background: the Water Improvement Program in Rural China

Since the 1950s, the Chinese government has made great strides in establishing the public water systems. By the 1980s, water treatment facilities had been built in almost all of the major cities to guarantee drinking water quality. But, the rural population, comprising more than 70 percent[6] of the total population in China, still had difficulty in accessing safe drinking water. The CHNS data shows that more than 70

---

[6] According to the 1982, 1990 and 2000 census data, the proportions are 79.1, 73.6 and 63.8 percent, respectively (China Statistical Yearbook, 2001).

percent of rural households were using untreated water from springs, rivers, lakes, or wells in 1989. In addition, human waste was often stored in open pits in household backyards, and livestock was reared within the village. According to the CHNS data, 68 percent of the households in rural areas used open pits as their toilet facilities and 64 percent raised livestock in 1989. The crude sanitation situation, along with poor hygiene practices, exacerbated the harmful influences of unsafe drinking water. It is estimated by that the average diarrheal incidence is 2.5 episodes per child per year among children under 5 years of age in rural China (Wei, 2008) and the diarrheal mortality rate of children under five is 14 times as great in rural areas as in urban areas (Tao, 2008). Regardless, diarrheal diseases are less prevalent in China than in other developing countries due to the fact that people customarily eat cooked food and drink boiled water (notably in making tea). Zhang et al. (2009) find that more than 85 percent of rural households boil water for drinking. A recent Chinese study by Chen (2009) has found, for example, that the diarrheal mortality rate in rural areas of seven Chinese provinces is 0.51 per thousand, which is much lower than the average (6.5 per thousand) in other developing countries.

In addition to diarrheal diseases, other diseases are caused by chemical impurities in water such as toxic metals and inorganic or organic compounds.[7] For example, fluorosis and arsenicosis, caused by high concentration of fluoride and arsenic in drinking water respectively, endanger tens of millions of people. A total of 1,115 counties and about 81.6 million people are at risk of fluorosis via drinking water, and 35 counties (385,000 people) are at risk of arsenicosis (Chinese National Health Statistics,

---

[7] Other endemic diseases, such as Keshan disease, Kashin-Beck Disease (KBD) and schistosomiasis, are considered in government yearly reports to be cured by improving water quality.

2007). Fluoride and arsenic are naturally occurring contaminants to a large extent. In addition, human activities have generated ever more severe water pollution in the process of China's industrialization. Toxic metals from industrial disposals and persistent organic pollutants from fertilizer and pesticides are jeopardizing human health through drinking water. In China, the quantity of industrial wastewater was 39.7 billion tons in 1997 and 49.7 billion tons in 2006 (China Water Resources Bulletin, 1997 and 2006). The consumption of chemical fertilizers increased from 17.75 million tons to 47.66 million in the period 1985 to 2005 and the use of nitrogenous fertilizers grew from 12.04 million tons to 22.29 million tons during the same period (China Statistical Yearbook, 2006). The World Bank's report (2007) estimates that in rural China about 66,000 people die from water pollution every year. One reason why the health damages of chemical pollutants in drinking water have not drawn much public attention is that it is hard to determine when small changes of chemical contents in nature become health risks, and relevant symptoms may need long-time exposure to pollutants to become detectable.

During the 1980s, the Chinese government started to launch a drinking water improvement program in rural areas. This program aims to build water plants to provide safe drinking water and pipeline systems to deliver it. The "Sanitary Standard for Drinking Water" and relevant guidelines for the program implementation in rural areas stipulate locations of water plants, safe drinking water standards (including aesthetic properties and general chemical, toxicological, bacteriological and radiative indexes), monitoring, etc. Water is treated using various technologies in four consecutive processes: coagulation, precipitation, filtration and disinfection.[8] Considering the

---

[8] Chlorination, one of the most popular methods to improve water quality in the world, is sometimes used in the disinfection process.

diversity of natural conditions, deep well pumps and rainwater harvesting systems have been installed as temporary substitutes in some areas. Nevertheless, appropriate water plant systems are the ultimate goal of the program. As shown in the fourth National Health Service Survey (2008), the proportion of beneficiaries is 85.3 percent, overall, whereas only 41.9 percent of rural people have access to water from water plants.[9] Currently in rural China, about three hundred million rural people still use unsafe drinking water. The construction of this program is still ongoing.

In terms of effectiveness of the water improvement program, Zhang et al. (2009) investigate water quality in rural households and find that, overall, water plants provide water with better quality than untreated water, as illustrated in Figure 2. This figure also shows that a greater proportion of households use drinking water that violates bacteria and coliform standards than violating other standards, which seems to contradict the argument that chemicals are the main pollutants in drinking water in China. However, as I mentioned in the beginning, the Chinese cultural norms of boiling water and cooking food can eliminate the pathogens to a large extent.

The program is financed through a variety of sources. The central and local governments, villages, rural households and other international organizations (such as UNDP, WHO, UNICEF, World Bank) all contribute to parts of the funds, but the ratios are quite different across regions. Poor areas are more reliant on outside funds from governments and international organizations, while in rich areas majority of funds come from beneficiaries directly and some private capital. From 1981 to 2002, it is estimated that total investment in the water improvement program was about 8.8 billion US dollars

---

[9] Ministry of Health (2009), Analysis on the Fourth National Health Service Survey, http://www.chinacdc.net.cn/n272442/n272530/n272742/29573.html (accessed on May 25, 2011).

(Meng et al., 2004). Overall, 25.7 percent of funds come from the central and local governments (in western regions of China, this proportion is as high as 50 percent), 26.9 percent from villages, 42.5 percent from beneficiaries and 4.9 percent as loans and donations from international organizations and other countries. The average cost of water plant and pipeline systems in this program is around 30 dollars per capita (Meng et al., 2004). The detailed information is presented in Table 1.

This study examines the impact of the water improvement program on the health status of adults and children in rural China by employing the CHNS longitudinal data. At least two possibilities could limit its impact. One is that the water quality from water plants may not be greatly improved relative to untreated water. This may be due to plant operations that do not meet government standards. The other issue is compliance; i.e., whether plant water is supplied continuously, 24 hours per day and 7 days a week, because water can also be contaminated by microbes during storage. Some studies show that mean coliform levels were considerably higher in households' water containers than in the original water sources (Fewtrell et al., 2005).

In the next three chapters, I describe the estimation strategies employed and evaluate the impacts of the water program on adults and children in rural China.

Chapter 2

Estimation on the Water Improvement Program in Rural China

2.1 Data and Variables

In this study I employ the China Health and Nutrition Survey (CHNS) dataset. The sample selection is based on a multistage, random cluster scheme in nine Chinese provinces: Liaoning, Heilongjiang, Jiangsu, Shandong, Henan, Hubei, Hunan, Guangxi, and Guizhou. In 1997 Liaoning was missing and Heilongjiang was included as a replacement. In later surveys, both provinces are covered. The ranks of per capita GDP of these nine provinces among 31 province-level administration regions recorded in the China Statistical Yearbook (2007) is the following: Liaoning(8), Heilongjiang(12), Jiangsu(5), Shandong(7), Henan(16), Hubei(17), Hunan(21), Guangxi(27), Guizhou(31). The average per capita GDP of these nine provinces is 16137 Chinese yuan (2024 US dollars), slightly smaller than the national average, 18662.52 Chinese yuan (2341 US dollars). In terms of the geographic regions, no provinces in North-west China and North China are included in this survey. Thus, the CHNS sample is not likely to be nationally representative. Regardless, in Figure 3 we can see that there is large variation across these provinces in the prevalence of fluorosis. And four counties are randomly selected from an income-stratified sample in each province generated from a weighted sampling scheme. The smaller sampling units, such as villages or towns, are then randomly drawn from each county.

The survey was taken in 1989, 1991, 1993, 1997, 2000, 2004, and 2006 so far, which allows me to explore the variation in program implementation during this period. As mentioned in Clasen et al. (2007), most trials and experiments can only focus on a small group and do not have sufficiently long follow-up periods (usually less than one year), which may lead to inaccurate estimates due to seasonality or the fact that some agents of infectious diarrhea are often delayed, such as campylobacteriosis. The desire to avoid this shortcoming is one reason why I choose to use the longitudinal data.

The CHNS data includes the samples in rural and urban areas and I restrict my analysis to the rural sample. In total, approximately 4,500 households and 152 villages or towns in rural China are included. In this study I do not distinguish between villages and towns.

One main advantage of the CHNS data in studying the health impact of water treatment plants is that it has multiple individual health indicators, including subjective (self-reported health status) and objective (weight and height) indicators, along with other demographic variables. In addition to the individual survey, a detailed community-level survey contains much information about infrastructure, which is useful in confirming that my results are not driven spuriously by variation in other infrastructure conditions across villages. The main problem with the longitudinal data is attrition, much of it being the result of substantial migration out of rural China since the 1980s. I find that the young, aged from 15 to 40, make up a large proportion of observations that attrit from the sample. This may cause bias if the migration is correlated with the water program implementation. In Section 3.2.5, I address this issue further.

The treatment variable in this study is defined based on the survey question answered by the households: “What is your water source?” The possible options include water plants, wells, springs, and rivers. Given the fact that households in a village live close together in rural areas, it is reasonable to expect that the water program is implemented at the village level and that there might not be much selection on the plant water coverage within the village. This is also justified by the CHNS data, showing that most of the proportions of households with plants as water sources in a village are very close to zero or one, as shown in Figure 4. Thus, in order to capture the water program implementation better and to avoid possible measurement error problems when using household reports as the treatment variable, I define the treatment at the village level by detecting a relatively large increase in the number of households who self-report access to water plants in a village. A dummy treatment variable—**water plant**—indicating whether the village is covered by the water improvement program is defined in the following ways:

- In the first survey year, the **water plant** = 1 if 80 percent or more of households in a village report a water plant as their water source.
- If **water plant** = 0 in the first wave, then in all subsequent waves **water plant** = 1 if there is more than a 20 percentage point increase in plant coverage for each year since the last wave. The reason for using the percentage change per year is the difference in time span between two consecutive waves of CHNS data. For example, if over 40 percent of households in a village report that their water sources switched to plants from 1989 to 1991, then the treatment variable—**water plant**—is set to 1.

- Once **water plant** = 1 in a given period, all subsequent periods are coded as 1.

The potential bias caused by measurement errors in household reports exists in OLS regressions; nevertheless, instrumenting for the treatment variable solves both the measurement error and the endogeneity problems, which provides me with a way to check whether the defined village-level treatment variable truly captures the implementation of the water improvement program. In Section 3.2.1 and 4.2.1, I show that the definition of treatment is reasonable by comparing the estimation results using household reports and my defined variable. It should also be mentioned that there are other possible water treatments (deep wells and rain harvesting systems) that serve as temporary substitutes for water plants in some areas, but they cannot be identified from the questionnaire. Thus, in this study I only focus on water plants, which may lead to underestimation of the impacts of the water improvement program.

2.2 Estimation Strategy

OLS regression with fixed effects

To estimate the causal effects of the drinking water infrastructure program on people's health status, the basic regression model is as follows:

$$Y_{ivt} = X_{ivt}\beta + T_{vt}\gamma + u_c + v_t + \varepsilon_{ivt} \qquad (1)$$

where $\mathrm{Y_{ivt}}$ is the health measure of person i in village v in year t. In empirical studies, it is always challenging to consider how to measure health precisely. In this study, three different health measures are employed.

I use a binary variable indicating whether the respondent has been sick during the last four weeks as one outcome variable. In the questionnaire, the survey question related to this variable is: “Have you been sick or injured within the last four weeks? Have you suffered from a chronic or acute disease?” Here I do not specify whether sickness is water related or not, based on the following considerations: 1) unsafe drinking water in China may cause multiple complicated symptoms, such as diarrhea, joint pains (from fluorosis) and skin rash (from arsenicosis and fluorosis); 2) the recording of symptoms of illness in the CHNS survey varies across waves (for example, diarrhea is combined with stomachache after 2000); 3) in the CHNS data, the incidence of diarrhea is less than one percent, which may result in inaccurate estimates of the program’s impacts when it is used as an outcome variable.

Considering that this indicator is self-reported, nutrition-based anthropometric measurements are also employed to estimate the impacts of drinking water improvement. Given the fact that lots of diseases besides diarrhea are caused by the water pollutants in China, these anthropometric outcomes may be able to capture the health gains in a more complete way. In addition, these measures are objective to large extent, or at least the measurement errors in them is less likely to be systematically correlated with respondents’ income than self-reported health measures (Strauss and Thomas, 1998).

The weight-for-height ratio adequately represents the long-run nutrition status of adults (Waterloo, 1972), which reflects both consumption and health status. Since the diseases caused by unsafe drinking water, whether diarrheal or other chronic diseases, usually lead to weight loss, weight-for-height is a reliable health outcome measure in this context.[10] For example, Cöl et al. (1999) state that the symptoms of acute arsenic poisoning include vomiting, diarrhea, and weight loss.

---

[10] Personal communication with Jin Jiang from Medical School, Tongji University, China, November, 2009.

In addition, the last four waves of CHNS (1997, 2000, 2004, and 2006) include self-reported health status of adults and this subjective measure is used as another adult health outcome as well.

For children (age from zero to 17), the outcome variable Y_{ivt} includes two anthropometric measures (weight-for-height and height), along with the illness incidence in the last four weeks. Anthropometric indicators of children are even more of interest in this context. Weight-for-height is considered to be an indicator of the short-term nutrition status of children, and height is a good way to measure the duration of children’s malnutrition and growth (de Onis, 2000). Briend (1990) states that there is no suggestive evidence of the causal relationship between diarrhea and children’s malnutrition since the catch-up growth reduces the harms caused by diarrhea. However, Humphrey (2009) shows that other water-related diseases, such as tropical enteropathy, can still cause malnutrition. In terms of chemical pollutants, Wang et al. (2007) present epidemiological evidence in China that that high concentration of fluoride and arsenic in drinking water is negatively correlated with children’s height, weight, IQ scores and lung capacity.

Child mortality is sometimes used as an outcome variable in other studies of improved water. However, in my sample fewer than 10 child deaths can be observed in each wave and, therefore, the impact of water from treatment plants on mortality must be small in terms of overall lives lost, and impossible to measure meaningfully given my sample sizes.

X_{ivt} represents characteristics of individuals, households and villages, such as age, sex, educational attainment of adults, household size, and distance to the nearest medical facility in adults’ regressions. When considering the impact on child health, I substitute parents' educational attainment for children's own education since the water program, as a health intervention, might also improve children's educational attainment (Bleakley, 2006; Politi, 2008). If parents' educational attainment is missing, I use that of the oldest male or female adults in the household as the proxies.

Considering that healthy people are able to work more productively and earn more money, I use income in the first wave when the household appears in the survey in order to avoid the possible endogeneity problem by using current income. Here, gross income instead of net income is employed because durable expenses are sometimes recorded in the survey which may lead to negative net income, as suggested by de Mel et al. (2007).

T_{vt} is a dummy variable indicating whether plant water is available in village v in year t (**water plant**). u_c and v_t are the region and year fixed effects respectively, and ε_{ivt} is the idiosyncratic error term. The average treatment effect γ can be consistently estimated by OLS regressions if the error term ε_{ivt} satisfies $E(T_{vt}\varepsilon_{ivt}|X_{ivt}, u_c, v_t) = 0$.

Instrumental variable strategy

A threat to the OLS regression validity in this context is the program placement issue. Consistent estimates of the causal effects of water quality improvement require that $E(T_{vt}\varepsilon_{ivt}|X_{ivt}, u_c, v_t) = 0$, which means that the installation of water plants and pipelines is exogenous or randomly assigned conditional on X_{ivt}, u_c and v_t. By employing the OLS strategy with region and time fixed effects, I am able to capture some unobservables that are constant across regions or years. However, considering the implementation and financing mechanisms of the water improvement program, some unobservables that vary within regions across time may affect the timing and locations of the program construction and, therefore, may generate an endogeneity problem.

One possible concern about the endogeneity of program placement is that local governments may prefer to carry out the program in the villages with high growth rates or with great potential to develop first (for example, villages near recently-built national highways). In these areas, well-established infrastructure can attract more investment. The governments have incentives to implement the program in those places to stimulate the local economy and to increase tax revenue. Furthermore, since, on average, the majority of funds come from villages and households directly, the program is also easier to fund in those places. People there may tend to be in better health than those in remote areas. If so, the positive relationship between program implementation and health implies that $E(T_{vt}\varepsilon_{ivt}|X_{ivt}, u_i, v_t) > 0$, which causes upward bias of OLS estimates. Program placement can also be negatively correlated with health, i.e., $E(T_{vt}\varepsilon_{ivt}|X_{ivt}, u_i, v_t) < 0$. Based on equity considerations, the central and local governments might be more likely to target locations by priority where people suffer from severe health impairments induced by drinking water. China's Eleventh Five-Year Plan (2006–2010) emphasizes expediting the water improvement program in areas suffering from high fluoride water, high arsenic water, high salinity water, and polluted water. The proportion of government investment in the program in western areas (which are relatively poor areas) can be as much as 50 percent. As a result, the OLS estimates are very like to be biased downward.

To address this endogeneity issue, I instrument for program placement using the topographic characteristics of villages (flat versus hilly or mountainous), which are assumed to influence the costs of the construction of water plants and pipeline systems in several ways. Fixed costs are higher in non-flat areas since it becomes more difficult to introduce pipes, and high-pressure water pumps must be installed to deliver water. As for variable costs, large amounts of electricity need to be consumed to pump water from plants to villages in hilly and mountainous areas. The system of regression equations, then, is:

$$\text{First stage: } T_{vt} = X_{ivt}\beta + Z_v\alpha + u_c + v_t + \mu_{ivt} \qquad (2)$$

$$\text{Second Stage: } Y_{ivt} = X_{ivt}\beta + T_{vt}\gamma + u_c + v_t + \varepsilon_{ivt} \qquad (3)$$

where Z_v is the instrument representing the topography of the villages.

The key identification assumption of the IV estimation strategy is that, conditional on demographic characteristics, household income, accessibility to medical facilities and the fixed effects, topographic characteristics of the villages should affect people's health status only through the quality of drinking water.

Topography, or land gradient, has been discussed in the literature as affecting agricultural productivity (Udry, 1996), crop types (Qian, 2008) and infrastructure construction (Duflo and Pande, 2007; Dinkelman, 2008; Donaldson, 2009). These factors may affect health mainly through household income. Therefore, controlling for household income in the regression can help the estimation satisfy the exclusion restriction when using the villages' topography as an instrument. In Section 3.2.3 and 4.2.2, I add other infrastructure information, including road construction, distance to schools, and electricity and telephone coverage as control variables in the regressions to test this assumption.

After using the instrument, the estimation parameter γ captures the local average treatment effect (LATE) and, more specifically, the weighted average of covariatespecific LATE (Angrist and Pischke, 2009). This refers to the weighted average impact of the water program on the villages whose implementation timing is affected by their topographic characteristics (compliers). The usefulness of IV estimates depends on the degree to which the impact on compliers can represent that on the total population. In my study the mechanism through which improved drinking water affects health is mainly a biological process, especially when chemical impurities are the major harmful contents. I would not expect that human behavior influences the program's impact to a great extent, which is also justified by the fact that I do not find heterogeneous treatment effects across income and educational groups in Section 3.3 and 4.3. Thus, the LATE on the compliers should be quite similar to the average treatment effect (ATE) on the whole population.

2.3 Summary Statistics

Table 2 presents the summary statistics of adult observations in all the waves. The illness incidence in last four weeks for adults is around 9.0 percent, on average, which is much lower than the national survey results. According to the National Health Service Survey (1998, 2003, 2008), the two-week morbidity rates in rural China were 13.7 percent in 1998, 13.9 percent in 2003 and 17.7 percent in 2008. The low rate in the 1989 wave is due to the fact that this question targets only individuals under age 7 or between 20 and 45. Despite this fact, this number in other waves remains low.

For the anthropometric measure of adults,[11] WHO regards a Body Mass Index (BMI) under 16 severely underweight and over 40 Obese Class III. In order to avoid errors in the survey data, I exclude 230 adult observations whose BMI is less than 10 or

---

[11] In the 1989 wave, the majority of height records are for preschool students and adults aged from 20 to 45.

greater than 46. The average weight-for-height ratio of adults is 35.485 kg/m, and the average BMI is 22.20. In my sample, 8.55 percent of adults are considered underweight (BMI<18.5), while 16.74 percent are considered overweight (BMI>25). Columns (2) and (3) show the means and standard deviations of the variables in villages with and without access to water from water plants for all of the waves respectively, and Column (4) compares their differences. Not surprisingly, people in the treated villages have higher socioeconomic status. We see that in the villages with plant water, adults are relatively wealthier, more highly educated and less likely to raise livestock than those in the villages without.

Table 3 presents the summary statistics of child observations when pooling all the waves. The illness incidence in last four weeks for children is 6.6 percent, on average. In order to eliminate the possibility of misreporting height, I employ the WHO growth table to calculate z scores and exclude the observations with z scores below -6 and above 6.[12] The outlier cutoffs recommended by WHO are around 5. According to Chang et al. (2006), which studies the growth characteristics of children under age 5 in China between 1990 and 2005, in rural China, 41.4 percent in 1990 and 13.1 percent in 2005 showed stunted growth. Therefore, I extend the normal range of children's height and exclude 146 observations. For the weight-for-height measure, I use similar cutoffs as adults and drop 138 child observations. The means of children's weight-for-height ratio and height are 21.429 kg/m and 124.338 cm, respectively, as shown in Table 3.

Columns (2) and (3) in Table 3 list means and standard deviations of the variables in villages with and without plant water for all of the waves, and Column (4) compares

---

[12] The means and standard deviations of the reference population are for age in months in the WHO growth table; here, I choose just the means and standard deviations in the 6th month for each age.

their differences. Children in the treated villages are healthier and live in households with higher socioeconomic status. For example, in the villages with plant water, children are heavier (0.484 kg/m) and taller (2.896 cm) and their parents have more years of education (0.436 for fathers and 0.933 for mothers) than those in the villages without.

Table 4 presents the descriptive statistics of some variables about households' environments and villages' infrastructure conditions. These variables are used in the robustness check to see if the baseline regression results are stable after controlling for them. They have been shown to influence individuals' labor supply decision (Dinkelman, 2008) and households' income. However, since their direct impact on health may be secondary, I do not include them as control variables in the baseline regressions. From Table 4, we can see that the means and standard deviations are very comparable in the adult and child samples. For example, the average distance to the nearest middle school is 1.831 km in the adult sample and 1.834 km in the child sample.

2.4 Program Implementation

In Tables 2 and 3, I compare the characteristics of individuals and households in villages covered by the water improvement program and with those in villages not covered by the program. Generally, people in treated villages are wealthier and have more educational attainment. However, no causality can be concluded through the simple comparison; this correlation can be explained by the fact that wealthier villages can afford the program or that improved water quality makes people healthier and, therefore, able to earn more money, or both.

In order to see if there is any selection rule when implementing the water improvement program, I compare the demographic and infrastructure characteristics of treated villages one to five years before the treatment with those of villages that had not been treated by 2006. Table 5 shows the means of the characteristics' differences (untreated years of treated villages one to five years before the treatment – untreated villages) after controlling for year fixed effects. We can see that there is no suggestive evidence that the program is more likely to be launched in richer areas or areas with better infrastructure conditions, except that the areas closer to a medical facility tend to be covered by the water program earlier.

Chapter 3

Results for Adults

3.1 OLS and IV Regression Results

3.1.1 OLS regression results with fixed effects

Table 6 presents regression results with different levels of fixed effects and for different outcome variables for adults. All standard errors are clustered at the village level. Generally, the regression coefficients are more pronounced when controlling for the county fixed effects than for the village fixed effects. With the county fixed effects, OLS regression results show significantly positive estimates of the program's impact on adult health. On average, a water treatment plant decreases adults' illness incidence by one percentage point and increases their weight-for-height ratio by 0.835 kg/m. Given the average illness incidence of about 9.0 percent, the estimate actually implies that the probability of illness in the last four weeks is reduced by about 11 percent when a water treatment plant is present. The self-reported health status also rises by around 0.027 point. Although this estimate is short of significance, its p-value is very close to 0.1.

The estimated coefficients on the other covariates in the county fixed-effect specification are sensible in signs and magnitudes. Health decreases with age. Males report better health status than females do. Men's illness incidence is one percentage point lower than for women; they have 0.063 point higher self-reported health status, and are 1.780 kg/m heavier in weight-for-height ratio. A one year increase in educational attainment is associated with a 0.1 percentage point reduction in illness incidence, a 0.014 point increase in self-reported health status and a 0.087 kg/m rise in weight-for-height. Married persons tend to feel healthier and are heavier than single people. Given constant household income, a larger household implies that fewer resources are allocated to each member, which may worsen household members' health. Nevertheless, the estimation results present mixed evidence: Household members in large families are less likely to be sick and evaluate their own health as being better, though they weigh a little less. The positive correlation between health and income may be due to non-market home production and the existence of economies of scale (public goods) in a household (Barten, 1964). Household income is negatively correlated with illness incidence and positively correlated with self-reported health status and weight-for-height. Raising livestock shows a minor negative correlation with adult health: It is associated with a 0.854 kg/m reduction in weight-for-height. The distance to a nearest medical facility does not have a significant impact on adult health.

Regression results with the village fixed effects are also presented in Table 6. The estimated coefficients of the treatment variable vary from ones with the county fixed effects in terms of magnitudes and significance, whereas most of other covariates show very similar relationships with health. The water improvement program is estimated to lead to a two percentage point reduction in adults' illness incidence, a 0.030 point increase in their self-reported health status, and 0.279 kg/m rise in their weight-for-height. However, none of these health benefits are statistically significant.

An important question here is which regression specification is better in this context when employing a difference-in-difference strategy: the county fixed effects or the village fixed effects?

Given the huge population and vast land area in rural China, the water improvement infrastructure program has been rolled out slowly in terms of coverage. Figure 1 shows that over 20 percent of villages in this sample already had access to plant water before the first wave and around 60 percent were not yet covered by this program by the last wave. The treatment status of those villages, which constitute a majority of the sample, stays constant during the survey period. As a result, the health status of residents of those villages does not contribute to the estimation of the magnitudes of treatment effects when employing village fixed effects, but does contribute to identification when county fixed effects are used instead. Not surprisingly, then, the standard errors of the estimates of the effects of water treatment plants using village fixed effects are larger for all three of the health outcomes, and more than doubled for two of them as compared to those using county fixed-effects This partially explains why the estimates on the water plant treatment variable are not statistically significant with the village fixed effects. Bootstrap Hausman tests are implemented to see whether those estimated treatment effects are different from the ones estimated using county fixed effects. As shown in Table 6, the p-values from the Hausman tests are over 0.1 for two of the outcomes—illness in the last four weeks and self-reported health status, but not for weight-for-height, where the point estimate of the effect of water treatment plants on weight-for-height is only one third as large when using village fixed effects, although still positive.

Furthermore, I restrict the sample to the villages that started to access plant water between the second wave (1991) and the final wave (2006) so that at least one of their pre- and post-periods can be observed. In total, 37 out of 152 villages and around 9000 adult observations are included. The estimation from the full sample with the village fixed effects comes mainly from the variation in these villages. Thus, if employment of the county fixed effects still leads to similar treatment effects in this restricted sample, it suggests that the baseline estimates from the full-sample country fixed effects are not driven by the simple comparison between villages treated before the first wave and those having not been treated by the last wave—that is, by sets of villages that may be very different from each other in unobservable ways.

Table 7 shows the estimated treatment effects of the water improvement program on the restricted sample. We can see that with county fixed effects, adults' illness incidence in last four weeks decreases by 2.1 percentage points with access to plant water, which is equivalent to a 25.8 percent deduction give the average incidence is 8.13 percentage points in this subsample. And the estimated treatment effect is significant at the 10 percent level. A water plant increases adults' self-report health status by 0.059 point, but not significantly, and raises their weight-for-height significantly, by 0.727 kg/m. To summarize, the impact of plant water on adult health in villages that started receiving it only during the survey period is comparable to the impact on the whole sample.

Considering the problems of drinking water in China, the long-term benefit of the program may be even larger because some chronic diseases caused by chemical impurities occur only after long-term exposure to unsafe drinking water. Inclusion of the villages with constant treatment status—that is, those that have been exposed for many years to water from treatment plants and those that have never been exposed—is very informative in evaluating this program. Thus, in the rest of this dissertation, I will focus on the OLS estimates with the county fixed effects.

To explore the impact patterns of the water improvement program and to check for the existence of a pre-existing health trend in the treated villages, I substitute leads and lags of treatment for the single treatment variable in the regressions. Due to the definition of the treatment and the time spans between CHNS survey waves, only several specific leads and lags can be identified. The estimated coefficients of the leads and lags and 95 percent confidence intervals are drawn in Figure 5. We can see that there is no clear evidence of a positive or negative health trend before the program implementation, in spite of a few significant coefficients. In most of the cases, the health benefits occur right after usage of plant water and remain persistent afterwards.

Although Table 5 and Figure 5 do not suggest any strong program placement problems, it is reasonable to expect that areas where people are suffering from water-related diseases may have priority in program implementation, according to the policy guidelines of the Chinese government. As a result, the estimates of treatment effects in the baseline regressions could be downward biased, something that is suggested by the results of the instrumental variables strategy. However, they are still meaningful since they likely inform us of the lower bounds of the program's impacts.

3.1.2 IV regression results

The first stage regression results of program assignment are presented in Table 8. Compared to flat areas, the probability of villages being covered under the water program in non-flat (hilly and mountainous) areas is 41.9 percentage points lower. Similar to the relationships between the covariates and the treatment shown in Table 2, household income and residents' educational attainment are positively conditionally correlated with the treatment variables. However, this conditional correlation does not shed light on the direction of causality. Table 8 shows that the villages with access to water plants also have more females. One possible explanation is that better infrastructure conditions can help females perform agricultural production alone, which may allow males to work outside of villages. The negative conditional correlation between the program placement and the indicator of livestock raised might be due to the fact that more urbanized areas are less likely to have the space or conditions to perform livestock and poultry farming. The F-statistic on the instrument is around 17.29, which implies that topography is not a weak instrument in this context. The rule of thumb suggested by Staiger and Stock (1997) is that F-statistic should be greater than 10 when there is only one endogenous regressor. It can guarantee that the maximum bias of Two Stage Least Squares (2SLS) estimates is less than 10 percent.

In the CHNS survey, communities' topography is described by three different categories: flat, hilly, and mountainous. Therefore, in theory, two dummy variables can be defined and used as the instruments for the treatment variable. However, the F-statistic in the first stage is around 9, so the weak instrument problem exists when using those two instruments. Moreira and Cruz (2005) and Mikusheva and Poi (2006) suggest that estimating confidence intervals inverted from fully-robust tests under weak instruments is also meaningful. Chernosukov and Hansen (2007) provide a new method to deal with heteroscedasticity of standard errors in this scenario, but their method becomes less powerful when the number of instruments exceeds that of endogenous variables. Therefore, they suggest making the model just identified by eliminating some instruments if the explanatory power of the remaining instruments does not decrease much. Based on the above considerations, I choose to use one combined instrument—non-flat—in this study. Furthermore, the estimates of treatment effects by using the combined instrument are very similar, in both magnitudes and significance, to the ones with the two instruments in the basic specification.

Table 9 presents results of instrumental variables regressions for different adults' outcome variables. Here, all standard errors are also clustered at the village level. As compared to OLS estimates, the IV strategy generates stronger and statistically significant effects on behalf of the water intervention. The probability of adults' illness incidence in the last four weeks decreases by 4.5 percentage points, or 50 percent, after villages are provided with plant water. Self-evaluation of health status increases by 0.144 and, objectively, adults' weight-for-height also shows a significant 2.580 kg/m gain, which is equivalent to saying that an individual who is 180 cm tall gains 4.68 kg. If adults' BMI index is employed as an outcome variable, the estimate on **water plant** is 1.26 and significant at the one percent level. The coefficients of the other covariates are very similar to the ones in OLS regressions in terms of magnitudes and significance.

From the above table, we see that the OLS and IV strategies both generate positive impacts of the water improvement program on the health status of adults; the difference is that IV estimates are larger in magnitudes. However, there is still a concern about the validity of using topography as the instrument in this context, i.e., whether the exclusion restriction holds. Flat areas are where town centers are located and, therefore, are more developed and have better infrastructure and social services, as discussed in Lipscomb et al. (2008). Thus, the instrumental variables strategy might lead to upward-biased estimates, providing us with the upper bounds of true estimates.

Combining both OLS and IV estimates, I find that the illness incidence of adults decreases by 11 to 50 percent and their weight-for-height increases by 0.835 to 2.580 kg/m following the program implementation. Chinese National Health Service surveys show that digestive diseases accounted for 26.8, 25.9, 23.6, and 16.1 percent of two-week morbidity in rural China in 1993, 1998, 2003, 2008, respectively. Excess amounts of fluoride and arsenic can also cause other kinds of diseases, such as skin and respiratory diseases (which accounted for 3.1 and 50.4 percent, respectively, of two-week morbidity in rural China in 2008). In addition, the average weight gains from the OLS and IV estimation are 1.57 and 4.85 kg, which imply 2.7 and 8.5 percent increases, given that the average adult weight is 57.323 kg. Milne et al. (2006) conduct meta-analysis on the effects of protein and energy supplementation on the elderly and show that their weight change is around 2.5 percent in the short term. Thus, it can be concluded that the estimated treatment effects from OLS and IV estimation provide us with reasonable ranges of the impacts of the water improvement program.

As I mentioned earlier, given the slow coverage of this water infrastructure program and the fact that some villages have just recently been included, I expect that the long-term benefit of the program may be even larger.

In this study, the main treatment variable—water plant—is a dummy variable indicating whether the village has access to water from water plants. And one of my outcome variables, illness incidence in last four weeks, is also binary. In the baseline specifications, I estimate the effect of having water from a treatment plant using 2SLS, assuming linearity of the first and second stages. Abadie (2003) proposes a way to deal with scenarios in which the instrument and endogenous regressor are both bivariate along with a dummy or continuous outcome variable. However, given the limited sample size, I am able to apply this method only in the regression with adults' weight-for-height as the outcome variable. The estimated health gain is 2.8 kg/m and significant at the one percent level.

Nonlinear profiles of age and household income when considering health have been the subject of much research. In this study I use different specifications for adults and children, which partially helps me avoid the nonlinear relationship between age and health. Since rural households are relatively poor, it is unlikely that a negative impact of income on health exists. To test the above hypothesis, I add age-squared and income-squared terms as control variables in the baseline regressions. The OLS and IV regression results are shown in Table 10. We can see that inclusion of these nonlinear terms barely changes the significance and magnitudes of the baseline estimates of the impact of plant water.

In the remainder of this dissertation, I conduct several robustness checks on both OLS and IV estimates to see whether they are stable and how they may vary with people's demographic characteristics.

3.2 Robustness Check

3.2.1 Justification of the definition of the treatment variable

In this robustness check, I tackle whether the defined village-level treatment variable is correct: that is, whether it reflects the implementation of the water improvement program in rural China. Since water plants and pipeline systems are constructed at the village level, all of the households in a village are assumed to be covered at almost the same time. Therefore, the coefficients of the treatment variable I define at the village level should be very similar to ones that use individual household reports as the treatment variable. I expect this to be true especially when I compare their IV estimates since the instrumental variables strategy helps correct the bias caused by measurement errors. In this check, I instead define the household-level treatment variable to be 1 if a household reports a plant as its water source, and 0 otherwise. It turns out that 14.8 percent of the household observations show a discrepancy in values between the household-level reports and my village-level constructed treatment variables.

Table 11 shows the OLS and IV results using these two treatment levels. We see that IV regressions generate very similar impacts on adults. For example, the estimate of household reports on adults' weight-for-height ratio is 2.683 kg/m versus 2.580 kg/m of the village-level treatment variable reported in Table 11, Column 6. Moreover, their OLS estimates are also close in magnitudes. The coefficient on adults' weight-for-height is 1.027 kg/m for the household-level definition and 0.835 kg/m for the village-level definition. In conclusion, the similarities between the OLS and IV estimates suggest that the defined village-level treatment variable—water plant—does reflect the program implementation to a large extent.

3.2.2 Sensitivity analysis on the cutoffs used to define the treatment variable

In this section, a sensitivity analysis is conducted that tests whether estimates of the impact vary with the criterion I use for defining when a village obtains plant water. As mentioned in section 2.2, I define a threshold number—20 percent—to detect the change in the proportion of households that have access to plant water in a village. Here I use five different cutoffs—10 percent, 15 percent, 20 percent, 25 percent and 30 percent, respectively—to construct a treatment variable. Only 15.6 percent of the household observations do not have the same values for those five treatment variables. Their kernel density plots, drawn in Figure 6, imply that they are quite similar.

The OLS and IV regression results for each treatment variable are presented in Table 12. For the OLS estimates, the significance and magnitudes do not change dramatically. For example, the probability of being ill for adults in the last four weeks decreases by 1.2 percentage points when using a 10 percent cutoff, which is quite similar to the 1.0 percentage point decline with a 20 percent cutoff and 1.3 percentage point decline with a 25 percent cutoff. All of these are statistically significant. When 15 percent and 30 percent are employed, the estimates are slightly smaller (0.6 and 0.9 percentage point) and become insignificant. The magnitude of the estimated impact of plant water on adults' weight-for-height ratio is slightly different, along with the criterion used, and varies from 0.751 kg/m to 1.087 kg/m, while the significance stays the same.

In addition, the point estimates and statistical significance of the IV estimates are also stable. An interesting pattern is observed: The impact becomes slightly stronger when a stricter criterion is applied. For example, the probability of illness for adults in last four weeks decreases with water treatment by 3.6 to 5.4 percentage points (corresponding to a 40 to 60 percent reduction), and their weight-for-height increases from 2.077 kg/m to 3.111 kg/m, when the cutoff increases from 10 to 30 percent. Therefore, both the OLS and the IV estimation results in Table 12 suggest that access to water from a water treatment plant does benefit adult health significantly, although the magnitudes of the influences vary slightly with the criteria employed. Ultimately, this exercise shows that 20 percent seems an appropriate cutoff to be applied in this study.

3.2.3 Omitted Variable Bias—Other Infrastructure Construction

The basic assumption of validity of the OLS in this study is that, conditional on covariates controlled in the regressions, the treatment variable should be uncorrelated with the error term. And the instrumental variable strategy also relies on the assumption that topographic characteristics of the villages should affect people's health only through water quality, when controlling for those covariates. Since poor sanitation conditions can lead to water-related diseases and topography has been argued to influence several kinds of infrastructure construction (such as road construction and electrification), in this section I add these controls in the regression specifications to see if the baseline estimates are still robust.

Sewage and sanitation environment

Most diarrheal diseases occur through oral-fecal or hand-to-mouth transmission. Therefore, when studying water interventions, the sewage and sanitation environments are also considered important factors affecting people's health since they may work interactively with drinking water. Thus, sanitation might be an omitted variable that could bias estimates of the effect of the water program.

In the public health literature, findings about the complementarity between water improvement and other interventions are somewhat limited. Fewtrell et al. (2005) point out that combined interventions do not have an advantage in reducing diarrheal incidences over those with a single focus. Zwane and Kremer (2007) and Clasen (2007) also present similar findings showing no statistically significant additional effects from combined interventions. They argue that it is consistent with epidemiological models that a large dose of pathogens can cause diseases. Once a single intervention reduces the volume of pathogens to a certain threshold, additional efforts may not generate extra benefits.

Besides water and sanitation, hygiene interventions are implemented to reduce water-related diseases in developing countries. A review paper by Curtis and Cairncross (2003) notes that hand washing reduces diarrhea risk by 47 percent. Hygiene education is executed alongside the water improvement program in rural China. However, I cannot disentangle its impact given the fact that there is no information about hygiene practices in the CHNS survey.

Since the 1980s the Chinese government has also promoted a sanitation improvement program—a disposal system called “Rural Ecological Sanitation”—in addition to pipeline flushing. In this setup, excrement flows into a sealed biogas tank under a household bathroom. Then, after biomass gasification, gas can be used as fuel and remains (which contain no bacteria) can be used as a safe fertilizer containing plenty of nitrogen, potassium, phosphorus and organic components. The reasons for promoting this waste disposal system are several. Rural people have used human and livestock waste as fertilizers for a long time, which can help them save money on chemical fertilizers. In addition, rural households traditionally used wood and straw as fuels, producing significant indoor pollution (smoke) and doing harm to people's health. Also, the use of traditional fuels may result in deforestation. The ecological sanitation system helps solve all of these problems at one time.

In terms of implementation, the sanitation program significantly lags the water improvement program. The percentage of total beneficiaries was about 33 percent in 2006, but it varies a lot across the country. In the northern and western provinces, this number is below 10 percent (The Ministry of Health of China, 2006). Since it is hard to clearly identify the sanitation improvement program from the CHNS household survey and its coverage is relatively low, in this study I only focus on the water program.

In the robustness check, in order to consider how much my baseline estimates may be affected by ignoring sanitation conditions, I control for households' toilet types and sanitation environments (interviewers' evaluation of the amount of excreta around households' dwellings) in the regressions. The estimation results are shown in Table 13. Overall, we see that adding sanitation controls to the model causes the coefficients on the water treatment plant variable to be smaller and less significant for both OLS and IV estimates. In Table 13, the OLS estimate, when the outcome is adults' illness incidence, decreases from 1.0 to 0.8 percentage point in the magnitude and becomes insignificant. And the estimate for adults' weight-for-height, while still significant, decreases from 0.835 to 0.467 kg/m. The coefficient for self-reported health status decreases by a small amount, from 0.027 to 0.023, and stays statistically insignificant. For the IV estimates, as compared to the OLS estimates, the decreased amounts are less and the changes in significance are the same. The illness incidence drops by 4.2 percentage points and is no longer significant. The estimate on adults' weight-for-height drops by a small amount, from 2.580 to 2.068 kg/m and keeps significant. One exception is that the coefficient for self-reported health status increases slightly, from 0.144 to 0.167.

Some caution must be exercised in interpreting these regression results since households' sewage facilities and sanitation environment may be endogenous to access to plant water or piped water. For example, Bennett (2008) proposes a possibility—a moral hazard issue that piped water could worsen the sanitation environment since the marginal health benefit of clean surroundings decreases. The opposite may also be true: I.e., access to piped water decreases the opportunity cost of households' use of flush toilets and cleaning of houses and surroundings. The CHNS data support the second argument: Households' adoption of flush toilets is positively correlated with access to plant water and, on average, happens 1.1 years after the water program is implemented. Therefore, the change of estimates here might be due to the fact that including sanitation implementation "over-controls" for the treatment effects of water plant implementation.

A poor sanitation environment is likely to counteract the impact of water improvement on diseases caused by microorganisms, but it is less likely to do so when water pollution comes, instead, from chemical impurities. As I showed earlier, the inclusion of sanitation information does have an influence on the coefficients of improved water, although not to a large extent.

Other infrastructure (roads, distance to schools, electricity, etc.)

Besides sanitation, some other infrastructure construction, such as electrification and roads, can also be affected by the land gradient. However, when considering determinants of health, the benefits from these infrastructure components may be secondary and may be captured to a large extent by household income. If this is true, the inclusion of other infrastructure conditions should not change my results much. In Table 14, I present such tests by controlling in the regressions for road construction (dirt, stone or paved), distances to schools (closest primary and middle schools), accessibility to trade areas, and telephone and electricity availability. We see that only a few of the coefficients of these infrastructure variables are significant. Regardless, the magnitudes and significance of OLS and IV estimates of the health impact of plant water for adults (a 1.2 percentage point reduction in illness incidence, a 0.033 point rise in self-reported health status and a 0.646 kg/m increase in weight-for-height from OLS regression results; a 4.9 percentage point reduction in illness incidence, a 0.176 point rise in self-reported health status and a 2.381 kg/m increase in weight-for-height from IV regression results) are almost the same as my baseline estimates (a 1.0 percent point reduction in illness incidence, a 0.027 point rise in self-reported health status and a 0.835 kg/m increase in weight-for-height from OLS regression results; a 4.5 percent points reduction in illness incidence, a 0.144 point rise in self-reported health status and a 2.580 kg/m increase in weight-for-height from IV regression results). This test shows that these infrastructure conditions do not seem to have first-order influence on the estimated impacts.

3.2.4 Channels Clarified—Safe Drinking Water versus Water Accessibility (Quality versus Quantity)

Unlike some studies focusing on piped water, in this study I define the treatment based on households' water sources to address the importance of water quality in China. But it is true that the introduction of pipelines in this program also improves households' access to water. This is an issue of water quality versus water quantity. Water quality affects people's health directly through microbial contents and toxic elements in drinking water. Access to water generally benefits people's health in a more indirect way. For example, access to water can enable people to save time, leading to increases in labor supply and, therefore, increases in household income. As a result, people's health is improved because more resources are allocated to the consumption of household members. Weak evidence has been found to link water quantity and health in the literature. One piece of research close to this study is Mangyo (2008), which uses the CHNS data and does not find any significant impact of water access on the health status of children under age 10 in China in the early 1990s. Clasen et al. (2007) point out one possible explanation that water supply interventions take effect only if there is direct-connection provision to households and water is used with no storage.

More importantly, an emphasis only on water access could actually exacerbate the poor quality of drinking water. The CHNS data show that in 1989, 67 percent of rural households had in-yard water, so called "optimal water access," and 66 percent of them used untreated well water. It has been found that the concentration of some chemical impurities (such as fluoride or arsenic) in underground water is even higher than in surface water since these elements come mainly from local soils or rocks. Furthermore, disentangling the effects of water quality improvement and water access in this context has an important policy implication. In Figure 1, we see that in 2006, around 98 percent of rural households could access water on their property. If the impacts of improved water that I estimated before were generated from better water access only, then the water program may not need to be continued.

In this section, I test the hypothesis that it is water quality, not access, that drives my results. Here the variable that refers to households' water accessibility—water access—is constructed in the same way as in Mangyo (2008). A question in the CHNS survey is asked at the household level: How does your household obtain drinking water? 1) in-house tap water; 2) in-yard tap water; 3) in-yard well; 4) other place. Then, the water access is coded as 1 if the answer is 1), 2), or 3) and as 0 otherwise. With this newly constructed variable, several sets of regressions have been run in the following way.

First, I add water access as one of the control variables. Panel A of Table 15 shows the estimated coefficients of both water quality and water access variables in the OLS and IV specifications. For both regression results, all estimates on the water quality variable stay almost the same as the baseline ones in terms of magnitudes and significance, while for water access, only the estimate for adults' weight-for-height ratio is significant in the OLS specification. The OLS estimated impact of having plant water on adults' illness incidence remains at one percentage point; the impacts on self-reported health status and on weight-for-height are a 0.026 point increase and a 0.795 kg/m gain, respectively. The IV regressions show that adults' illness incidence decreases by 4.5 percentage points with exposure to treated water, their self-reported health status rises by

0.136 point, and their weight-for-height ratio increases by 2.614 kg/m. This suggests that the impacts estimated in the baseline regressions come from water treatments in water plants.

Next, I restrict the sample to households that can access in-yard water (optimal water access) in all of the waves when they exist and see how the existence of water from a treatment plant affects adults' health in these households. The regression results are presented in Panel B of Table 15. We see that the estimates of water improvement are quite similar to those from the baseline regressions, which also supports the hypothesis that it is quality, not access, that improves health. For example, the OLS estimate for adults' weight-for-height is 0.996 kg/m and the IV estimate is 3.154 kg/m, as compared to 0.835 and 2.580 kg/m. To summarize, the health gain predicted by the baseline regressions come mainly from the improvement of water quality, which plays a more important role than water access in the context of Chinese drinking water problems.

3.2.5 Attrition Bias—Migration and the Data Attrition Problem

In China, there has been much migration out of rural areas since the late 1980s because, beginning at that time, the old household registration (Hukou) system and consequent legal urban-rural segregation were relaxed (Zhao, 2003). According to the 2000 census data, there are 12.46 million migrants, comprising 10.6 percent of the total population, and 58.9 percent of those migrants come from rural areas (Cai and Wang, 2003). The CHNS data show that 40 percent of individuals cannot be tracked during the whole period from 1989 to 2006. Figure 7 shows the age distribution of observations that attrit from the sample; we see that people aged 15 to 40 account for a large proportion of them.

Sample attrition is of concern in this study if it is correlated with the treatment variable. The sign of the bias from attrition is theoretically ambiguous. If access to plant water in a village makes people healthier the young, healthiest group in the entire population may be more likely to move out to look for a job, compared to those in villages without access to water plants. As a result, the estimates would be downward biased. The opposite situation may also be true, however: The younger people in untreated villages may be willing and able to leave home to escape a dirty environment. Thus, the impacts estimated from the regressions would be over-estimated.

To check if the sample attrition sorts on the treatment, I regress the probability of adults' not being present for the next survey wave on the treatment and demographic characteristics in the current survey year, using a variety of different specifications.[13] Table 16 shows that this probability is positively correlated with the bivariate treatment variable, but not statistically significantly. This provides relatively weak evidence that sample attrition sorts on the implementation of the water improvement program in either direction.

Another way to test whether the treatment effects are biased is to use inverse probability weighted (IPW) estimators, which assume that attrition can be explained by observables. To correct for attrition bias, this method puts more weight on the observations that have characteristics similar to those who end up leaving the survey. Wang (2008) employs this procedure for the CHNS attrition problem. Here I focus only

---

[13] Besides checking the probability of being missing, I also compare the demographic characteristics of missing people in the treated and untreated villages. No significant differences show up, except for marital status.

on adults present in the first wave and implement the procedure in the following two steps: First, I estimate the probability of staying in each wave after 1989, based on the individuals' characteristics in the first wave (1989), and then I use the inverse of these probabilities as weights to rerun my basic regressions. Table 17 shows the estimated treatment effects with and without correcting for data attrition. The IPW estimates are basically the same as those without weighting. For example, with the IPW correction, the water improvement program increases the weight-for-height ratio of adults in the first wave by 1.208 kg/m with the OLS strategy and 2.860 kg/m with the IV strategy. These results are only slightly less than those from my basic specifications (1.132 and 2.891 kg/m) for the same adults. In conclusion, sample attrition does not appear to cause much bias since there is little observed sorting of migrants on whether villages have access to plant water.

3.2.6 Two Placebo Tests

Placebo Test 1: Treatment effects on the incidences of water-related and other kinds of diseases

If the water improvement program in rural China does benefit people's health and reduce illness incidence, then the treatment variable should affect only water-related diseases. In the CHNS data, diagnoses of illness in last four weeks were recorded if patients visited a medical facility. Based on the suggestions of Rachel Rosenberg,[14] an expert on the toxicology of drinking water, and on my knowledge, I divide the diagnoses into two categories:

---

[14] Personal communication with Rachel Rosenberg from School of Public Health, University of Maryland at College Park, February 2nd, 2010.

Water-related diseases include: infectious/parasitic disease, tumor, respiratory disease, endocrine disorder, hematology/blood disease, mental retardation, neurological disorder, eye/ear/nose/throat/teeth disease, digestive disease, urinary disease, neonatal disease, dermatological disease, and hereditary disease.

Other kinds of diseases include those less likely to be caused by poor drinking water quality: heart disease, injury, alcohol poisoning, mental/psychiatric disease, sexual disorder, muscular/rheumatological disease, and old-age/mid-life syndrome.

I exclude from classification two categories—obstetrical/gynecological disease and other—both of which are non-specific enough to make it impossible to classify them as being related (or not) to water quality. The average adults' illness incidences within the two categories of diseases are 3.70 and 1.32 percent, respectively. It is expected that the coefficients of the treatment variable should be significant when using the incidence of water-related diseases as an outcome variable, but not when employing the other group, which is less likely to be caused by the poor quality of drinking water. The regression results in Table 18 are consistent with this. Plant water reduces the incidence of water-related diseases for adults by 0.5 percentage point in the OLS specification (although this estimate is not statistically significant) and 2.9 percentage points in the IV specification, respectively. These results imply a 13.5 to 78.4 percent reduction in water-related diseases when a water improvement program is launched in a village. Meanwhile, the placebo test reveals that plant water has no significant impact on other kinds of diseases; the point estimate for both the OLS and IV specifications are small (0.002).

Placebo Test 2: Treatment Effects on Adult Height

Another interesting falsification check is to use adults' height as an outcome variable since plant water should not change young adults' height. If there are other unobserved factors that affect people's health, then the impacts of these omitted variables could lead to a spuriously positive and statistically significant estimated effect of the coefficient on **water plant**. Here, I exclude adults who benefitted from water improvement when they were children and those over 50 because the elderly's height may change with their health status. The sample size decreases to around 18,000. The regression results are presented in Table 19. The results are not conclusive. The OLS estimate is 0.363 cm and the IV estimate is a very large 9.441 cm, but both are insignificant.

3.3 Heterogeneous Treatment Effects

The previous estimates present only the average treatment effects on adult health. However, heterogeneous treatment effects are of interest since their impacts may vary with beneficiaries' socioeconomic characteristics, both because their knowledge can influence whether usage of water is effective and because their health endowment may affect the marginal gains from the program.

In this section, I explore heterogeneous treatment effects across income and education groups. The sample is divided into three groups (poor, middle and rich) based on the village's average income in the first survey year. The reason for using the average income at the village level is to avoid the endogeneity between households' income and their members' health status. Panel A of Table 20 shows the OLS and IV regression results across these three income groups. The estimates are sometimes imprecise, but are qualitatively similar. For example, when employing the IV strategy, the coefficient on adults' weight-for-height is zero for the middle income group; however, the coefficients for low and high income groups are similar (2.489 and 3.498 kg/m) and both significant, and they are also consistent with the baseline IV estimate (2.580 g/cm). Meanwhile, the OLS estimates for the same outcome are 0.779 kg/m for the low income group, 0.391 kg/m for the middle income group and 0.974 kg/m for the rich group. All of these OLS estimates are significant. From the above table, we can see that there is no clear evidence showing the existence of heterogeneous treatment effects across different income groups.

Another interesting hypothesis is that education is a complement to usage of safe drinking water. As in the case study by Ahmed et al. (1998) in Bangladesh, safe drinking water can be contaminated if households still use untreated surface water to wash containers. Better-educated individuals may be more aware of the importance of drinking water quality and have better hygiene practices, so their water is less likely to be contaminated. To test this hypothesis, I place adults into four education groups: Illiterate (years of education=0), Primary school (0<years of education<=6), Lower middle school (6<years of education<=9), and Upper middle school and above (years of education>9). Panel B of Table 20 shows the treatment effects on adult health across their own educational groups. The results indicate that access to plant water does lead to differential benefits in different measures across these educational groups, but these estimates do not vary much in terms of magnitudes and are not very different from the average treatment effects in the baseline regressions. For example, the OLS estimates for weight-for-height vary from 0.568 to 1.454 kg/m, while the IV estimates vary from 1.437 to 3.434 kg/m.

In summary, there is little evidence of heterogeneous treatment effects across income and educational groups. This may not be surprising considering that chemical impurities are the major problem with drinking water in China, and that they are hard to eliminate through human hygiene behavior such as boiling, the point-of-use water treatment generally employed in Chinese daily life.[15]

---

[15] According to Zhang et al. (2009), only 5.11 percent of rural households conduct other treatments.

Chapter 4

Results for Children

4.1 OLS and IV Regression Results

Table 21 presents OLS regression results for children.[16] OLS regressions with the county fixed effects indicate positive and significant impacts of the water program on children's weight-for-height and height. They predict gains for children of 0.446 kg/m in weight-for-height and 0.962 cm in height. In terms of other covariates, older children are healthier, and boys are heavier and taller than girls. Higher educational levels of fathers and mothers also significantly benefit children's health status. Larger household size worsens children's health status, presumably because children in large households obtain fewer household resources. Raising livestock has a negative and significant influence on children's anthropometric measures. Other controls—income and distance to a medical facility—do not show any statistically significant impact on child health.

Similar to what we see in adult health, the estimates of the impacts of plant water become smaller and insignificant when using the village fixed effects. The bootstrap Hausman tests are performed and their p-values are all above 0.2, which provides evidence that the OLS and IV estimates are not statistically different. In Table 22 I report results where I restrict the sample to the villages whose water treatment status

---

[16] In this study, children are defined as individuals aged from zero to 17. The regression results do not qualitatively change when I exclude infants aged from zero to two from the child sample.

changed during the survey period from 1989 to 2006 in order to explore the treatment effects on child health in those villages. As a result, around 3,000 children are included. None of the estimated coefficients are significant, which might be due to the limited sample size.

Next, I use same the instrumental variables strategy as I used for adults in order to deal with the potential endogenous placement of the water treatment plants—the instrument is the indicator for whether the village is in a non-flat area. The first stage regression results of program assignment for the child sample are presented in Table 23, and not surprisingly, the coefficient on the instrument (-0.431) is very similar to that in the first stage for the adult sample. The estimates of the relationship between other covariates and the treatment variable are similar to those in Table 3. For example, children in the treated villages tend to have better-educated parents, to live in a family that is less likely to raise livestock and to live closer to a medical facility. The F-statistic on the instrument is 17.36 and, therefore, finite sample bias resulting from a weak instrument employed in an IV regression is not a concern. Table 24 presents the results of OLS and IV strategies and for different outcome variables of children. All standard errors are clustered at the village level. The IV estimates for children’s anthropometric measures are almost twice as large as the OLS estimates: 0.754 kg/m in weight-for-height and 2.489 cm in height, while the OLS estimates are 0.446 kg/m in weight-for-height ratio and 0.962 cm in height. While both are statistically significant at the one percent level, the treatment effect on children’s illness incidence remains insignificant. The estimated coefficients on other covariates in the IV specifications are comparable to the ones in the OLS specifications. Considering, as discussed in the adult analysis, the potential biases generated by the OLS and IV estimation, I take the results as the lower and upper bounds of the true treatment effects of the drinking water improvement program.

To summarize, the predicted gains of plant water on children are 0.446 to 0.754 kg/m in weight-to-height and 0.962 to 2.489 cm in height. In addition, the children’s estimated health benefit in weight is 0.932 to 1.400 kg. Kanani and Poojara (2000) present epidemiological evidence that adolescent Indian girls aged 10 to 18 gain 0.83 kg with three months of iron and folic acid supplementation. In the meta-analysis by Brown et al. (2002), a set of studies shows that the height gain of children under 10 years old varies from -0.26 to 1.70 cm after treatment with zinc supplements for 6.8 months, on average. Habicht et al. (1995) find that in Guatemala, three-year zinc treatment had a cumulative effect of up to 2.5 cm on the height of children under the age of three. Thus, we can see that the estimated impacts of the water improvement program in this dissertation are in line with studies in the areas of nutrition and public health.

In Table 25, I include the age-square and income-square as control variables to see if they are crucial determinants in this context. The estimates are very similar to the baseline ones, as we see in the results for adults. For example, the OLS coefficient for children’s height with the nonlinear terms as controls is 1.043 cm, and the IV coefficient is 2.241 cm, similar to the results without these controls (0.962 and 2.489 cm). Therefore, it can be concluded that the exclusion of nonlinear terms of the age and income variables does not affect the estimation of the treatment effects of plant water.

The OLS and IV estimates both support the positive impacts of the water infrastructure program on child health, and the estimation from the IV specifications demonstrates a stronger influence than from OLS specifications. In the following sections in this chapter, I follow the structure of Chapter 3, checking the robustness of baseline estimates and exploring heterogeneous treatment effects across different demographic groups for children.

4.2 Robustness Check

4.2.1 Sensitivity analysis with regard to the definition of the treatment variable

First, I compare the regression results with different levels of definitions of treatment variables—the household level and the village level (**water plant**), in order to see whether the defined treatment variable reflects the actual program implementation because IV can correct for bias caused by omitted variables and measurement errors. In fact, 14.8 percent of households in the child sample have different values of those two variables. Table 26 presents the coefficients of the treatment variables at the household and village levels. We see that the regression results are very comparable when using different levels of treatment variables, regardless of whether they are derived from the OLS or the IV estimation.

For example, when children’s weight-for-height is the outcome variable, the OLS estimates of the household-level and the village-level treatment variables are 0.465 and 0.446 kg/m, respectively, and the IV estimates of these two variables are 0.749 and 0.754 kg/m. For children’s height, the OLS estimates are 0.875 cm at the household level and 0.962 cm at the village level, while the IV estimates are 2.708 and 2.489 cm, respectively. Thus, we can conclude that the definition of the treatment variable is reasonable in terms of reflecting reality.

Table 27 shows regression results with different percentage increases used to detect the timing of program implementation. The cutoffs and the treatment variables defined by those cutoffs are the same as the ones used in the analysis for adults. I do observe that the impacts estimated in the IV specifications become slightly stronger when a stricter criterion is applied. For example, the IV estimates on children's weight-for-height ratio and height change from 0.638 to 0.874 kg/m and from 2.108 to 2.878 cm, respectively. Moreover, the OLS estimates do not show such a pattern. For example, the coefficient of the treatment variable defined by a 10 percent cutoff for children's weight-for-height is 0.485 kg/m, while one defined by a 30 percent cutoff is 0.407 kg/m. In terms of significance, the estimates do not change across these numbers, except for children's height with both the OLS and IV specifications. Nevertheless, we can conclude that the benefits of the drinking water infrastructure program to child health are also relatively stable and robust across different definitions of the treatment variable.

4.2.2 Omitted Variable Bias—Other Infrastructure Construction

In this section, I address a similar consideration that some possible omitted variables may exist in the baseline regression. How the sanitation environment could affect the way that drinking water improvement works is especially important when we consider child health since children are more vulnerable to diarrheal diseases or diseases caused by microorganisms. In Table 28, I control for households' toilet types and sanitation environment (excreta around households' dwellings as evaluated by interviewers) in the regressions. Overall, both the OLS and IV estimates for children's anthropometric indicators fall to almost half of the baseline estimates. And they are all short of statistical significance except for the OLS estimate on weight-for-height ratio. The health gain decreases from 0.446 to 0.286 kg/m in weight-for-height and from 0.962 to 0.500 cm in height in the OLS regressions, and from 0.754 to 0.458 kg/m and from 2.489 to 1.771 cm, respectively, in the IV regressions.

As discussed in Section 3.2.3, the weakness of the estimates when controlling for sanitation variables does not necessarily imply that the baseline estimates of the impacts of the water improvement program are overstated. Sanitation improvement, whether due to a government program or to households' own decisions, may rely on the availability of plant water or piped water. Therefore, the change in the estimates here might be due to the regressions "over-controlling" for the impact evaluation of plant water through the inclusion of the sanitation variables. Furthermore, when the sanitation variables are included as controls, the amounts of the reductions in the magnitudes of estimated treatment effects imply the extent to which microorganisms influence health through drinking water. Not surprisingly, the changes in coefficients of plant water are relatively larger for child health than for adult health, which suggests that children are more vulnerable to bacteria and germs in drinking water.

In Table 29, I add the information of some other infrastructure construction to see if the baseline estimates for child health are robust. Here new control variables include roads (dirt, stone or paved), schools (distances to closest primary and middle schools), accessibility to trade areas, and telephone and electricity availability. The OLS estimates (0.427 kg/m for weight-for-height and 0.759 cm for height) are similar to the baseline estimates (0.446 kg/m for weight-for-height and 0.962 cm for height). The IV estimate for children's weight-for-height stays at the same magnitude (0.761 kg/m) and significant at the 10 percent level, while the estimate for children's height drops a little, from 2.489 to 2.178 cm, and is not significant (its p-value is still close to 0.1) now. Therefore, I can conclude that other infrastructure conditions only show secondary influences on health in this context and omitting them does not change my estimate substantially.

4.2.3 Channels Clarified—Safe Drinking Water versus Water Accessibility (Quality versus Quantity)

In order to explore the channel through which this program benefits child health—water quality or water quantity, I again run regressions, including water access as a control variable, and the results are presented in Panel A of Table 30. Here the variable—water access—is constructed in the same way as in the adult analysis. We can see that the estimated effects of water improvement generated by the program stay very similar to the baseline estimates: 0.457 versus 0.446 kg/m in weight-for-height and 0.943 versus 0.962 cm in height from the OLS estimation; and 0.849 versus 0.754 kg/m in weight-for-height and 2.589 versus 2.489 cm in height from the IV estimation. And they are also significant, with the exception of the IV estimate on child height, while water access does not generate any statistically significant influences. In Panel B, we restrict the sample to children whose households always have optimal water access. For them, the health gain from plant water becomes slightly greater and remains significant. For instance, the OLS coefficient of the treatment for children's weight-for-height increases from 0.446 to 0.508 kg/m, and the coefficient for height rises from 0.962 to 1.443 cm.

This change may reflect only the treatment effect for this sample, but the similarity between these estimates and the baseline ones again suggests that the benefits of this program come from the improvement of water quality. In Panel C of Table 30, I also focus on the children under age 10 in the first three waves that Mangyo (2008) studies, and consider how water improvement affects their weight, height and BMI. The results suggest that children gained 0.560kg in weight and 1.380 cm in height in the OLS regressions and 1.825 kg and 2.214 cm in the IV regressions, after plant water is accessible (although only the OLS results are statistically significant). These estimated benefits are much more pronounced than those of water access. For children who always have optimal water access, the health gain from having access to water from a treatment plant is almost double as shown in Panel D of Table 30.

Next, I attempt to replicate Mangyo's (2008) study in which children are less than 10 years old and show up in all of the first three waves. The sample size I obtain is slightly different from what he uses: 904, 1007, and 708 children for height, weight, and BMI as outcome variables, respectively, as compared to 1094, 1192 and 816 in his sample. The regressions results for this replication sample are presented in Table 31. For the very limited sample, plant water generates effects that are comparable to the full sample results, while water access does not. For example, the OLS estimates are 0.466 kg for weight and 1.622 cm for height, and the IV estimates are 1.937 kg and 3.426 cm. However, when restricting the sample to observations with optimal water access in all three waves, the sample size shrinks by half, and the almost all of the estimates are doubled in magnitude and statistically significant.

4.3 Heterogeneous Treatment Effects

In this section, I explore heterogeneous treatment effects on children across income and parents' educational groups. The sample is divided into three groups (poor, middle and rich) according to the village's average income in the first survey year. Table 32 presents the estimates across these income groups in Panel A. The estimates are sometimes imprecise. For example, the IV coefficient on illness incidence for the low income group is positive (0.049) and statistically significant, while the coefficients for the middle and high income groups are negative and insignificant. When using anthropometric measures (weight-for-height and height) as the outcome variables, the program has a significant impact only on poor children. Children in low income villages shows, on average, a 0.739 kg/m gain in weight-for-height and a 2.723 cm increase in height from the OLS estimation and 1.284 kg/m and 3.251 cm from the IV estimation after the water infrastructure program covers these villages.

Next, I test whether the impact of this program on child health is a function of their parents' educational attainment. Panels B and C of Table 32 present the treatment effects across mothers' and fathers' education groups, respectively. We can see that the estimates stay stable across mothers' educational groups, although they are sometimes imprecise. The estimated impact on children's weight-for-height is quite similar: around 0.4 kg/m for OLS estimates and around one kg/m for IV estimates when mothers' highest education varies from illiterate (0.424 and 0.970 kg/m), to primary school (0.438 and 0.954 kg/m), to lower middle school (0.483 and 0.878 kg/m) to upper middle school or above (0.486 and 1.198 kg/m). However, if we look at the same estimates across fathers' educational groups, they vary across a larger range. This can be explained by the findings from many other empirical studies that mothers’ education plays a more essential role than fathers’ in child growth. However, this may be also due to the fact that for one third of children in my sample their fathers are missing and that I use the education of the oldest male in their households as a substitute.

Chapter 5

Conclusion

Providing people with safe drinking water is one of the most important health-related infrastructure programs in the world. The most prevalent water pollutants in the world—microorganisms—can be partially eliminated by the Chinese tradition of drinking boiled water and eating cooked food. As a result, chemical impurities likely are the main threat to drinking water quality in China. Such impurities are a result of geography—high concentration of chemical elements in natural soil and rocks—and human activities due to vast disposal of industrial waste and excess usage of fertilizers during rapid industrialization.

Since the 1980s, the Chinese government has implemented a water improvement program in rural areas, constructing water plants and pipeline systems to provide people with safe drinking water. Those water plants install equipment and employ clean water technology to eliminate contaminants in drinking water, and the pipeline systems deliver treated water to households directly. It has been almost thirty years since the government launched the drinking water improvement program, which now covers around half of China's rural population. The impact of this program on people's health has important policy implications.

My dissertation uses the CHNS data to estimate the impact of the drinking water improvement program in rural China on the health of adults and children. Here two estimation strategies are employed: Ordinary Least Squares strategy with fixed effects and Instrumental Variables. Given that the program may have been launched first in areas having unsafe drinking water, the impacts estimated from the OLS specifications may be under-estimated. Moreover, the IV estimates when using villages' topography as the instruments for the possible endogenous program placement may imply the upper bounds of the treatment effect. This may be due to upward bias caused by other unobservable conditions that are better in villages in flat areas. Thus, in this study combining these two sets of estimates can help us identify the range where the treatment effects are located.

The estimated effects of plant water are that the illness incidence of adults decreases by 10 to 50 percent, and that their weight-for-height increases by 0.835 to 2.580 kg/m. Adults also self-evaluate their own health to be better when they have access to treated plant water. Children's weight-for-height and height rise by 0.446 to 0.754 kg/m and 0.962 to 2.489 cm, respectively, after the program is launched. These health gains for adults and children are consistent with studies in the areas of nutrition and public health. Given the fact that some villages have only recently been covered, the long-term benefits to health might be even greater.

I show that the estimated impacts are fairly robust and are not driven by measurement errors, omitted variable bias from obvious candidates, or attrition bias. The OLS and IV estimates are not sensitive to the definitions of the treatment variable, regardless of whether it is defined at the household or the village level and what cutoffs are used. Inclusion of sanitation reduces the program impacts but sanitation may itself be endogenous to water treatment plan access. Adding villages' other infrastructure conditions as controls barely changes the coefficients of plant water, implying that these conditions do not have first-order influences on health.

I also confirm that sample attrition does not sort on the treatment, and the inverse probability weighted estimates are quite similar to the results from the unweighted specifications. Finally, the mechanism through which the program takes effect is via improved water quality rather than simply via increased access to water. Placebo tests show that plant water decreases the illness incidence of water-related diseases among adults, but not of other diseases that are less likely to be caused by unsafe drinking water. Furthermore, this water program does not generate any statistically significant effects on adults' height, supporting the validity of the estimation strategies employed in this study. The heterogeneous treatment effects across income and educational groups are sometimes imprecise, but are qualitatively similar. This is consistent with the fact that the main threats to drinking water quality in China are chemical impurities.

My results clearly indicate that the construction and implementation of water plants in rural China has resulted in short-term health benefits for adults and children. To the extent that these water treatment plants are costly to construct, and to the extent that we are still only able to see short-run benefits on health, a full analysis of the health benefits awaits future research.

Figure 1 Coverage of Water Plant versus Water Access from 1989 to 2006 (CHNS)

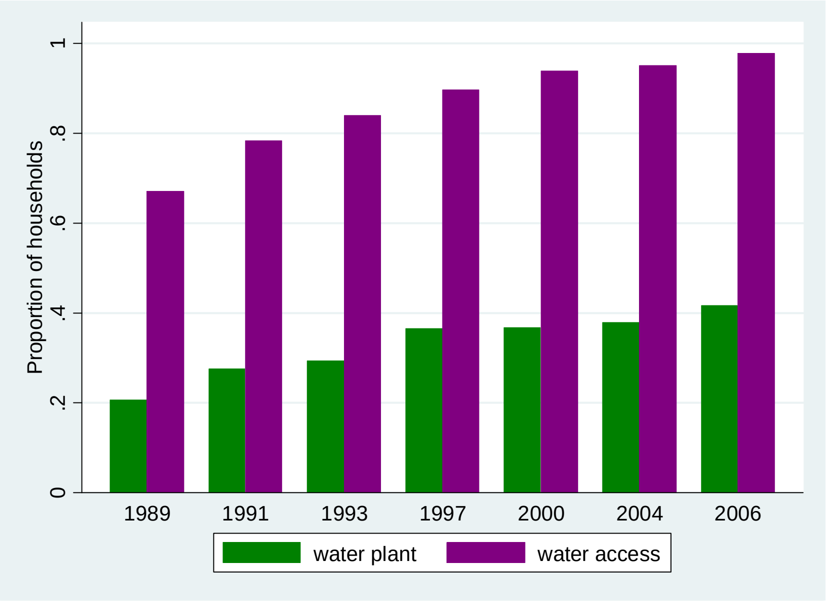
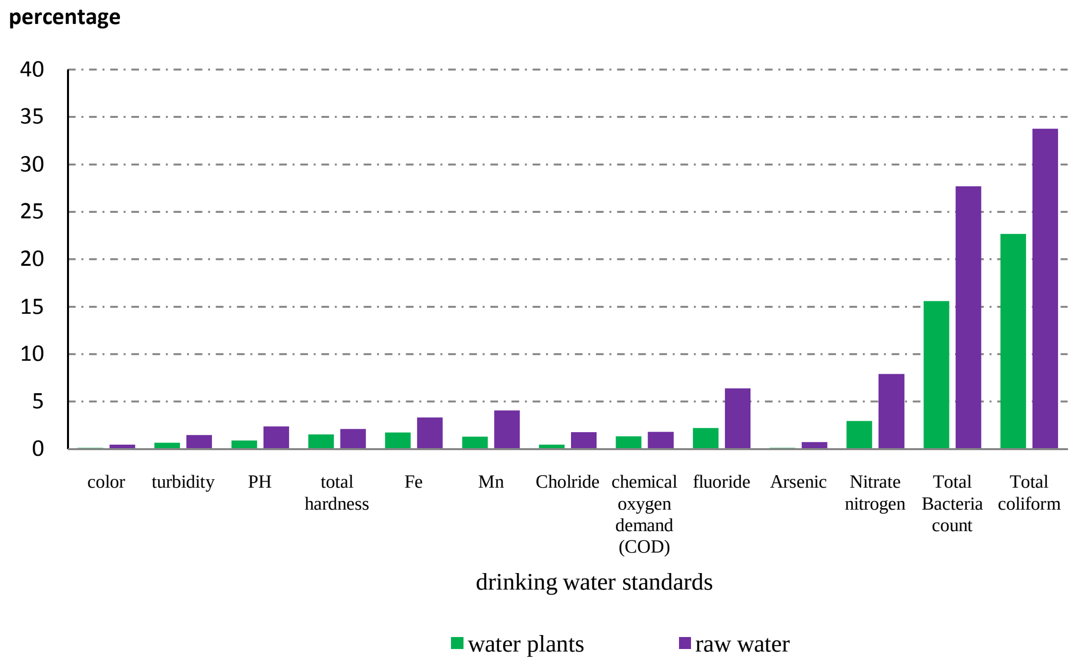


Figure 2 Proportions of Households Violating the Standards of Drinking Water in China in 2006



Source: Zhang et.al., “Current Situation Analysis on China Rural Drinking Water Quality”, *Journal of Environment and Health*, Jan 2009, 26(1)

Figure 3 Map of Population Drinking High Fluoride Water in China in 2006

Source: Chinese National Health Statistics, 2007

Figure 4 Fraction of Households reporting plants as water source in a village

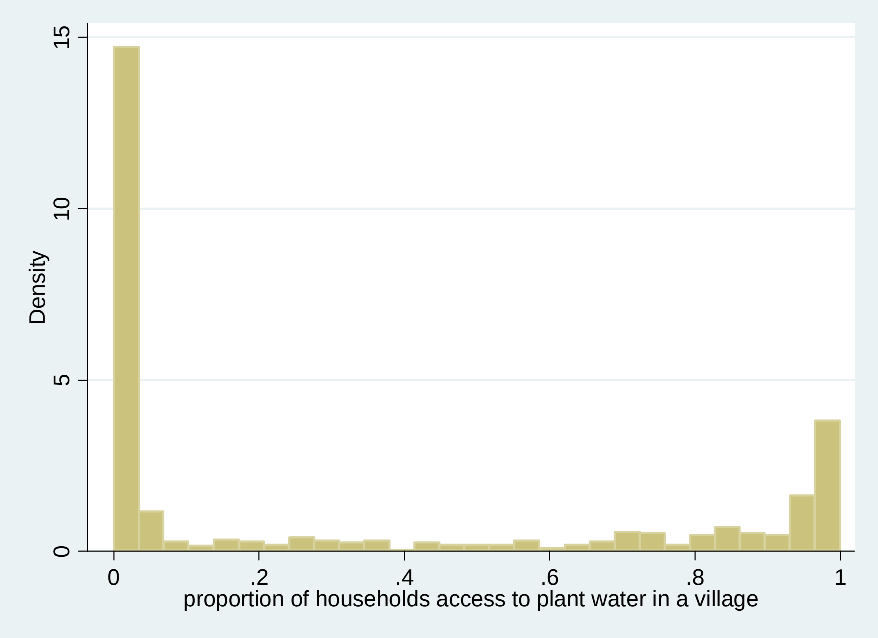


Figure 5 Coefficients of the Treatment’s Leads and Lags

Notes: The estimated coefficients of the treatment’s leads and lags are drawn in solid lines and their 95 percent confidence intervals are in dash lines. If time distance is greater than 5 years, I code the leads and lags as -5 or 5 depending on it’s before or after program implementation.

Figure 6 Kernel Densities of the Treatment Variables with Different Cutoffs

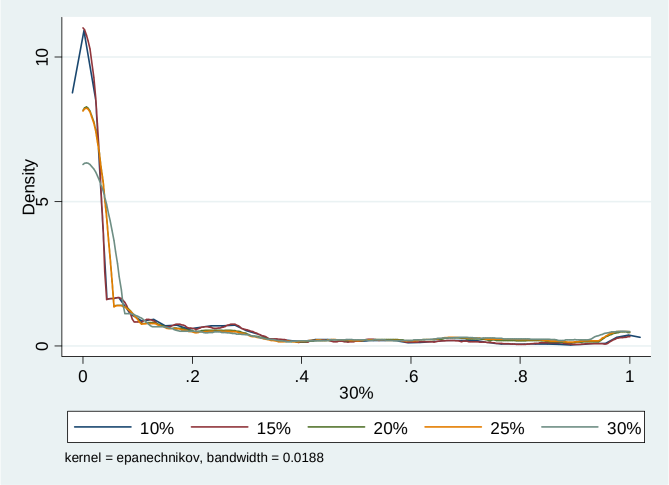


Figure 7 Age Distribution of Missing Observations (CHNS)

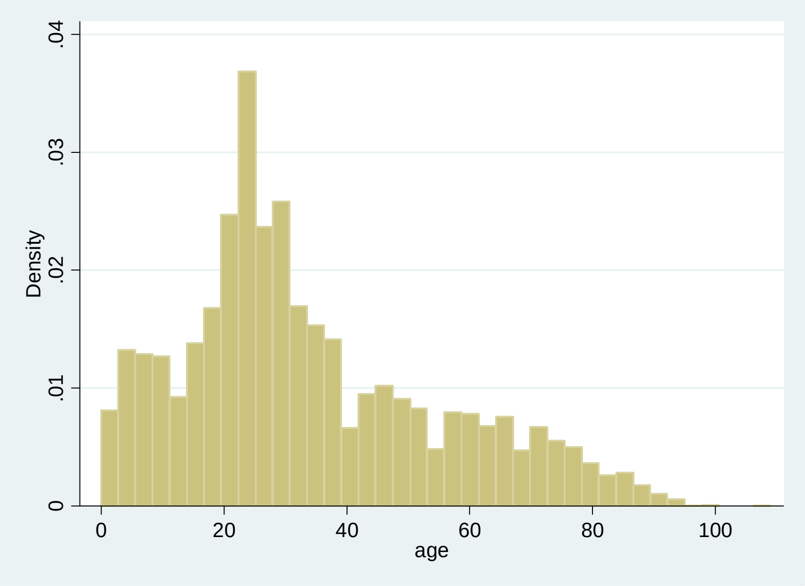


Table 1 the Cost of the Construction of Water and Sewage Systems

	water pipeline system	deep well pump	rainwater harvesting system	household sewage setup	public and school sewage setup
Cost	<$30 per capita	$5-$10 per capita	$50-$80 per capita	$90-$120 per setup	$500-$850 per sitting

Source: Meng et al (2004) “water supply and sanitation environment in rural China: promote service to the poor,” Poverty Reduction Conference in Shanghai, China (2004).

Table 2 Descriptive Statistics of Adults

Variables	All (1)	Treatment (2)	Control (3)	Δ_{T-C} (4)
Illness in last four weeks	0.090	0.088	0.091	-0.003
	(0.286)	(0.284)	(0.287)	(0.008)
Self-reported health status	1.598	1.616	1.589	0.027
(1=poor, 2=fair, 3=excellent)	(0.603)	(0.597)	(0.606)	(0.024)
Weight-for-height	35.485	35.810	35.347	0.463
(kg/m)	(5.640)	(5.791)	(5.568)	(0.385)
Age	41.627	43.186	40.968	2.218***
	(15.740)	(16.307)	(15.447)	(0.482)
Female	0.505	0.510	0.504	0.006
	(0.500)	(0.500)	(0.500)	(0.004)
Educational attainment	6.267	6.820	6.033	0.787***
(years)	(3.947)	(4.110)	(3.853)	(0.227)
Married	0.770	0.761	0.773	-0.012
	(0.421)	(0.427)	(0.419)	(0.013)
Household size	4.566	4.497	4.595	-0.098
	(1.688)	(1.754)	(1.658)	(0.134)
Log household annual income in	8.699	8.805	8.654	0.151**
first wave	(1.014)	(0.987)	(1.023)	(0.073)
Whether households raise	0.503	0.346	0.570	-0.224***
livestock	(0.500)	(0.476)	(0.495)	(0.057)
Distance to the nearest medical	0.364	0.292	0.394	-0.102
facility (km)	(0.930)	(0.529)	(1.053)	(0.079)
Observations	39517	11738	27779	

Notes: Column (1) displays sample means and standard deviations (in parentheses) for adult observations in all waves, and Column (2) and (3) by their treatment status. The mean differences between column (2) and (3) and their standard errors in parentheses (clustered at the village level) are shown in column (4). *** p<0.01, ** p<0.05, * p<0.1

Table 3 Descriptive Statistics of Children

Variables	All (1)	Treatment (2)	Control (3)	Δ_{T-C} (4)
Illness in last four weeks	0.066	0.057	0.069	-0.012
	(0.247)	(0.231)	(0.253)	(0.008)
Weight-for-height	21.429	21.785	21.301	0.484
(kg/m)	(7.067)	(7.107)	-7.048	(0.379)
Height (cm)	124.338	126.466	123.570	2.896**
	(27.767)	(27.876)	-27.688	(1.272)
Age	8.879	9.150	8.785	0.365**
	(5.020)	(4.971)	-5.033	(0.166)
Female	0.466	0.472	0.464	0.008
	(0.499)	(0.499)	-0.499	(0.017)
Father’s education	7.365	7.689	7.253	0.436*
(years)	(3.405)	(3.544)	-3.349	(0.234)
Mother’s education	5.755	6.447	5.514	0.933**
(years)	(3.890)	(3.997)	-3.824	(0.423)
Household size	4.855	4.817	4.868	-0.051
	(1.511)	(1.590)	-1.482	(0.142)
Log household annual income	8.660	8.785	8.617	0.168*
in first year	(1.010)	(1.003)	-1.009	(0.087)
Whether households raise	0.543	0.406	0.591	-0.185***
livestock	(0.498)	(0.491)	-0.492	(0.063)
Distance to the nearest medical	0.372	0.329	0.386	-0.057
facility (km)	(0.927)	(0.556)	-1.024	(0.091)
Observations	14494	3737	10757	

Notes: Column (1) displays sample means and standard deviations (in parentheses) for child observations in all waves, and Column (2) and (3) by their treatment status. The mean differences between column (2) and (3) and their standard errors in parentheses (clustered at the village level) are shown in column (4). *** p<0.01, ** p<0.05, * p<0.1

Table 4 Descriptive Statistics of Infrastructure Variables

Variables	Little excreta	Some excreta	Much excreta	No bathroom	Flush toilet
Adult	0.293	0.181	0.013	0.017	0.176
Sample	(0.444)	(0.385)	(0.112)	(0.128)	(0.381)
Child	0.298	0.213	0.013	0.017	0.142
Sample	(0.458)	(0.410)	(0.114)	(0.130)	(0.349)

	Non-flush toilet	Open pit	Dirt road	Stone road	Distance to the nearest primary school (km)
Adult	0.160	0.625	0.263	0.292	0.291
Sample	(0.367)	(0.484)	(0.440)	(0.455)	(0.855)
Child	0.145	0.669	0.313	0.291	0.281
Sample	(0.352)	(0.471)	(0.464)	(0.454)	(0.884)

	Distance to the nearest middle school (km)	Trade area near the village (yes/no)	Telephone availability in the village (yes/no)	Electricity availability in the village (yes/no)	
Adult	1.831	0.264	0.719	0.983	
Sample	(4.673)	(0.441)	(0.450)	(0.131)	
Child	1.834	0.237	0.682	0.981	
Sample	(4.262)	(0.425)	(0.466)	(0.138)	

Notes: The table displays sample means of the variables referring to households' infrastructure conditions for adult and child sample, respectively. The standard deviations are in parentheses. Households' sanitation environment evaluated by the interviewers is divided into four categories: no excreta (omitted), little excreta, some excreta and much excreta. Five types of households' toilet facilities are no bathroom, flush toilet (in- and outside house), non-flush toilet (in- and outside house), open pit (cement and earth) and other (omitted). Road conditions around the villages are described by three categories: dirt, stone and paved (omitted).

Table 5 Mean Differences between Characteristics of Treated and Untreated Villages

	Log household annual income (1)	Adults' educational attainment (2)	Proportion of female (3)	Distance to nearest medical facility (km) (4)	Paved road (5)	Distance to a middle school (km) (6)
Mean differences	0.042 (0.098)	0.060 (0.069)	-0.003 (0.007)	-0.362* (0.186)	0.054 (0.084)	-0.709 (0.880)

Notes: the means of the treated villages are the average of their characteristics in five years before the treatment. The mean differences are adjusted for year fixed-effects and the standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6 Treatment Effects on Adults' Health Status

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	-0.010*	-0.020	0.027	0.030	0.835***	0.279
	(0.006)	(0.012)	(0.016)	(0.067)	(0.174)	(0.199)
Age	0.003***	0.003***	-0.011***	-0.010***	-0.003	-0.013***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	(0.004)
Female	0.010***	0.009**	-0.063***	-0.062***	-1.780***	-1.906***
	(0.004)	(0.004)	(0.009)	(0.009)	(0.114)	(0.115)
Educational attainment (years)	-0.001**	-0.002**	0.014***	0.014***	0.087***	0.033**
	(0.001)	(0.001)	(0.002)	(0.002)	(0.016)	(0.015)
Married	-0.005	-0.004	0.057***	0.058***	1.437***	1.524***
	(0.004)	(0.004)	(0.012)	(0.012)	(0.122)	(0.117)
Household size	-0.005***	-0.005***	0.009**	0.009**	-0.059**	-0.057**
	(0.001)	(0.001)	(0.004)	(0.004)	(0.029)	(0.028)
Log income in first year	-0.005**	-0.005**	0.017***	0.015***	0.215***	0.210***
	(0.002)	(0.002)	(0.005)	(0.005)	(0.058)	(0.056)
Livestock	0.004	0.011**	-0.010	-0.012	-0.854***	-0.186**
	(0.005)	(0.004)	(0.012)	(0.013)	(0.109)	(0.080)
Kms to the nearest medical facility	-0.000	0.001	-0.001	-0.011	-0.018	0.000
	(0.004)	(0.004)	(0.011)	(0.012)	(0.064)	(0.048)
Constant	0.088***	0.150***	1.708***	1.664***	34.988***	37.749***
	(0.026)	(0.026)	(0.065)	(0.092)	(0.704)	(0.626)
County fixed effect	Yes	No	Yes	No	Yes	No
Village fixed effect	No	Yes	No	Yes	No	Yes
Observations	39,278	39,278	21,308	21,308	33,116	33,116
R-squared	0.059	0.069	0.178	0.194	0.203	0.237
P value (bootstrap Hausman test)	0.390		0.963		0.024	

Notes: each column lists coefficient estimates with standard errors in parentheses (clustered at the village level) from separate regressions of a health outcome. In addition to the covariates listed above, each regression also controls for year fixed-effects. The bootstrap Hausman tests are based on 1000 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1

Table 7 Treatment Effects on Adults' Health Status in the Restricted Sample

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	-0.021*	-0.027	0.059	0.035	0.727***	0.404*
	(0.010)	(0.017)	(0.037)	(0.075)	(0.240)	(0.213)
Constant	0.145***	0.170***	1.731***	1.707***	35.032***	35.074***
	(0.043)	(0.044)	(0.134)	(0.146)	(1.443)	(1.384)
County fixed effect	Yes	No	Yes	No	Yes	No
Village fixed effect	No	Yes	No	Yes	No	Yes
Observations	9,248	9,248	4,632	4,632	7,782	7,782
R-squared	0.057	0.062	0.183	0.192	0.233	0.244

Notes: the other covariates controlled for in each regression are the same as ones in Table 6. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 8 Assignment to Treatment for Adult Sample—OLS (First Stage)

Treatment	Water plant (1)
Non-flat	-0.419***
	(0.101)
Age	0.003***
	(0.001)
Female	0.025***
	(0.007)
Educational attainment (years)	0.012***
	(0.003)
Married	-0.021**
	(0.009)
Household size	0.001
	(0.004)
Log income in first year	0.005
	(0.009)
Livestock	-0.182***
	(0.038)
Kms to the nearest medical facility	-0.043***
	(0.016)
Constant	0.449
	(0.283)
Observations	35,752
R-squared	0.380
F-stat on instruments	17.29
Prob>F	0.0001

Notes: the regression also controls for county and year fixed-effects. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 9 Treatment Effects of Water Program on Adults' Health Status

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	-0.010*	-0.045**	0.027	0.144**	0.835***	2.580***
	(0.006)	(0.022)	(0.016)	(0.065)	(0.174)	(0.756)
Age	0.003***	0.003***	-0.011***	-0.011***	-0.003	-0.012**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.005)	(0.006)
Female	0.010***	0.010**	-0.063***	-0.067***	-1.780***	-1.762***
	(0.004)	(0.004)	(0.009)	(0.011)	(0.114)	(0.121)
Educational attainment (years)	-0.001**	-0.001	0.014***	0.012***	0.087***	0.057***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.016)	(0.021)
Married	-0.005	-0.007	0.057***	0.060***	1.437***	1.435***
	(0.004)	(0.004)	(0.012)	(0.013)	(0.122)	(0.132)
Household size	-0.005***	-0.005***	0.009**	0.008*	-0.059**	-0.063**
	(0.001)	(0.001)	(0.004)	(0.004)	(0.029)	(0.031)
Log income in first year	-0.005**	-0.005**	0.017***	0.018***	0.215***	0.179***
	(0.002)	(0.002)	(0.005)	(0.005)	(0.058)	(0.062)
Livestock	0.004	-0.006	-0.010	0.014	-0.854***	-0.414**
	(0.005)	(0.007)	(0.012)	(0.018)	(0.109)	(0.204)
Kms to the nearest medical facility	-0.000	-0.001	-0.001	-0.003	-0.018	0.054
	(0.004)	(0.004)	(0.011)	(0.014)	(0.064)	(0.086)
Constant	0.088***	0.096***	1.647***	1.667***	34.988***	35.177***
	(0.026)	(0.029)	(0.063)	(0.076)	(0.704)	(0.783)
Observations	39278	35538	21308	17890	33116	29763
R-squared	0.059	0.058	0.178	0.172	0.203	0.183

Notes: each regression also controls for county and year fixed-effects. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 10 Treatment Effects on Adult Health with and without Controlling for Nonlinear Terms of Age and Income

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	Without	With	Without	With	Without	With
	(1)	(2)	(3)	(4)	(5)	(6)
			OLS Estimates			
Water plant	-0.010*	-0.011*	0.027	0.028*	0.835***	0.864***
	(0.006)	(0.006)	(0.016)	(0.016)	(0.174)	(0.175)
			IV Estimates			
Water plant	-0.045**	-0.045**	0.144**	0.142**	2.580***	2.630***
	(0.022)	(0.022)	(0.065)	(0.064)	(0.756)	(0.768)

Notes: the nonlinear terms are age^2 and log household income in the first $year^2$. The other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 11 Regression Results with Treatment Variables at Different Levels

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Household-level Treatment	-0.011**	-0.047**	0.029*	0.131**	1.027***	2.683***
	(0.005)	(0.021)	(0.016)	(0.057)	(0.124)	(0.643)
Village-level Treatment (**water plant**)	-0.010*	-0.045**	0.027	0.144**	0.835***	2.580***
	(0.006)	(0.022)	(0.016)	(0.065)	(0.174)	(0.756)

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 12 Treatment Effects on Adult Health across Different Cutoffs

	OLS Estimates			IV Estimates		
Cutoffs	Illness in last four weeks (1)	Self-reported Health Status (2)	Weight-for-height (3)	Illness in last four weeks (4)	Self-reported Health Status (5)	Weight-for-height (5)
10%	-0.012**	0.010	1.087***	-0.036**	0.108**	2.077***
	(0.006)	(0.017)	(0.176)	(0.016)	(0.048)	(0.463)
15%	-0.006	0.011	1.035***	-0.041**	0.130**	2.415***
	(0.006)	(0.017)	(0.186)	(0.020)	(0.059)	(0.610)
20%	-0.010*	0.027	0.835***	-0.045**	0.144**	2.580***
(**water plant**)	(0.006)	(0.016)	(0.174)	(0.022)	(0.065)	(0.756)
25%	-0.013**	0.027*	0.751***	-0.054**	0.180**	3.107***
	(0.006)	(0.016)	(0.181)	(0.026)	(0.079)	(1.133)
30%	-0.009	0.021	0.772***	-0.054**	0.181**	3.111***
	(0.006)	(0.016)	(0.185)	(0.026)	(0.080)	(1.134)

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 13 Treatment Effects on Adult Health Controlling for Households’ Sanitation Facilities and Environment

		OLS Estimates			IV Estimates		
Variables		Illness in last four weeks	Self-reported Health Status	Weight-for-height	Illness in last four weeks	Self-reported Health Status	Weight-for-height
		(1)	(2)	(3)	(4)	(5)	(6)
Water plant		-0.008	0.023	0.467***	-0.042	0.160**	2.068***
		(0.006)	(0.017)	(0.153)	(0.027)	(0.081)	(0.759)
Sanitation Environment	Little excreta	0.006	-0.014	-0.397***	0.004	-0.005	-0.330***
		(0.004)	(0.015)	(0.092)	(0.005)	(0.017)	(0.112)
	Some excreta	0.021***	-0.042**	-0.360***	0.018***	-0.027	-0.245*
		(0.006)	(0.017)	(0.110)	(0.006)	(0.018)	(0.130)
	Much excreta	0.060***	-0.224***	-0.355	0.062***	-0.273***	-0.304
		(0.018)	(0.064)	(0.273)	(0.019)	(0.073)	(0.367)
Toilet type	No bathroom	-0.011	0.048	0.154	-0.005	0.009	0.080
		(0.015)	(0.070)	(0.584)	(0.015)	(0.072)	(0.659)
	Flush toilet	-0.013	0.105	0.885***	-0.004	0.053	0.331
		(0.013)	(0.066)	(0.318)	(0.014)	(0.067)	(0.407)
	Non-flush toilet	0.003	0.077	0.243	0.006	0.054	0.073
		(0.013)	(0.065)	(0.324)	(0.013)	(0.063)	(0.362)
	Open pit	-0.007	0.094	-0.520*	-0.007	0.083	-0.539*
		(0.012)	(0.063)	(0.301)	(0.012)	(0.060)	(0.312)
Constant		0.079***	1.573***	36.011***	0.090***	1.541***	35.960***
		(0.028)	(0.089)	(0.716)	(0.031)	(0.094)	(0.780)
	Observations	38,672	20,943	32,626	34,992	17,573	29,322
	R-squared	0.061	0.180	0.211	0.060	0.172	0.194

Notes: the omitted group of sanitation environment is “no excreta” and that of toilet types is “other”. The other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 14 Treatment Effects on Adult Health Controlling for Other Infrastructure

	OLS Estimates			IV Estimates		
	Illness in last four weeks	Self-reported Health Status	Weight-for-height	Illness in last four weeks	Self-reported Health Status	Weight-for-height
Control variables	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	-0.012*	0.033**	0.646***	-0.049*	0.176**	2.381***
	(0.006)	(0.017)	(0.173)	(0.025)	(0.076)	(0.820)
Dirt roads around villages	-0.002	0.011	-0.381***	-0.007	0.032	-0.173
	(0.005)	(0.022)	(0.135)	(0.006)	(0.029)	(0.187)
Stone roads around villages	-0.004	0.028	-0.511***	-0.008	0.047**	-0.229
	(0.005)	(0.017)	(0.101)	(0.006)	(0.023)	(0.166)
Kms to the nearest primary school	-0.002	0.003	-0.013	-0.002	0.005	-0.001
	(0.002)	(0.010)	(0.045)	(0.003)	(0.014)	(0.055)
Kms to the nearest middle schools	0.000	-0.001	-0.021*	-0.000	0.000	-0.009
	(0.000)	(0.001)	(0.011)	(0.000)	(0.001)	(0.009)
Trade areas nearby	-0.000	0.013	0.229**	-0.000	-0.002	0.129
	(0.007)	(0.017)	(0.109)	(0.008)	(0.024)	(0.150)
Telephone availability	0.002	-0.024	0.273**	0.005	-0.049*	0.031
	(0.006)	(0.022)	(0.126)	(0.007)	(0.026)	(0.160)
Electricity	-0.014	-0.018	0.266	-0.018	0.048	0.670*
	(0.014)	(0.032)	(0.368)	(0.014)	(0.058)	(0.361)
Constant	0.104***	1.673***	34.831***	0.117***	1.633***	34.595***
	(0.032)	(0.074)	(0.775)	(0.036)	(0.101)	(0.833)
Observations	39,193	21,266	33,039	35,495	17,890	29,728
R-squared	0.059	0.178	0.206	0.058	0.170	0.187

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 15 Treatment Effects Controlling for Water Accessibility

	PANEL A: Controlling for Water Access					
	OLS Estimates			IV Estimates		
	Illness in last four weeks (1)	Self-reported Health Status (2)	Weight-for-height (3)	Illness in last four weeks (4)	Self-reported Health Status (5)	Weight-for-height (6)
Water plant	-0.010	0.026	0.795***	-0.045*	0.136**	2.614***
	(0.006)	(0.016)	(0.177)	(0.025)	(0.066)	(0.821)
Water access	-0.010	0.033	0.425***	-0.004	0.018	0.023
	(0.007)	(0.022)	(0.155)	(0.009)	(0.027)	(0.237)
Observations	38,939	21,082	32,827	35,237	17,702	29,512
R-squared	0.060	0.178	0.203	0.058	0.173	0.182
	PANEL B: the Sample always with Optimal Water Access					
	OLS Estimates			IV Estimates		
	Illness in last four weeks (1)	Self-reported Health Status (2)	Weight-for-height (3)	Illness in last four weeks (4)	Self-reported Health Status (5)	Weight-for-height (6)
Water plant	-0.013*	0.019	0.996***	-0.036	0.168	3.154***
	(0.007)	(0.018)	(0.190)	(0.035)	(0.102)	(0.772)
Observations	23,944	13,881	20,395	20,900	11,061	17,660
R-squared	0.061	0.172	0.190	0.061	0.163	0.172

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 16 Regression Results of Adults' Probability of Leaving the Sample

	(1)	(2)	(3)
Water plant	0.017	0.017	0.018
	(0.021)	(0.021)	(0.021)
Age		-0.000	-0.005***
		(0.000)	(0.001)
Age^2			0.000***
			(0.000)
Female		0.003	0.002
		(0.002)	(0.002)
Years of education		-0.000	-0.000
		(0.001)	(0.001)
Married		-0.009**	0.011**
		(0.004)	(0.005)
Household size		-0.007***	-0.008***
		(0.002)	(0.002)
Log income in first year		-0.004	-0.017
		(0.003)	(0.025)
Log income in first $year^2$			0.001
			(0.001)
Livestock		-0.024***	-0.023***
		(0.008)	(0.008)
Constant	0.447***	0.530***	0.658***
	(0.019)	(0.032)	(0.109)
Observations	39,517	39,517	39,517
R-squared	0.162	0.164	0.166

Notes: in addition, each regression controls for village and year fixed effects. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 17 Treatment Effects on Adult Health with and without Correcting for Data Attrition

Dependent Variables	Illness in last four weeks		Self-reported health status		Weight-for-height	
	without	with	without	with	without	with
	(1)	(2)	(3)	(4)	(5)	(6)
			OLS Estimates			
Water plant	-0.010	-0.002	0.051**	0.051*	1.132***	1.208***
	(0.008)	(0.010)	(0.025)	(0.026)	(0.238)	(0.258)
Observations	19,960	19,960	10,575	10,575	17,387	17,387
R-squared	0.061	0.073	0.167	0.180	0.198	0.200
			IV Estimates			
Water plant	-0.035	-0.034	0.198**	0.223**	2.891***	2.860***
	(0.027)	(0.027)	(0.081)	(0.086)	(0.840)	(0.818)
Observations	19,587	19,587	10,352	10,352	17,077	17,077
R-squared	0.072	0.061	0.175	0.160	0.189	0.187

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 18 Treatment Effects on Adults' Illness Incidence

Dependent Variables	Water-related diseases		Other kinds of diseases	
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Water plant	-0.005	-0.029**	0.002	0.002
	(0.003)	(0.014)	(0.002)	(0.004)
Constant	0.043***	0.045**	0.001	0.001
	(0.016)	(0.018)	(0.009)	(0.009)
Observations	39,288	35,548	39,281	35,541
R-squared	0.022	0.020	0.018	0.016

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 19 Treatment Effects on Adults' Height

VARIABLES	Height OLS (1)	Height IV (2)
Water plant	0.363	9.441
	(0.273)	(5.914)
Constant	168.526***	168.858***
	(1.033)	(2.545)
Observations	18,500	16,700
R-squared	0.546	0.432

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 20 Heterogeneous Treatment Effects on Adult Health across Income and Educational Groups

PANEL A: Income Groups

		OLS Estimates			IV Estimates		
		Illness in last four weeks	Self-reported Health Status	Weight-for-height	Illness in last four weeks	Self-reported Health Status	Weight-for-height
Income Group		(1)	(2)	(3)	(4)	(5)	(6)
Poor	**Water plant**	-0.022**	0.006	0.779***	-0.013	0.152	2.489**
		(0.011)	(0.033)	(0.275)	(0.031)	(0.102)	(0.972)
Middle	**Water plant**	-0.013	0.048*	0.391*	-0.069**	0.220	-0.037
		(0.008)	(0.025)	(0.229)	(0.033)	(0.151)	(0.569)
Rich	**Water plant**	-0.017	0.037	0.974**	-0.070	0.233	3.498**
		(0.015)	(0.025)	(0.371)	(0.066)	(0.234)	(1.550)

PANEL B: Adults' Educational Groups

		OLS Estimates			IV Estimates		
		Illness in last four weeks	Self-reported Health Status	Weight-for-height	Illness in last four weeks	Self-reported Health Status	Weight-for-height
Educational Groups		(1)	(2)	(3)	(4)	(5)	(6)
Illiterate	**Water plant**	-0.008	0.036	1.454***	-0.074*	0.160	1.437
		(0.011)	(0.037)	(0.357)	(0.038)	(0.133)	(0.989)
Primary school	**Water plant**	-0.003	0.013	1.036***	-0.046	0.199**	3.045***
		(0.010)	(0.028)	(0.228)	(0.036)	(0.090)	(1.067)
Lower middle school	**Water plant**	-0.016***	0.031*	0.568***	-0.026	0.101	3.434***
		(0.006)	(0.017)	(0.178)	(0.020)	(0.068)	(1.063)
Upper middle school	**Water plant**	-0.010	0.042**	0.805***	-0.062*	0.098**	1.957**
		(0.008)	(0.018)	(0.267)	(0.032)	(0.048)	(0.911)

Notes: the other covariates in each regression are the same as ones in Table 9. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 21 Treatment Effects on Child Health

Dependent Variables	Illness in last four weeks		Weight-for-height		Height	
	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	-0.004	0.015	0.446***	0.355	0.962***	0.318
	(0.007)	(0.016)	(0.129)	(0.255)	(0.352)	(0.662)
Age	-0.006***	-0.006***	1.190***	1.185***	5.488***	5.468***
	(0.001)	(0.001)	(0.012)	(0.012)	(0.024)	(0.024)
Female	0.002	0.003	-0.446***	-0.486***	-1.895***	-1.840***
	(0.004)	(0.004)	(0.075)	(0.074)	(0.158)	(0.151)
Father’s education (years)	-0.002***	-0.002***	0.010	0.000	0.106***	0.069**
	(0.001)	(0.001)	(0.014)	(0.014)	(0.028)	(0.028)
Mother’s education (years)	-0.001	-0.000	0.022	0.006	0.189***	0.122***
	(0.001)	(0.001)	(0.015)	(0.015)	(0.030)	(0.028)
Household size	-0.003	-0.003	0.034	0.043	-0.272***	-0.260***
	(0.002)	(0.002)	(0.032)	(0.033)	(0.083)	(0.081)
Log income in first year	-0.000	-0.002	0.052	0.088	0.088	0.065
	(0.003)	(0.003)	(0.060)	(0.062)	(0.113)	(0.117)
Livestock	0.005	0.004	-0.388***	-0.173*	-0.705***	0.108
	(0.005)	(0.005)	(0.096)	(0.103)	(0.224)	(0.212)
Kms to the nearest medical facility	0.002	0.005	0.043	0.028	0.119	0.009
	(0.005)	(0.006)	(0.048)	(0.054)	(0.100)	(0.085)
Constant	0.276***	0.306***	12.272***	15.354***	77.893***	79.906***
	(0.039)	(0.031)	(0.744)	(0.585)	(1.364)	(1.329)
County fixed effect	Yes	No	Yes	No	Yes	No
Village fixed effect	No	Yes	No	Yes	No	Yes
Observations	14,394	14,394	12,141	12,141	12,075	12,075
R-squared	0.066	0.081	0.710	0.718	0.927	0.930
P value (bootstrap Hausman test)	0.243		0.697		0.309	

Notes: each column lists coefficient estimates with standard errors in parentheses (clustered at the village level) from separate regressions of a health outcome. In addition to the covariates listed above, each regression also controls for year fixed-effects. The bootstrap Hausman tests are based on 1000 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1

Table 22 Treatment Effects on Child Health in the Restricted Sample

Dependent Variables	Illness in last four weeks		Weight-for-height		Height	
	(1)	(2)	(3)	(4)	(5)	(6)
Water plant	0.001	-0.012	0.326	0.119	-0.127	-0.392
	(0.015)	(0.024)	(0.269)	(0.261)	(0.597)	(0.712)
Observations	3,197	3,197	2,701	2,701	2,686	2,686
R-squared	0.058	0.066	0.709	0.714	0.927	0.928

Notes: the other covariates controlled for in each regression are the same as ones in Table 21. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 23 Assignment to Treatment for Child Sample—OLS (First Stage)

Treatment	Child Sample Water plant (3)
Non-flat	-0.431***
	(0.103)
Age	0.004***
	(0.001)
Female	0.005
	(0.007)
Father’s education (years)	0.005**
	(0.002)
Mother’s education (years)	0.006**
	(0.003)
Household size	-0.000
	(0.005)
Log income in first year	0.008
	(0.011)
Livestock	-0.149***
	(0.034)
Kms to the nearest medical facility	-0.038**
	(0.018)
Constant	0.508*
	(0.280)
Observations	13,321
R-squared	0.371
F-stat on instruments	17.36
Prob>F	0.0001

Notes: the regression also controls for county and year fixed-effects. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 24 Treatment Effects of Water Program on Children' Health Status

Dependent Variable	Illness in last four weeks		Weight-for-height		Height	
	OLS	IV	OLS	IV	OLS	IV
	(2)	(3)	(5)	(6)	(8)	(9)
Water plant	-0.004	-0.023	0.446***	0.754*	0.962***	2.489*
	(0.007)	(0.026)	(0.129)	(0.426)	(0.352)	(1.433)
Age	-0.006***	-0.006***	1.190***	1.177***	5.488***	5.474***
	(0.001)	(0.001)	(0.012)	(0.013)	(0.024)	(0.026)
Female	0.002	0.002	-0.446***	-0.407***	-1.895***	-1.864***
	(0.004)	(0.004)	(0.075)	(0.074)	(0.158)	(0.169)
Father's education (years)	-0.002***	-0.002***	0.010	0.003	0.106***	0.088***
	(0.001)	(0.001)	(0.014)	(0.015)	(0.028)	(0.032)
Mother's education (years)	-0.001	-0.000	0.022	0.008	0.189***	0.172***
	(0.001)	(0.001)	(0.015)	(0.017)	(0.030)	(0.037)
Household size	-0.003	-0.002	0.034	0.053	-0.272***	-0.257***
	(0.002)	(0.002)	(0.032)	(0.033)	(0.083)	(0.090)
Log income in first year	-0.000	-0.001	0.052	0.071	0.088	0.004
	(0.003)	(0.003)	(0.060)	(0.058)	(0.113)	(0.128)
Livestock	0.005	0.002	-0.388***	-0.315**	-0.705***	-0.264
	(0.005)	(0.006)	(0.096)	(0.134)	(0.224)	(0.388)
Kms to the nearest medical facility	0.002	0.001	0.043	0.047	0.119	0.124
	(0.005)	(0.006)	(0.048)	(0.056)	(0.100)	(0.117)
Constant	0.276***	0.283***	12.272***	12.114***	77.893***	78.300***
	(0.039)	(0.039)	(0.744)	(0.713)	(1.364)	(1.521)
Observations	14394	13234	12,141	11,114	12,075	11,055
R-squared	0.066	0.065	0.710	0.717	0.927	0.929

Notes: each regression also controls for county and year fixed-effects. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 25 Treatment Effects on Child Health with and without Controlling for Nonlinear Terms of Age and Income

Dependent Variables	Illness in last four weeks		Weight-for-height		Height	
	Linear Model	Nonlinear Model	Linear Model	Nonlinear Model	Linear Model	Nonlinear Model
	(1)	(2)	(3)	(4)	(5)	(6)
			OLS Estimates			
Water plant	-0.004	-0.004	0.446***	0.416***	0.962***	1.043***
	(0.007)	(0.007)	(0.129)	(0.128)	(0.352)	(0.334)
			IV Estimates			
Water plant	-0.023	-0.021	0.754*	0.847**	2.489*	2.241
	(0.026)	(0.026)	(0.426)	(0.427)	(1.433)	(1.377)

Notes: the nonlinear terms are age^2 and log household income in the first $year^2$. The other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 26 Regression Results with Treatment Variables at Different Levels

Dependent Variables	Illness in last four weeks		Weight-for-height		Height	
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Household-level Treatment	-0.004	-0.025	0.465***	0.749**	0.875***	2.708*
	(0.005)	(0.027)	(0.111)	(0.377)	(0.262)	(1.376)
Village-level Treatment (**water plant**)	-0.004	-0.023	0.446***	0.754*	0.962***	2.489*
	(0.007)	(0.026)	(0.129)	(0.426)	(0.352)	(1.433)

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 27 Treatment Effects on Child Health across Different Cutoffs

	OLS Estimates			IV Estimates		
Cutoffs	Illness in last four weeks (1)	Weight-for -height (2)	Height (3)	Illness in last four weeks (4)	Weight-for -height (5)	Height (6)
10%	-0.005	0.485***	1.071***	-0.019	0.638**	2.108*
	(0.007)	(0.114)	(0.295)	(0.021)	(0.307)	(1.102)
15%	-0.004	0.555***	1.142***	-0.022	0.726*	2.393*
	(0.007)	(0.118)	(0.309)	(0.025)	(0.369)	(1.289)
20% (water plant)	-0.004	0.446***	0.962***	-0.023	0.754*	2.489*
	(0.007)	(0.129)	(0.352)	(0.026)	(0.426)	(1.433)
25%	-0.007	0.435***	1.024***	-0.026	0.872*	2.870
	(0.006)	(0.137)	(0.365)	(0.029)	(0.518)	(1.741)
30%	-0.006	0.407***	1.052***	-0.027	0.874*	2.878
	(0.007)	(0.139)	(0.382)	(0.030)	(0.520)	(1.745)

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 28 Treatment Effects on Child Health Controlling for Households' Sanitation Facilities and Environment

		OLS Estimates			IV Estimates		
Dependent Variables		Illness in last four weeks	Weight-for-height	Height	Illness in last four weeks	Weight-for-height	Height
		(1)	(2)	(3)	(4)	(5)	(6)
Water plant		-0.003	0.286**	0.500	-0.026	0.458	1.771
		(0.008)	(0.123)	(0.322)	(0.032)	(0.437)	(1.502)
Sanitation Environment	Little excreta	0.004	-0.204**	-0.418*	0.005	-0.224**	-0.401
		(0.006)	(0.096)	(0.225)	(0.007)	(0.102)	(0.256)
	Some excreta	0.002	-0.118	-0.784***	0.000	-0.116	-0.672**
		(0.008)	(0.115)	(0.266)	(0.009)	(0.121)	(0.310)
	Much excreta	0.029	0.005	-1.042	0.033	0.013	-0.919
		(0.025)	(0.286)	(0.801)	(0.025)	(0.296)	(0.736)
Toilet type	No bathroom	-0.005	0.280	-0.571	-0.006	0.389	-0.268
		(0.031)	(0.404)	(0.785)	(0.033)	(0.432)	(0.716)
	Flush toilet	-0.027	0.633**	1.288**	-0.018	0.531	0.838
		(0.020)	(0.289)	(0.626)	(0.023)	(0.346)	(0.766)
	Non-flush toilet	-0.022	0.197	-0.046	-0.021	0.152	-0.053
		(0.020)	(0.241)	(0.587)	(0.021)	(0.251)	(0.581)
	Open pit	-0.021	0.004	-0.814	-0.021	0.018	-0.670
		(0.019)	(0.209)	(0.530)	(0.019)	(0.211)	(0.514)
Constant		0.294***	12.430***	79.213***	0.303***	12.253***	79.388***
		(0.043)	(0.764)	(1.356)	(0.043)	(0.751)	(1.499)
	Observations	14,161	11,951	11,880	13,019	10,943	10,879
	R-squared	0.068	0.711	0.927	0.066	0.718	0.929

Notes: the omitted group of sanitation environment is "no excreta" and that of toilet types is "other". The other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 29 Treatment Effects Controlling for Other Infrastructure

	OLS Estimates			IV Estimates		
Dependent Variables	Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Water plant	-0.002	0.427***	0.759**	-0.020	0.761*	2.178
	(0.007)	(0.133)	(0.341)	(0.029)	(0.459)	(1.495)
Dirt roads around villages	0.013	-0.227**	-1.013***	0.012	-0.205	-0.784***
	(0.009)	(0.109)	(0.198)	(0.010)	(0.125)	(0.250)
Stone roads around villages	-0.004	-0.260**	-0.574***	-0.008	-0.209	-0.283
	(0.007)	(0.123)	(0.208)	(0.008)	(0.147)	(0.275)
Kms to the nearest primary school	0.006*	-0.016	-0.038	0.007*	-0.001	-0.044
	(0.004)	(0.052)	(0.116)	(0.004)	(0.050)	(0.108)
Kms to the nearest middle school	-0.000	0.003	-0.017	-0.000	-0.001	0.003
	(0.000)	(0.007)	(0.018)	(0.000)	(0.007)	(0.016)
Trade areas nearby	-0.002	0.014	-0.111	-0.001	-0.026	-0.157
	(0.010)	(0.133)	(0.243)	(0.012)	(0.146)	(0.288)
Telephone availability	-0.016**	-0.251**	0.166	-0.016	-0.319**	0.001
	(0.008)	(0.114)	(0.226)	(0.010)	(0.127)	(0.276)
Electricity	-0.021	0.002	0.308	-0.015	-0.041	1.197*
	(0.030)	(0.302)	(0.655)	(0.034)	(0.315)	(0.649)
Constant	0.303***	12.599***	78.050***	0.305***	12.518***	77.579***
	(0.048)	(0.815)	(1.418)	(0.051)	(0.802)	(1.527)
Observations	14,363	12,111	12,045	13,222	11,102	11,043
R-squared	0.068	0.711	0.927	0.067	0.717	0.929

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 30 Treatment Effects on Child Health with Controlling for Water Accessibility

PANEL A: the Whole Sample

	OLS Estimates			IV Estimates		
	Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Water plant	-0.003	0.457***	0.943***	-0.019	0.849*	2.589
	(0.007)	(0.133)	(0.344)	(0.029)	(0.498)	(1.587)
Water access	-0.004	-0.137	0.392	-0.001	-0.211	0.059
	(0.010)	(0.134)	(0.319)	(0.011)	(0.180)	(0.406)
Observations	14,262	12,031	11,963	13,109	11,016	10,955
R-squared	0.067	0.711	0.927	0.066	0.718	0.928

PANEL B: Sample always with Optimal Water Access

	OLS Estimates			IV Estimates		
	Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Water plant	-0.011	0.508***	1.443***	-0.003	1.509**	2.725**
	(0.009)	(0.181)	(0.367)	(0.034)	(0.709)	(1.319)
Observations	7,951	6,758	6,718	7,032	5,949	5,913
R-squared	0.063	0.706	0.925	0.063	0.718	0.928

PANEL C: Children under Age 10 in the First Three Waves

	OLS Estimates			IV Estimates		
	Weight (1)	Height (2)	BMI (3)	Weight (4)	Height (5)	BMI (6)
Water plant	0.560**	1.380***	-0.066	1.825	2.214	0.435
	(0.267)	(0.440)	(0.189)	(1.106)	(1.539)	(0.588)
Water access	-0.193	0.517	-0.112	-0.482	0.340	-0.241
	(0.217)	(0.330)	(0.158)	(0.331)	(0.455)	(0.212)
Observations	4,160	4,023	4,068	4,092	3,955	4,000
R-squared	0.587	0.890	0.167	0.582	0.889	0.164

PANEL D: Children under 10 always with Optimal Water Access in the First Three Waves

	OLS Estimates			IV Estimates		
	Weight (1)	Height (2)	BMI (3)	Weight (4)	Height (5)	BMI (6)
Water plant	1.030**	2.295***	0.204	2.769**	2.160	1.705*
	(0.409)	(0.560)	(0.270)	(1.197)	(1.808)	(0.957)
Observations	2,142	2,079	2,100	2,097	2,034	2,055
R-squared	0.546	0.891	0.162	0.538	0.891	0.144

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 31 Treatment Effects on Child Health for Mangyo (2008)'s Sample

PANEL A: the Replicated Sample

	OLS Estimates			IV Estimates		
	Weight (1)	Height (2)	BMI (3)	Weight (4)	Height (5)	BMI (6)
Water plant	0.466*	1.622***	-0.144	1.937*	3.426**	0.266
	(0.278)	(0.479)	(0.220)	(1.032)	(1.665)	(0.647)
Water access	-0.097	0.335	-0.125	-0.471	-0.106	-0.233
	(0.195)	(0.388)	(0.182)	(0.328)	(0.536)	(0.243)
Observations	2,780	2,506	1,966	2,723	2,452	1,918
R-squared	0.618	0.889	0.127	0.610	0.887	0.126

PANEL B: Children always with Optimal Water Access in the Replicated Sample

	OLS Estimates			IV Estimates		
	Weight (1)	Height (2)	BMI (3)	Weight (4)	Height (5)	BMI (6)
Water plant	1.156*	3.366***	0.107	4.527***	6.573***	2.130***
	(0.640)	(0.830)	(0.408)	(1.052)	(2.357)	(0.611)
Observations	1,018	937	748	983	906	719
R-squared	0.587	0.883	0.197	0.561	0.878	0.168

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Table 32 Heterogeneous Treatment Effects on across Income and Parents' Educational Groups

PANEL A: Income Groups

Income Group		OLS Estimates			IV Estimates		
		Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Poor	**Water plant**	-0.006	0.739***	2.723***	0.049***	1.284*	3.251***
		(0.016)	(0.249)	(0.505)	(0.018)	(0.726)	(1.052)
Middle	**Water plant**	-0.013	0.161	0.070	-0.027	0.388	-0.025
		(0.011)	(0.184)	(0.488)	(0.026)	(0.485)	(1.211)
Rich	**Water plant**	0.016	0.644	1.326	-0.143	1.826	5.924
		(0.014)	(0.403)	(0.837)	(0.217)	(1.335)	(7.980)

PANEL B: Mothers' Educational Groups

Educational Groups		OLS Estimates			IV Estimates		
		Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Illiterate	**Water plant**	-0.008	0.424**	0.525	-0.093*	0.970	2.173
		(0.012)	(0.195)	(0.773)	(0.054)	(0.631)	(2.028)
Primary school	**Water plant**	-0.009	0.438**	0.946**	-0.051*	0.954*	1.316
		(0.009)	(0.172)	(0.473)	(0.030)	(0.558)	(1.539)
Lower middle school	**Water plant**	0.002	0.483**	0.828*	0.027	0.878*	3.739**
		(0.012)	(0.189)	(0.433)	(0.045)	(0.511)	(1.605)
Upper middle school and above	**Water plant**	-0.015	0.486	2.408***	-0.042	1.198**	2.790***
		(0.017)	(0.296)	(0.653)	(0.031)	(0.575)	(0.931)

PANEL C: Fathers' Educational Groups

Educational Groups		OLS Estimates			IV Estimates		
		Illness in last four weeks (1)	Weight-for-height (2)	Height (3)	Illness in last four weeks (4)	Weight-for-height (5)	Height (6)
Illiterate	**Water plant**	-0.012	0.503	1.051	0.138***	-0.706	5.631*
		(0.025)	(0.395)	(0.700)	(0.051)	(1.046)	(3.239)
Primary school	**Water plant**	-0.009	0.322*	0.610	-0.059	0.466	0.917
		(0.011)	(0.186)	(0.518)	(0.038)	(0.521)	(1.525)
Lower middle school	**Water plant**	0.006	0.265**	1.109***	-0.001	1.099*	3.632**
		(0.009)	(0.125)	(0.368)	(0.039)	(0.608)	(1.458)
Upper middle school and above	**Water plant**	-0.019	0.773**	1.091	-0.059	0.639	2.605
		(0.014)	(0.319)	(0.707)	(0.038)	(0.930)	(2.496)

Notes: the other covariates in each regression are the same as ones in Table 24. The standard errors in parentheses are clustered at the village level. *** p<0.01, ** p<0.05, * p<0.1

Bibliography

Abadie, Alberto, “Semiparametric Instrumental Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, Apr. 2003, 113(2): 231-263.

Angrist, Joshua, “Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice,” *Journal of Business & Economic Statistics*, Jan. 2001, 19(1): 2-16.

Angrist, Joshua D. and William N. Evans, “Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size,” *The American Economic Review*, Jun. 1998, 88(3): 450-477.

Angrist, Joshua and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: an Empiricist’s Companion*, Princeton University Press, 2009.

Ahmed SA, BA Hoque and A Mahmud, “Water Management Practices in Rural and Urban Homes: A Case Study from Bangladesh on Ingestion of Polluted Water,” *Public Health*, Mar. 1998, 112(5): 317-321.

Arnold, Benjamine and John M. Coldford JR., “Treating Water with Chlorine at Point-of-use To Improve Water Quality and Reduce Child Diarrhea in Developing Countries: A Systematic Review and Meta-analysis,” *American Journal of Tropical Medicine and Hygiene,* Feb. 2007, 76(2): 354-364.

Aturupane, Harsha, Anil B. Deolalikar, and Dileni Gunewardena, “The Determinants of Child Weight and Height in Sri Lanka: A Quantile Regression Approach,” May 2008, UNU-WIDER Research Paper No. 2008/53.

Banerjee, Abhijit, Esther Duflo and Nancy Qian, “The Route to Success: The Effect of Infrastructure on Economic Growth,” 2009, working paper.

Baker, Michael, Mark Stabile and Catherine Deri Armstrong, (2004), “What Do Self-Reported, Objective, Measures of Health Measure?,” *Journal of Human Resources*, Autumn 2004, 39(4): 1067-1093.

Barten, Anton P. “Family Composition, Prices and Expenditure Patterns.” In *Econometric Analysis for National Economic Planning*, edited by Peter E. Hart, Gordon Mills, and John K. Whitaker. London: Butterworths, 1964.

Benette, Daniel, “Clean Water Makes You Dirty: Water Supply and Sanitation Behavior in Philippines,” 2007, Working paper.

Bleakley, Hoyt, “Disease and Development: Evidence from Hookworm Eradication in the American South,” *Quarterly Journal of Economics*, Feb. 2007, 122 (1): 73-117.

Bound, J., D.A. Jaeger, and R.M. Baker, "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, Jun. 1995, 90(430): 443-450.

Braudel, Fernand, *Civilization and Capitalism: 15th-18th Century*, University of California Press, 1992.

Briend, Andre, "Is Diarrhoea a Major Cause of Malnutrition among the Under-fives in Developing Countries? A Review of Available Evidence," *European Journal of Clinical Nutrition*, Sept. 1990, 44: 611-628.

Cai, Fang and Dewen Wang, "Migration as Marketization: What Can We Learn from China's 2000 Census Data?," *The China Review*, Fall 2003, 3(2): 73-93.

Cameron, A. Colin, and Pravin K. Trivedi, *Microeconometrics: Methods and Application*, Cambridge University Press, May 2005.

Case, Anne, Darren Lubotsky and Christina Paxson, "Economic Status and Health in Childhood: The Origins of the Gradient," *American Economic Review*, American Economic Association, Dec. 2002, 92(5): 1308-1334.

Carrasco Raquel, "Binary Choice with Binary Endogenous Regressors in Panel Data: Estimating the Effect of Fertility on Female Labor Participation," *Journal of Business & Economic Statistics*, Oct. 2001, 19(4): 385-394.

Chang, Suying, Wu He and Chunming Chen, "The Growth Characteristics of Children under 5 in the Past 15 Years," *Journal of Hygiene Research*, Nov. 2006, 35(6): 768-771.

Chattopadhyay, Raghabendra and Esther Duflo, "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *Econometrica*, Sept. 2004, 72(5): 1409-1443.

Chernozhukov, Victor and Christian Hansen, "The Reduced Form: A Simple Approach to Weak Inference," 2007, University of Chicago, Graduate School of Business.

Cöl, M., C. Cöl, A. Soran, B.S. Sayli, and S. Oztürk, "Arsenic-related Bowen's Disease, Palmar Keratosis, and Skin Cancer," *Environmental Health Perspectives*, Aug. 1999, 107(8): 687-689.

Curtis, V., and S. Cairncross, "Effect of Washing Hands with Soap on Diarrhoea Risk in the Community: A Systematic Review," *The Lancet Infectious Diseases*, May 2003, 3: 275-281.

Cutler, David and Grant Miller, "The Role of Public Health Improvement in Health Advances: The 20th Century United States," *Demography*, Feb. 2005, 42(1): 1-22

de Mel, Suresh, David J. McKenzie and Christopher Woodruff, "Measuring Microenterprise Profits: Must We Ask How the Sausage is Made?," *Journal of Development Economics*, Jan. 2009, 88(1): 19-31.

de Onis, Mercedes, "Measuring Nutritional Status in Relation to Mortality", *Bulletin of the World Health Organization*, 2000, 78(10): 1271-4.

Dinkelman, Taryn, "The Effects of Rural Electrification on Employment: New Evidence from South Africa," Jan. 2008, Working paper.

Donaldson, Dave, "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," 2009, London School of Economics, working paper.

Duflo, Esther and Rohini Pande, "Dams," *Quarterly Journal of Economics*, May 2007, 122 (2): 601-646.

Duflo, Esther, Michael Greenstone, and Rema Hanna, "Indoor Air Pollution, Health and Economic Well-being," Feb. 2008, Working paper.

Ebenstein, Avraham, "Water Pollution and Digestive Cancers in China," Aug. 2008, Working paper.

Esrey, Steven A., "Water, Waste, and Well-Being: A Multicountry Study," *American Journal of Epidemiology*, 1996, 143(6): 608-623.

Fewtrell, Lorna, David Kay, Mark Wyer and Gary O'Neill, "An Investigation into the Possible Links between Shigellosis and Hepatitis A and Public Water Supply Disconnections," *Public Health*, May 1997, 111(3): 179-181.

Fewtrell, Lorna, Rachel B kaufmann, David Kay, Wayne Enanoria, Laurence Haller and John M. Colford, Jr., "Water, Sanitation and Hygiene Interventions to Reduce Diarrhoea in less developed countries: A Systematic Review and Meta-analysis," *The Lancet Infectious Diseases*, Jan. 2005, 5(1): 42-52.

Foster, Andrew D., "Poverty and Illness in Low-income Rural Areas," *The American Economic Review*, May 1994, 84(2): 216-220. (papers and proceedings of the hundred and sixth Annual meeting of the American Economic Association (May 1994)).

Freedman David A., "Statistical Models and Shoe Leather," *Sociological Methodology*, 1991, 21: 291-313.

Galiani, Sebastian, Paul Gertler and Ernesto Schargrodsky, “Water for Life: The Impact of the Privatization of Water Services on Child Mortality,” *Journal of Political Economy,* Feb. 2005, 113(1): 83-120.

Galiani, Sebastian, Martin Gonzalez-Rozada, and Ernesto Schargrodsky, “Water Expansions in Shantytowns: Health and Savings,” *Economica*, Oct. 2009, 76(304): 607-622,

Gamper-Rabindran, Shanti, Shakeeb Khan and Christopher Timmins, “The Impact of Piped Water Provision on Infant Mortality in Brazil: A Quantile Panel Data Approach,” 2008, NBER Working Papers No.14365.

Gleick, Peter H., “Dirty Water: Estimated Deaths from Water-Related Diseases 2000–2020,” Aug. 2002, Pacific Institute Research Report.

Habicht, J.P., R. Martorell and J.A. Rivera, “Nutritional Impact of Supplementation in the INCAP Longitudinal Study: Analytic Strategies and Inferences,” *Journal of Nutrition*, Apr. 1995, 125 (4 Suppl): 1042S-1050S.

Humphrey, J.H., “Child Undernutrition, Tropical Enteropathy, Toilets, and Handwashing,” *The Lancet*, Sept. 2009, 374: 1032-1035.

Jacoby, Hanan and Limin Wang, ‟Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach,” March, 2004, World Bank Policy Research Working Paper No.3241.

Jalan, Jyotsna, and Martin Ravallioin, “Does Piped Water Reduce Diarrhea for Children in Rural India?,” *Journal of Econometrics*, Jan. 2003, 112 (1): 153-173.

Jensen, Robert and Emily Oster, “The Power of TV: Cable Television and Women’s Status in India,” Sept. 2008, Working paper.

Kling, Jeffrey R., “Interpreting Instrumental Variables Estimates of the Returns to Schooling,” *Journal of Business & Economic Statistics*, Jul. 2001, 19(3): 358-364.

Kremer, Michael, Jessica Leino, Edward Miguel and Alix Peterson Zwane, “Spring Cleaning: A Randomized Evaluation of Source Water Quality Improvement,” Mar. 2009, Working paper.

Lee, Lung-fei, Mark R. Rosenzweig, and Mark M. Pitt, “The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High Mortality Populations,” *Journal of Econometrics*, Mar. 1997, 77 (1): 209-235.

Lipscomb, Molly, A. Mushfiq Mobarak, and Tania Barham, “Returns to Electricity: Evidence from the Quasi-Random Placement of Hydropower Plants in Brazil,” Jul. 2008, Working paper.

Logan, Trevon, “Economies of Scale in the Household: Puzzles and Patterns from the American Past,” Mar. 2008, NBER Working paper No.13869.

Mangyo, Eiji, “The Effect of Water Accessibility on Child Health in China,” *Journal of Health Economics*, Apr. 2008, 27(5): 1343-1356.

McKeown, Thomas, *The Role of Medicine: Dream, Mirage or Nemesis*, Princeton University Press, 1979.

Merrick, Thomas W, “The Effect of Piped Water on Early Childhood Mortality in Urban Brazil, 1970 to 1976,” *Demography*, Feb. 1985, 22 (1): 1-24.

Meng et al., “Water Supply and Sanitation Environment in Rural China: Promote Service to the Poor,” 2005, Poverty Reduction Conference, Shanghai, China.

Mikusheva, A. and C.B. Poi, “Tests and Confidence Sets with Correct Size in the Simultaneous Equations Model with Potentially Weak Instruments,” *The Stata Journal*, Sept. 2006, 6(3): 335-347.

Ministry of Health of the People’s Republic of China, *Chinese National Health Statistics (2006)*, Peking Union Medical College Press, 2006.

Ministry of Health of the People’s Republic of China, *Chinese National Health Statistics (2007)*, Peking Union Medical College Press, 2007.

Ministry of Health of the People’s Republic of China, “The implementation of the water and sanitation improvement program in 2006,” *Patriotic Health Brief*, May 2007.

Moreira, M. and L.M. Cruz, “On the Validity of Econometric Techniques with Weak Instruments: Inference on Returns to Education Using Compulsory School Attendance Laws,” *Journal of Human Resources*, 2005, 40 (2): 393-410.

National Bureau of Statistics of China, *China Statistical Yearbook 2001*, China Statistics Press, 2001.

National Bureau of Statistics of China, *China Statistical Yearbook 2006*, China Statistics Press, 2006.

National Bureau of Statistics of China, *China Statistical Yearbook 2007*, China Statistics Press, 2007.

Pelletier, D.L., E.A. Frongillo and J.P. Habicht, “Epidemiologic Evidence for a Potentiating Effect of Malnutrition on Child Mortality,” *American Journal of Public Health*, 1993, 83(8): 1130-1133.

Pelletier, D.L. “The Relationship between Child Anthropometry and Mortality in Developing Countries: Implications for Policy, Programs and Future Research,” *The Journal of Nutrition*, Oct. 1994, 124 (10 Suppl): 2047-2081.

Qian, Nancy, “Missing Women and the Price of Tea in China: The Effect of Relative Female Income on Sex Imbalance,” *Quarterly Journal of Economics*, Aug. 2008, 123 (3): 1251-85.

Rabie, T. and Curtis, V., “Handwashing and Risk of Respiratory Infections: A Quantative Systematic Review,” *Tropical Medicine and International Health*, 2006, 11(3): 258-267.

Ravenscroft, Peter, “Arsenic in Drinking Water: a Global Threat to Health,” University of Cambridge, 29 Aug. 2007, www.admin.cam.ac.uk/news/dp/2007082901 (accessed on May 23, 2011).

Schieber, George, Lisa Fleisher and Pablo Gottret, “Getting Real on Health Financing,” *Finance and Development*, Dec. 2006, 43(4): 46-50.

Schmidt, Wolf-peter and Sandy Cairncross, “Household Water Treatment in Poor Populations: Is There Enough Evidence for Scaling up Now?,” *Environmental Science Technology*, Jan. 2009, 43(4): 986-992.

Staiger, D. and J.H. Stock, “Instrumental Variables Regression with Weak Instruments,” *Econometrica*, May 1997, 65(3): 557-586.

Stanton, Bonita, John D Clemens and Tajkera Khair, “Educational Intervention for Altering Water-sanitation Behavior to Reduce Childhood Diarrhea in Urban Bangladesh: Impact on Nutritional Status,” *The American Journal of Clinical Nutrition*, 1988, 48(5): 1166-1172.

Stevens, Philip, *Diseases of Poverty and the 10/90 Gap*, Nov. 2004, Hanway Print Centre, United Kingdom.

Stock, J.H. and M. Yogo, Testing for Weak Instruments in Linear IV Regression, in D.W.K. Andrews and J.H. Stock (eds.), *Identification and Inference for Econometric Models, Essays in Honor of Thomas Rothenberg*, Cambridge University Press, 2005.

Stock, J.H. and Mark W. Watson, "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression,” 2006, NBER Technical Working Paper No. 323.

Strauss, John and Duncan Thomas, “Health, Nutrition, and Economic Development," *Journal of Economic Literature*, Jun. 1998, 36(2): 766-817.

Tao, Yong, "Drinking Water and Health in Rural China," The Chinese Hydraulic Engineering Society Annual Conference 2008, http://www.chinawater.net.cn/zt/08nianhui/CWSNews_View.asp?CWSNewsID=30077 (accessed on May 25, 2011).

Troesken, Werner, "Lead Water Pipes and Infant Mortality in Turn-of-the-century Massachusetts," 2003, NBER Working Paper 9549.

Udry, Christopher, "Gender, Agricultural Production, and the Theory of the Household," *Journal of Political Economy*, Oct. 1996, 104(5): 1010-1046.

Wang, S., "Credit Constraints, Job Mobility and Entrepreneurship: Evidence from a Property Reform in China," Dec. 2008, Working Paper.

Wang, S., Z. Wang, X. Cheng, J. Li, Z. Sang, X. Zhang, L. Han, X. Qiao, Z. Wu and Z. Wang, "Arsenic and Fluoride Exposure in Drinking Water: Children's IQ and Growth in Shanyin County, Shanxi Province, China," *Environmental Health Perspectives,* Apr., 2007, 115(4): 643-647.

Wang, J., J. Huang, and S. Rozelle, "Evolution of Tubewell Ownership and Production in the North China Plain," *Australian Journal of Agricultural and Resource Economics*, Jun. 2005, 49(2): 177-196.

Wang, J., J. Huang, A. Blanke, Q. Huang, and S. Rozelle, "The Development, Challenges and Management of Groundwater in Rural China," In: Giordano M, Villholth KG (eds) *The Agricultural Groundwater Revolution: Opportunities and Threats to Development*, International Water Management Institute, 2006.

Waterloo, J.C., "Classification and Definition of Protein-Calorie Malnutrition," *British Medical Journal*, Sept. 1972, 5826(3): 566-569.

Werner, Troesken, "Lead Water Pipes and Infant Mortality at the Turn of the Twentieth Century," *Journal of Human Resources*, Summer 2008, 43(3): 553-575.

Wooldridge, J, "Cluster-Sample Methods in Applied Econometrics," *The American Economic Review*, Vol. 93, No. 2, Papers and Proceedings of the One Hundred Fifteenth Annual Meeting of the American Economic Association, Washington, DC, January 3-5, 2003 (May 2003): 133-138.

Wei, Chengyu, "General Status of Prevention and Control of Diarrhea, and Principal Measures in New China," *China Tropical Medicine*, 2008, 59(6): 883-888.

World Bank, *Cost of Pollution in China: Economic Estimates of Physical Damages*, World Bank, 2007.

World Health Organization, *The World Health Report 2005: Make Every Mother and Child Count*, World Health Organization, 2005.

World Health Organization, *A Safer Future: Global Public Health Security in the 21st Century*, World Health Organization, 2007.

World Health Organization Multicentre Growth Reference Study Group, *WHO Child Growth Standards: Length/height-for-age, Weight-for-age, Weight-for-length, Weight-for-height and Body Mass Index-for-age: Methods and Development*, World Health Organization, 2006.

Yang, Yunyan, “The New Characteristics of China’s Migration Since 1990s,” *South China Population*, Sept. 2004, 19(03): 13-20.

Zhang, Rong, Hongxing Li, Xiaofeng Wu, et al., “Current Situation Analysis on China Rural Drinking Water Quality”, *Journal of Environment and Health (Chinese Version)*, Jan. 2009, 26(1): 3-5.

Zwane, Alix Peterson and Michael Kremer, “What Works In Fighting Diarrheal Diseases In Developing Countries? A Critical Review,” May 2007, NBER Working Papers No. 12987.