

# Surface Water Quality and Infant Mortality in China

Guojun HE, Jeffrey PERLOFF

HKUST IEMS Working Paper No. 2015-32

May 2016

---

HKUST IEMS working papers are distributed for discussion and comment purposes. The views expressed in these papers are those of the authors and do not necessarily represent the views of HKUST IEMS.

More HKUST IEMS working papers are available at:  
<http://iems.ust.hk/WP>

---



# Surface Water Quality and Infant Mortality in China

Guojun HE, Jeffrey PERLOFF

HKUST IEMS Working Paper No. 2015-32  
May 2016

## Abstract

Surface water pollution has a significant, non-monotonic effect on the infant mortality rate in China. As surface water quality deteriorates, the infant mortality rate first increases and then decreases. Thus, moderate levels of pollution are the most dangerous.

## Authors' contact information

Guojun He  
Division of Social Science, Division of Environment, Department of Economics  
Hong Kong University of Science and Technology  
E: [gjhe@me.com](mailto:gjhe@me.com)  
Click here to enter text.

Jeffrey Perloff  
Department of Agricultural and Resource Economics  
UC Berkeley and The Gianini Foundation  
E: [jperloff@berkeley.edu](mailto:jperloff@berkeley.edu)

# Surface Water Quality and Infant Mortality in China

Authors:

Guojun He, Division of Social Science, Division of Environment and Department of Economics,  
The Hong Kong University of Science and Technology; Phone number: +(852)-3469-2423;  
Email address: [gjhe@me.com](mailto:gjhe@me.com).

Jeffrey M. Perloff, Department of Agricultural and Resource Economics, UC Berkeley and The  
Gianini Foundation; Email Address: [jperloff@berkeley.edu](mailto:jperloff@berkeley.edu).

Acknowledgement:

We are extremely grateful to Michael Anderson and William Dow for extensive comments on an earlier version of this paper when it was part of the first author's dissertation. We also thank Elisabeth Sadoulet, Meredith Fowlie, David Roland-Holst, Maximilian Auffhammer, Jeremy Magruder, Yuanli Liu, Jing Ma and Maoyong Fan for their constructive suggestions. We are grateful to the seminar participants in the Harvard School of Public Health and the Department of Agricultural and Resource Economics at UC Berkeley. We thank Shuming Bao, Gordon Liu and Avraham Ebenstein for their help in the process of collecting the data. Qilun Zhou provided superior research assistance. Any errors are the authors'.

# Surface Water Quality and Infant Mortality in China

## Abstract

Surface water pollution has a significant, non-monotonic effect on the infant mortality rate in China. As surface water quality deteriorates, the infant mortality rate first increases and then decreases. Thus, moderate levels of pollution are the most dangerous.

**Keywords:** water quality; water pollution; infant mortality

**JEL Classification:** Q53, I1

## I. Introduction

China's rapid industrialization led to severe water pollution in many areas due to massive industrial wastewater discharges and extensive use of agricultural fertilizer. Even though about 60 to 70 percent of river water is unsafe for human consumption (World Bank 2006), many people in poor areas rely on the surface water systems for daily use including drinking.

Consequently, this water pollution endangers the health of the rural population, especially that of infants. Using data on surface water quality from monitoring sites in 19 provinces and county-level infant mortality rate data from the 2000 Census, we estimate the effects of surface water quality on the infant mortality rate in China.

We hypothesize that if surface water becomes slightly degraded, people do not notice the pollution and continue consuming it. Consequently, the infant mortality rate initially rises as water pollution increases. However, as the pollution gets worse, people begin to notice the

pollution using visual and other clues and reduce their consumption of surface water. As the degree of pollution becomes very pronounced and hence very apparent, the infant mortality rate falls.

We discuss the literature on the effects of surface water quality on health in Section 2 and Chinese surface water quality in Section 3. In Section 4, we describe the data set, define the key variables, and provide summary statistics. We analyze whether we are likely to face a sample selection problem and describe an ordered-probit model to deal with that issue in Section 5. In Section 6, we present our estimates of the effects of surface water quality on the infant mortality rate using an ordered-probit selection model. We draw conclusions in Section 7.

## **II. The Effects of Water Pollution on Health**

The health effects of water pollution has been an important research topic in environmental science, epidemiology, environmental economics, and health economics for centuries. The earliest research in this area dates to the 1850s (see Freedman 1991). Using a natural experiment based on the distribution of water in London, John Snow demonstrated that unsanitary water caused cholera outbreaks. He found that the death rate for dirty-water users were over eight times higher than that for the clean-water users.

Many recent studies reported various connections between water pollution and diseases and other public health measures. Some studies focused on water pollution and water-borne diseases, such as typhoid (Cutler and Grant 2005) and diarrhea (Jalan and Rovalion 2003). Other studies explored the relationship between water pollution and cancer, such as Cantor (1997), Davis and Masten (2004), Chen et al. (2005), and Ebenstein (2012).

Much of the research focused on infant and child mortality. Galiani et al. (2005) found that privatizing water service improves water quality, and reduces child mortality. Merrick (1985),

Lavy et al. (1996), and Lee et al. (1997) found positive associations between water quality and infant health. Greenstone and Hanna (2014) investigated the effects of air quality and water quality on the infant mortality rate in India, and found no significant relationship between the infant mortality rate and water quality. Brainerd and Menon (2012) reported a negative effect of fertilizer agrichemicals in water on infant and child health in India. Currie et al. (2013) found that contaminated drinking water has large and statistically significant effects on birth weight and gestation of infants born to less educated mothers.

Several studies have examined people's pollution avoidance behavior (Mansfield et al. 2006; Shimshack et al. 2007; Neidell 2009; Moretti and Neidell 2011; Zivin et al. 2011). For example, Neidell (2009) found that when smog alerts were issued, attendance at major outdoor facilities in Los Angeles decreased significantly. Zivin et al. (2011) found that bottled water sales significantly increased when consumers were informed about tap water quality deteriorations (or violations) in Northern California and Nevada. These findings suggest that the provision and public dissemination of information about pollution can encourage the public to engage in avoidance behaviors to decrease exposure and minimize health risks.

### **III. Surface Water Quality in China**

In China, the overall surface water quality is graded based on chemical pollutant indicators, including the pH-value and the concentrations (measured by mg/L) of dissolved oxygen, biochemical oxygen demand, ammonia, and nitrogen. The overall surface water quality is graded on a 6-degree scale, where Type I water is the best quality water and Type VI is the worst.

According to the China Ministry of Water Resources, Type I water is an "Excellent" source of potable water. Type II water is a "Good" source of potable water. Type III water is "Fair."

Because Type II and III water may have pathogenic bacteria and parasites ova, drinking that water may cause diseases. Type II and III water should be purified and treated (such as by boiling) before drinking. Type IV water is polluted and unsafe to drink without advanced treatment, which is only possible at water supply plants. Type V is seriously polluted and can never be used as for human consumption. Type VI water is called “Worse than Type V Water,” and any direct contact with it is harmful to humans.<sup>1</sup>

Individuals can easily distinguish clean water from heavily polluted water, such as Types V and VI. Very polluted water is murky and smelly and may have algal blooms on the surface. Distinguishing “excellent” water from “good” or “fair” water is more challenging.

Not all the harms from consuming polluted water are immediately apparent. Some effects occur at once, while others appear only after toxins have accumulated in the body. For example, consuming contaminated water can cause malaria outbreaks within days, but it may take decades for water pollution to cause cancers. In this study, we focus on infant deaths that occur within a year.

#### **IV. Data and Summary Statistics**

China established a nationwide water quality monitor water system in the 1980s. Each year, the Ministry of Water Resource publishes the China National Water Resource Yearbook, which provides water quality information for major lakes, rivers and reservoirs. Many provinces also publish province-level water-body quality measures. The national and provisional publications are the sources of water quality data used in this paper.

---

<sup>1</sup> See the online appendix Table A1 for the limit values of water quality indicators.

We identified each water quality monitoring site's location and matched it with the 2000 Census data at the county-level.<sup>2</sup> Our sample includes 461 counties in 19 provinces. The sampled provinces include Anhui, Beijing, Chongqing, Fujian, Guangxi, Guizhou, Hainan, Hebei, He'nan, Jiangsu, Jiangxi, Liaoning, Ningxia, Shandong, Shanghai, Sichuan, Tianjin, Yunnan and Zhejiang. Guangxi and Ningxia are in the Autonomous Region; Beijing, Chongqing, Shanghai and Tianjin are directly controlled municipalities; and the rest are governed at the provincial level. Northern provinces are relatively under-sampled because we could not find water quality data for several Northern provinces.

The distribution of the six types of water quality is presented in Table 1. Nearly half (47 percent) of the rivers and other bodies of water in our sample are seriously polluted (Types IV, V, and VI).

The infant mortality rate is the annual number of deaths among infants (less than one year old) per thousand live births. The overall Chinese infant mortality rate has been decreasing at a rapid pace over the past forty years, from about 150 per thousand in the 1960s to around 20 per thousand in the 2000s.

Neonatal death is the major component of infant deaths. About 60 to 70 percent of infant deaths occurred within the first month after they were born. Most of the deaths during the neonatal period are due to endogenous causes (inherited defects), such as congenital anomalies, gestational immaturity, birth complications, and other physiological problems. Lack of proper

---

<sup>2</sup> There are five administrative levels in China. The highest level is the provincial level, which include provinces, autonomous regions, direct-controlled municipality, and special administrative regions (Hongkong and Macao). The second highest level of government is the prefecture level which includes prefectures, autonomous prefectures, prefecture-level cities and Leagues. Next is the county level, which includes counties, autonomous counties, county-level cities, and city districts. Below the county is the township level and the (informal) village level. The Census data are not available for these last two levels.



care during the pregnancy can cause neonatal deaths. For example, if pregnant women drink polluted water, the neonatal mortality may rise. The post-neonatal mortality rate describes the death rate of infants one month to one year. The vast majority of post-neonatal deaths are from exogenous causes, such as injuries, environmental and nutritional factors, especially as they interact with infectious disease like gastroenteritis and pneumonia. Water pollution may cause both neonatal and post-neonatal mortality.

The infant mortality rate summary statistics are presented in Table 2. In our sample, the average the infant mortality rate across counties is 19.2 per thousand. The variation in infant mortality rates is large, with a standard deviation of 15.8. The distribution of infant mortality rate is skewed to the left, with a few counties out in the tail with considerably higher infant mortality rates than the average. The lowest infant mortality rate in our data is less than 1 per thousand, while the highest infant mortality rate is 81.2 per thousand.

In general in China, female infants are more likely to die than are male infants. The female infant mortality rate is about 22 per thousand, and male infant mortality rate is about 17 per thousand. The difference is striking because male infants are usually more vulnerable and thus more likely to die than female infants on purely medical grounds. This difference is consistent with the popular argument that rural Chinese people prefer boys to girls so they invest more on male infants' health. The variance of the female infant mortality rate is also higher than that of the male infant mortality rate.

To isolate the effect of water pollution on health, we include a set of control variable from the 2000 Census of China: the percentage of the population that is non-agricultural, illiteracy rate, average rooms per home, and per capita housing area (square meters). We also include a few social-economic variables in the 2000 China Statistical Yearbook: per capita GDP (Chinese

Yuan), per capita government expenditure, and the number of beds in medical institutions per 10,000 people. We show the summary statistics of these variables in Table 3.

The non-agricultural share of the population that lives in non-agricultural areas (including the mobile population) to the total population. Urbanization, which reduces the non-agricultural share, has both positive and negative health effects, and the net impact on population is not obvious (see, for example, Van de Poel et al. 2009). It is generally believed that the positive consequences of urbanization outweigh the negative ones for infants' health. Urbanization is associated with better sanitation and medical treatments, easier access to tap water and infant care, all of which play important roles in improving infant health. So, we expect a negative association between the percentage of non-agricultural population and the infant mortality rate.

The illiteracy rate variable is the share of the total population over 15 who are illiterate (have not completed an elementary-school education). Many studies have found that infants are less likely to die the more educated are their parents, especially their mothers (Delgado et al. 2002; Behrman et al. 2003; Basu and Stephenson 2005; Frost et al. 2005; Miguel 2005; Boyle 2006; Cowell 2006; Cutler and Lleras-Muney 2006).

Poor housing and overcrowding are negatively associated with infant health (Martin 1967; Brennan and Lancashire 1978; Victora, et al. 1988). We expect that as either of our measures of living conditions—the average number of rooms per household and the per capita housing area (square meters)—increases, the infant mortality rate falls.

We expect per capita county GDP to be negatively correlated with the infant mortality rate. We use per capital government expenditure to approximate the local government's investments on social welfare programs, such as public health insurance, sanitation maintenance, tap water

provision, waste management, and pollution treatment. We expect it to be negatively correlated with the infant mortality rate.

We use the number of beds per 10,000 people in medical institutions as a measure of the availability of medical treatment. In regions where hospitals are readily available, the infant mortality rate should be lower.

The reliability of the data collected by Chinese government is often questioned by academic researchers. Local governments in China have been criticized for hiding news from the public about water pollution accidents. Consequently, we examined the consistency of the data in three ways.

In China, multiple agencies (e.g. the Ministry of Water Resource, the Environment Protection Agency, and the Center for Disease Control and Prevention) collected data on surface water quality. Since we do not know the raw data sources of different water quality reports, we first checked for consistency between reports of water quality data between national water resources reports and provincial water resources. We found that 98 percent of these paired reports are consistent, and the few differences are all within 1 water quality level. In the few case where they differ, we rely on the data from the provincial water resource reports.<sup>3</sup>

Second, we compared a subsample of the our 2004 water quality data with 2004 water quality data provided by the World Bank, which is used in Ebenstein's (2012) paper. The data are almost always identical for comparable monitoring sites.

Third, several monitoring sites' data are provided by the River Basin Water Quality Reports. Again, we found no substantial difference between these and other sources.

---

<sup>3</sup> If we drop these observations, we obtain the same qualitative results reported below.

Thus, the various reports on water quality are consistent. In the past several years, the government has trumpeted its efforts to promote the transparency of surface water quality information. For example, the central government started to release weekly water quality report for 100 national monitoring sites to the public in 2004, and it started to publicize real-time water quality data in 2009.

## **V. Sample Selection**

To estimate the effects of water pollution on health consistently, we need to avoid sample selection bias due to endogenous migration. Ebenstein (2012) suggests that China provides an ideal context to estimate the health effects of water pollution because of the Household Registration System (Hukou). This system prevents people from moving from rural to urban areas. It also makes it relatively difficult to move within rural and urban areas. A variety of benefits, such as health care and social security, are associated with the Household Registration System. If the system were to prevent people from migrating, we could treat people's location and thus the water quality they face as exogenous.

Although this argument may have held in the past, the share of people migrating increased by more than an order of magnitude between the 1982, when it was 0.66 percent according to the Census, and 2000 when it was 7.9 percent. Given that many rural migrant workers are of childbearing age, if the migration decision is correlated with water pollution levels, we may face a selection problem.

We address the potential sample selection problem using a Heckman-type selection model. In the first step, we estimate the relationship between the endogenous water pollution on the instrument variables. Conditional on the estimates from the first step, we estimate the effects of water pollution on the infant mortality rate in the second step.

We use wastewater dumping and precipitation as instruments for surface water quality. Industrial wastewater dumping degrades surface water quality. We use the amount of untreated wastewater discharged into the water system at the prefectural level as an instrument for the water quality in a given county. We treat the prefecture-level wastewater dumping as an exogenous variable at the county level because a county takes the total amount of discharged wastewater as given.<sup>4</sup>

Precipitation is one of the most important factors that influence surface water quality. Where rainfall is heavy, surface water quality is typically good. Precipitation affects surface water quality through two primary channels. First, rain directly dilute the concentration of water pollutants, and thus improves water quality. Second, rain causes the river to flow faster, which carries away water pollutant quickly and makes the river less prone to eutrophication (Zhong et al. 2005).

China has seven major river basins: Changjiang (Yangtze) River Basin, Huang River Basin, Zhujiang River Basin, Huai River Basin, Songhuajiang River Basin, Liao River Basin and Hai River Basin. Surface water quality in the southern river basins is better than that in the northern river basins because southern China has more rain. In our sample, the average yearly precipitation in the northern provinces (Beijing, Tianjin, Hebei, Liaoning and Ningxia) is 458 mm, with a standard deviation of 148 mm. In contrast, the average precipitation in the rest of the provinces is 1,286 mm with a standard deviation of 417 mm.

In Figure 1, we show the proportion of the river segments that were seriously polluted (Type IV, V, and VI) from 1991–2005. Roughly 60 to 70 percent of rivers in northern China

---

<sup>4</sup> A prefecture usually includes dozens of counties. Ideally, we would like to use the total wastewater discharged in all other counties within the same prefecture as the instrument variable for a particular county. Unfortunately, we cannot obtain county-level wastewater data.

(Huai River, Hai River, Songhuajiang River, Liao River) were severely polluted. In contrast, in southern river basins, such as the Changjiang River basin and Zhujiang River basin, the proportion of polluted river segments was much lower.

Although precipitation has a large effect on surface water quality, for it to be a valid instrument, it must affect surface water quality but not be correlated with the error term in our infant mortality rate equation. Our primary concern is that rainfall fluctuations may affect infant health through other channels besides surface water effects. For example, some studies have argued that an increase in rainfall increases agricultural production, lowering food prices, and increasing nutrient intake and hence health.<sup>5</sup> Although this argument sounds plausible, it is unlikely to be a problem in our study for two reasons.

First, small shocks in agricultural production induced by rainfall variations do not cause infant to die, unless the households are extremely poor and heavily depend on the food or the resulting income to survive. Recent studies that investigate the relationship between rainfall and health outcomes found a mixture of positive, negative, and non-significant results.<sup>6</sup> Studies that found a significant effect of rainfall fluctuations (within its usual range) on the infant mortality rate always focused on poor and arid or semiarid regions. In contrast, all the counties in our sample are located in relatively water-abundant areas, where the irrigation process mainly relies on the surface water system instead of rainfall. So it is unlikely that rainfall affects the infant mortality rate through its impact on agricultural production.

---

<sup>5</sup> The link between rainfall and agricultural income has been investigated in the literature. For example, Levine and Yang (2006) showed that more rainfall increases rice output in Indonesia. Duflo and Udry (2004) looked at how men and women's income and spending change when the yields of different crops vary due to their different sensitivity to rainfall.

<sup>6</sup> The details of these studies are in appendix Table A2.

When we regress agriculture production per capita on precipitation, the estimated coefficient on precipitation is -24.5 (a 100 mm increase in precipitation is associated with a 24.5 Yuan decrease in per capital agricultural production) with a *t*-statistic of  $-0.99$ .<sup>7</sup> Thus, we conclude that rainfall is not correlated with agriculture production in our sample.

Second, even in arid and semiarid regions, where water scarcity reduces agricultural production, the primary channel by which rainfall affects the infant mortality rate is not through its impact on agriculture production. For example, Bhalotra (2010) found that in rural areas of India income shocks have significant negative effects on the infant mortality rate. However, when he controlled for rainfall, the income effect did not change, suggesting that the effect of aggregate income on rural the infant mortality rate is not driven by agriculture income. In Rocha and Soares's (2012) study, irrespective of how they introduced agricultural production in the regression, the impact of rainfall variations on health at birth was not affected, also suggesting that agricultural income does not affect the infant mortality rate. Instead, they found that the negative impact of rainfall on the infant mortality rate would be greatly reduced by using piped water. Thus if people no longer use contaminated water, rainfall itself would not harm infant health.

Ebenstein (2012) investigated the effects of water pollution on digestive cancer in China and found that precipitation has only a very weak relationship with other disease mortality rates except for digestive and lung cancer rates, and found almost no relationship between rainfall and cancer rates in areas with high rate of tap water. His results also showed that the effect of rainfall variations on people's health is primary due to its impact on the surface water.

---

<sup>7</sup> Regression results are reported in online appendix Table A3.

In this study, if we include per capita agriculture production as a control variable, we find that it is not statistically significant and that the estimates of the effect of water quality on the infant mortality rate are unchanged. Thus, we believe that precipitation is a valid instrument and that rainfall affects infant health through its impact on surface water quality. We also verified that in 1999 and 2000 no catastrophic natural disaster such as severe droughts or floods occurred in the sampled counties.

We use a two-step sample selection model to estimate the infant mortality rate conditional on surface water quality. In the first step, we estimate the water quality using an ordered probit model. If people move in response to the local water quality, the water pollution level is endogenous.

Because the number of observations of Type I water is relatively small (Table 1), we aggregate Type I and Type II water into a single group, which leaves us with five water quality categories. Each County  $i$  has water quality in Category 1 (Type I and Type II), Category 2 (Type III), Category 3 (Type IV), Category 4 (Type V), or Category 5 (Type VI), ranging from good quality to bad. We estimate the water quality using an ordered-probit:

$$water_i^* = \alpha' \mathbf{z}_i + u_i \quad (1)$$

$$water_i = \begin{cases} 1 & \text{if } -\infty < water_i^* \leq \mu_1, \\ 2 & \text{if } \mu_1 < water_i^* \leq \mu_2, \\ 3 & \text{if } \mu_2 < water_i^* \leq \mu_3, \\ 4 & \text{if } \mu_3 < water_i^* \leq \mu_4, \\ 5 & \text{if } \mu_4 < water_i^* < \infty \end{cases} \quad (2)$$

where  $water_i^*$  is the unobserved latent selection variable (actual water quality),  $\mathbf{z}_i$  are a set of variables that affect water quality,  $u_i$  is a normal disturbance;  $water_i$  is the observed 5-degree water quality scale, and the unobserved cutoffs satisfy  $\mu_1 < \mu_2 < \mu_3 < \mu_4$ .



The infant mortality rate  $IMR_i$  is a linear function of the independent variables  $\mathbf{x}_i$ , the demographic, social-economic variables, conditional on the water quality level. We estimate separate coefficients of  $\mathbf{x}_i$  for each category  $water_i$ :

$$IMR_i = \begin{cases} \beta'_1 \mathbf{x}_i + \epsilon_{i1} & \text{if } water_i = 1, \\ \beta'_2 \mathbf{x}_i + \epsilon_{i2} & \text{if } water_i = 2, \\ \beta'_3 \mathbf{x}_i + \epsilon_{i3} & \text{if } water_i = 3, \\ \beta'_4 \mathbf{x}_i + \epsilon_{i4} & \text{if } water_i = 4, \\ \beta'_5 \mathbf{x}_i + \epsilon_{i5} & \text{if } water_i = 5 \end{cases} \quad (3)$$

where for each water quality category  $j$ ,  $\epsilon_{ij}$  has mean 0 and variance  $\sigma_j^2$ , and is bivariate normal with  $u_i$ . The correlation between  $\epsilon_{ij}$  and  $u_i$  is  $\rho_j$  for group  $j$ . We assume that  $\epsilon_{ij}$  and  $u_i$  are independently and identically distributed across observations.

We estimate this model using a two-step estimation procedure (Greene 2002) which is a generalization of Heckman's (1979) binary-model estimator.<sup>8</sup> Define:

$$\begin{aligned} \lambda_i \equiv E(u_i | water_i, \mathbf{z}_i) &= \frac{\int_{\mu_j}^{\mu_{j+1}} (water_i^* - \alpha' \mathbf{z}_i) \phi(water_i^* - \alpha' \mathbf{z}_i) dwater_i^*}{\Phi(\mu_{j+1} - \alpha' \mathbf{z}_i) - \Phi(\mu_j - \alpha' \mathbf{z}_i)} \\ &= \frac{\phi(\mu_j - \alpha' \mathbf{z}_i) - \phi(\mu_{j+1} - \alpha' \mathbf{z}_i)}{\Phi(\mu_{j+1} - \alpha' \mathbf{z}_i) - \Phi(\mu_j - \alpha' \mathbf{z}_i)} \end{aligned} \quad (4)$$

---

<sup>8</sup> There are two popular approaches in estimating the probit selection model: the full information maximum likelihood (FIML) approach and the two-step approach. In a binary selection case, Puhani (2000) found that FIML is usually more efficient than the two-step estimator. However, in an ordered probit selection model, such as we use, Chiburis and Lokshin (2007) found that the two-step estimator is more robust and is the better choice for almost all practical applications.

where  $\phi$  is the standard normal density function, and  $\Phi$  is the standard normal cumulative distribution function. Then the expectation of infant mortality rate, conditional on all the observed factors, is

$$E[IMR_i | water_i, \mathbf{z}_i, \mathbf{x}_i] = \beta'_j \mathbf{x}_i + E(\epsilon_{ij} | water_i = j, \mathbf{z}_i) = \beta'_j \mathbf{x}_i + \rho_j \sigma_j \lambda_i \quad (5)$$

Thus, if we only regress  $IMR_i$  on  $\mathbf{x}_i$  over the subsample  $\{i: water_i = j\}$ , without taking into account of  $\lambda$ , the estimation will be inconsistent if  $\rho_j \neq 0$ .

For the ordered-probit selection model to be identified,  $\mathbf{z}$  must contain at least one variable that is not in  $\mathbf{x}$ . That is, we must have at least one instrument  $z$  for the selection variable *water* (observed water quality) that is a significant determinant of water quality yet satisfies the exclusion restriction  $\text{Cov}(z, \epsilon_j) = 0$  for all  $j$ . We use wastewater dumping, rainfall, and their squares as instruments for water quality level.

In the first step, we estimate (2) by an ordered probit of *water* on  $\mathbf{z}$ , yielding the consistent estimates  $\hat{\alpha}$  and  $\hat{\mu}_j$ . Define  $\widehat{water}_i^* = \hat{\alpha}' \mathbf{z}_i$ . Using (4), we consistently estimate  $\lambda_i$  by

$$\hat{\lambda}_i = \frac{\phi(\hat{\mu}_j - \widehat{water}_i^*) - \phi(\hat{\mu}_{j+1} - \widehat{water}_i^*)}{\Phi(\hat{\mu}_{j+1} - \widehat{water}_i^*) - \Phi(\hat{\mu}_j - \widehat{water}_i^*)} \quad (6)$$

for  $j = water_i$ .

By using the observations  $i$  for which  $j = z_i$ , an OLS regression of  $IMR$  on  $\mathbf{x}$  and  $\hat{\lambda}$  provides a consistent estimate of  $\beta'_i$ .

Moreover,  $\sigma_j$  can be estimated by

$$\begin{aligned} \hat{\sigma}_j &\equiv \frac{1}{n_j} \left( RSS_j - \hat{C}_j^2 \sum_{i:j=j} \frac{\partial \hat{\lambda}_i}{\partial \widehat{water}_i^*} \right) \\ &= \frac{RSS_j}{n_j} - \frac{\hat{C}_j^2}{n_j} \left\{ \frac{((\hat{\mu}_j - \widehat{water}_i^*)\phi(\hat{\mu}_j - \widehat{water}_i^*) - (\hat{\mu}_{j+1} - \widehat{water}_i^*)\phi(\hat{\mu}_{j+1} - \widehat{water}_i^*))}{\Phi(\hat{\mu}_{j+1} - \widehat{water}_i^*) - \Phi(\hat{\mu}_j - \widehat{water}_i^*)} \right\} \end{aligned}$$

(7)

where  $n_j$  is the number of observations in which equation  $j$  is observed,  $\hat{C}_j$  is the coefficient on  $\hat{\lambda}$ , and  $RSS_j$  is the residual sum of squares for the regression. Because  $\hat{C}_j$  is a consistent estimator of  $\rho_j\sigma_j$ , we have a consistent estimator for  $\rho_j$ :

$$\hat{\rho}_j \equiv \frac{\hat{C}_j}{\hat{\sigma}_j}. \quad (8)$$

## VI. Estimations

We report the regression results in Table 4. In the first step, we regress water quality on the explanatory variables  $\mathbf{z}_i$ , which include our four instrument variables (precipitation, wastewater dumping and their squares) and all the demographic, social-economic  $\mathbf{x}_i$ . The estimated coefficients of the four instrument variables are all statistically significant. Both precipitation and wastewater dumping have strong effects on surface water quality.<sup>9</sup>

In the second step, we estimate the relationship between  $IMR_i$  and the independent variables  $\mathbf{x}_i$ , taking into account the first-step estimates  $\hat{\lambda}$ .

The estimated coefficients of  $\mathbf{x}_i$  vary across different groups. The percentage of non-agricultural population is statistically significant in Type I or II, Type IV and Type V areas. It is negatively correlated with the infant mortality rate, as expected. On average, a 10 percent increase in non-agricultural population is associated with roughly a 1.9 per thousand drop in the infant mortality rate in Type I or II areas. The estimated coefficients for Type IV and V areas are 1.6 and 2.4 per thousand, respectively.

---

<sup>9</sup> The relationship between water quality and the instrumental variables are similar if we treat water quality as a continuous variable and estimate the relationship using OLS. See appendix Table A4 for the regressions results.

A higher illiteracy rate statistically significantly increases the infant mortality rate in all regressions. As the illiteracy rate increases by 1 percent, the infant mortality rate falls by roughly 1 per thousand.

Housing and living conditions are typically negatively correlated with the infant mortality rate. Both the average number of rooms per household and the per capita housing area are statistically significantly associated with the infant mortality rate. For example, if each household in Type I or II areas has one more room at home, the infant mortality rate will fall by 3.8. If per capita housing area in Type I or II areas increases by 10 square meters, the infant mortality rate will decrease by 4.0 per thousand. As per capita GDP goes up, the infant mortality rate goes down. An increase of one thousand Yuan (about 150 US dollars) in per capita GDP is associated with a 0.17 to 0.64 per thousand fall in the infant mortality rate, depending on which group is chosen.

Given that the sample selection terms ( $\lambda_i$ ) are statistically significant in the Type IV and Type VI equations, if we did not explicitly take sample selection into account by using OLS, our estimates of the effects of water pollution on the infant mortality rate would be biased.

We can predict the expected infant mortality rate  $\overline{IMR}_i = \hat{\beta}'_j x_i$  for each category, and calculate the averages and differences of the predicted the infant mortality rate for each category. Using the estimates in Table 4, we find the average predicted the infant mortality rate  $\bar{y}$  for the five types of water are respectively 20.3, 24.0, 14.9, 12.1, and 5.8 per thousand. That is, the highest the infant mortality rate is associated with Type III water. In the cleanest areas (Type I or II) and most polluted areas (Type IV V and VI), the infant mortality rate is lower. The relationship between water quality and the infant mortality rate is non-monotonic, and the most polluted areas (Type VI) have the lowest the infant mortality rate. On average, the infant

mortality rate in the Type I or II areas is 3.7 per thousand lower than that in the Type III areas; and the infant mortality rate in the Types IV, V and VI areas are respectively 9.1, 11.9, and 18.2 per thousand lower than that in the Type III areas.

Our goal is to estimate the effects of water quality on the infant mortality rate. An alternative, perhaps better, method is to calculate the counterfactual  $\widehat{IMR}_j$  for equation  $j$ , if all observations were to switch to category  $j$ . Specifically, we predict:

$$\widehat{IMR}_j = \hat{\beta}'_j \mathbf{x}_i + \hat{\rho}_j \hat{\sigma}_j \hat{\lambda}_i \quad (9)$$

where  $\hat{\lambda}_i$  is calculated as in (6), using the actual  $water_i$ .

The counterfactual  $\widehat{IMR}_j$  tells us what the infant mortality rate would be if all observations were to switch to different levels of water pollution. We report the five predicted counterfactual average the infant mortality rate in Row (1) of Table 5. If we switch all observations to the Type I or II area, the average the infant mortality rate would be 18.7 per thousand. The corresponding mortality rates are 24.8 per thousand if we switch all observations to the Type III area, the infant mortality rate 16.9 per thousand for the Type IV area, 15.1 for the Type V area, and 4.6 for the Type VI area.

We still use the Type III areas as the reference group, and compare its the infant mortality rate with that in other groups. The differences in Row (2) in Table 5 show the estimated effects of water quality on the infant mortality rate in the selection model. If water quality changes from Type I or II to Type III, the infant mortality rate will increase by about 6.0 per thousand. Changing water quality from Type III to Type IV decreases the infant mortality rate by 7.9 per thousand. The most polluted areas (Type VI) have the lowest the infant mortality rate rates. If water quality deteriorates from Type III to Type VI, the infant mortality rate drops by 20.2 per thousand.

We also estimated the ordered probit selection model by gender. We calculated the counterfactual average the infant mortality rate for each category and the differences between them, using Type III areas as the reference group. The results for male infants are reported in Row (4) and (5) in Table 5. Changing water quality from Type I or II to Type III would increase male the infant mortality rate by roughly 9.7 per thousand. As water quality deteriorates from Type III to Type IV, V and VI, male the infant mortality rate will drop by 7.1, 8.1 and 15.1 per thousand, respectively. The estimates for female infants are reported in Row (7) and (8) in Table 5. The results are slightly different. Changing water quality from Type I or II to Type III would not significantly increase female the infant mortality rate: The difference is only about 1.1 per thousand. However, as water quality becomes more polluted, changing from Type III to Type IV, V and VI, female the infant mortality rate would drop dramatically, with the respective estimated magnitudes of  $-8.8$ ,  $-11.6$ , and  $-26.3$  per thousand. The estimate of the effect of water pollution on the female infant mortality rate is greater than that for the male infant mortality rate. Female infants benefit more than males from water quality falling from fair to polluted.

We compared our sample-selection model results to an OLS regression model:

$$IMR_i = \beta_0 + \beta_1 D1_i + \beta_2 D2_i + \beta_3 D3_i + \beta_4 D4_i + \beta_5' \mathbf{x}_i + \varepsilon_i, \quad (10)$$

where  $IMR_i$  is the infant mortality rate in county  $i$ ,  $\mathbf{x}_i$  is a vector of covariates,  $D1$  is a dummy equal to 1 is a county has *Type I* or *II* water,  $D2 = 1$  if it has *Type IV* water,  $D3 = 1$  if it has *Type V* water, and  $D4 = 1$  if it has *Type VI* water. Type III water is the reference group.

The regression results from OLS are reported in Table 6. The relationship between water quality and the infant mortality rate is non-monotonic in the OLS regressions. Based on the OLS estimates, changing water quality from Type I or II to Type III is associated with 4.0 per thousand drop in the infant mortality rate; and as water quality deteriorates from Type III to Type

IV, V, and VI, the associated the infant mortality rate would drop by 5.7, 7.7 and 8.0 per thousand, respectively.

The less flexible OLS model underestimates the effects of water pollution on the infant mortality rate, especially for the more polluted areas. For example, as water quality deteriorates from Type III to Type VI, the OLS estimates indicate the infant mortality rate would decrease by only 8 per thousand, whereas the ordered profit selection model predicts it would drop by 20 per thousand.

Assuming that people cannot perceive water quality changes between Type I or II and Type III, the associated 6.0 per thousand increase in the infant mortality rate can be interpreted as the pure health effect of water pollution. In other words, we find that a one-degree deterioration of water quality leads to about 30% increase in the infant mortality rate. Is this result plausible? To answer this question, we conduct a crude calculation on the magnitude of water quality changes from Type I or II to Type III.

According to the Environmental Surface Water Quality Standard in China, a reduction in water quality from Type I to Type III approximately corresponds to (using Type I as reference) a 33-percent decrease in the concentration of dissolved oxygen (DO), a 200-percent increase in the concentration of Potassium permanganate (KMnO<sub>4</sub>), a 33-percent increase in the concentration of chemical oxygen demand (COD), a 560-percent increase in the concentration of Ammonia Nitrogen (NH<sub>3</sub>-N), a 900-percent increase in the concentration of Total phosphorus (TP), a 400-percent increase in the concentration of Total nitrogen (TN). The changes in the concentrations for other pollutants are even more pronounced. For example, the maximum allowed number of fecal coliform for Type I water is 200 per liter, and it is 2,000 per liter for Type II water, and this maximum number increases to 10,000 per liter for Type III water. Thus, the actual change in

water quality from Type I to Type II to Type III, even though cannot be visually detected, is huge.

The large effect of water pollution on the Chinese infant mortality rate is consistent with the findings of studies covering other countries. Cutler and Miller (2005) argued that the adoption of clean water technologies such as filtration and chlorination was responsible for up to a 75 percent reduction in the infant mortality rate in early twentieth century America. Galiani et al. (2005) found that privatization of water supply in Argentina reduced the mortality of children under age 5 by 26 percent. Brainerd and Menon (2012) found that a 10 percent increase in the average of fertilizer chemicals in water in the month of conception increased the infant mortality rate by 4 percent, and neo-natal mortality by 7 percent.

Our study has two data limitations. We lack data on non-water pollution by county and on alternative sources of drinking water. However, we believe that our key qualitative result of a non-monotonic relationship between surface water and the infant mortality would hold even if such data were available.

First, if other types of pollution are correlated with water pollution, our estimated effects for surface water may reflect the combined effects of exposure to a range of environmental pollutants and risks. However, other types of pollution, such as air pollution and toxin accumulation, usually have a monotonic effect on health. While people can protect themselves from being harmed by polluted water at relatively low cost (such as by boiling the water or using ground water), avoiding the harms of polluted air and many other types of pollution can be costly and less effective. Given that water pollution is positively correlated with other types of pollution, the non-monotonic relationship we find between polluted water and infant mortality is unlikely to be a result from other types of pollution.



Second, due to a lack of detailed data on alternative water sources such as ground, tap, and bottled water at the county level, we are unable to explore the exact channels through which people can mitigate the harms of polluted surface water. If detailed information about alternative water sources were available, we could estimate how people substitute between surface water and the various more expensive alternative sources of water. However, lacking such data, our study shows that people substitute away from extremely dangerous surface water—we just do not know to which alternative. If they did not substitute, the death rates would be very high.

## **VII. Conclusions**

We show that the relationship between surface water quality and the infant mortality rate is non-monotonic in China. As surface water quality deteriorates, the infant mortality rate first increases and then decreases. The infant mortality rate is the highest in areas where surface water quality is fair. This finding is robust to a variety of specifications and models.

Our explanation is that, as surface water deteriorates from a good level, people do not detect a quality change and continue to consume the water, so more infants die. As the water pollution increase more, the low quality becomes obvious, so people reduce their usage of polluted water, and more infants survive.

We find strong evidence supporting this argument. Both the OLS and the ordered-probit selection models show that the infant mortality rate is highest when water quality is fair. In regions with cleaner or more polluted water, the infant mortality rate is lower. The infant mortality rate is the lowest in regions with the most polluted surface water (Type V and Type VI), suggesting that avoidance behavior significantly mitigates the health risks from water pollution.

An important policy implication of this study is that the Chinese government should intervene in regions where the surface water is moderately polluted. The government could provide health information or, even better, provide safe tap water, to those regions in which public awareness of water pollution is low but health risks are high.

## References

- Basu A.M., Stephenson R. 2005. “Low Levels of Maternal Education and the Proximate Determinants of Childhood Mortality: a Little Learning Is Not a Dangerous Thing.” *Social Science & Medicine*, 60(9): 2011–2023.
- Behrman, J. R., Hoddinott, J., Maluccio, J. A., Quisumbing, A., Martorell, R., and Stein, A. D. 2003. “The Impact of Experimental Nutritional Interventions on Education into Adulthood in Rural Guatemala: Preliminary Longitudinal Analysis.” *Filadelfia, Washington, Atlanta: University of Pennsylvania, IFPRI, Emory*.
- Bhalotra, S. 2010. “Fatal Fluctuations? Cyclicity in Infant Mortality in India.” *Journal of Development Economics*, 93(1), 7-19.
- Boyle M.H. 2006. “The Influence of Economic Development Level, Household Wealth and Maternal Education on Child Health in the Developing World.” *Social Science & Medicine*, 63(8): 2242–2254.
- Brainerd, E., and N., Menon. 2012. “Seasonal Effects of Water Quality on Infant and Child Health in India.” *IZA Discussion Paper*, No. 6559.
- Brennan M.E. and Lancashire R. 1978. “Association of Childhood Mortality with Housing Status and Unemployment.” *Journal of Epidemiology & Community Health*, 32(1): 28–33.
- Cantor, K. P. 1997. “Drinking Water and Cancer.” *Cancer Causes and Control*, 8(1): 292–308.
- Chen, K., Yu W., Ma X. et al. 2005. “The Association between Drinking Water Source and Colorectal Cancer Incidence in Jiashan County of China: a Prospective Cohort Study.” *European Journal of Public Health*, 15(6): 652–656.
- Chiburis, R. and Loxsin, M. 2007. “Maximum Likelihood and Two-Step Estimation of an Ordered-Probit Selection Model.” *The Stata Journal*, 7(2):167–182.

- Cowell, A.J.O. 2006. "The Relationship Between Education and Health Behavior: Some Empirical Evidence." *Health Economics*, 15(2):125–146.
- Cutler, D. and Lleras-Muney, A. 2006. "Education and Health: Evaluating Theories and Evidence." *NBER Working Paper 12352*. Cambridge, MA, National Bureau of Economic Research.
- Cutler, D. and Miller, G. 2005. "The Role of Public Health Improvements in Health Advances: the 20<sup>th</sup> Century United States." *Demography*, 42(1): 1–22
- Currie, J., Graff Zivin, J., Meckel, K., Neidell, M., and Schlenker, W. 2013. "Something in the Water: Contaminated Drinking Water and Infant Health." *Canadian Journal of Economics/Revue Canadienne D'économique*, 46(3), 791-810.
- Davis, M. L. and Masten, S. J. 2004. *Principles of Environmental Engineering and Science*, McGraw-Hill. 2004.
- Delgado, J., Ramirez-Cardich, M. E., Gilman, R. H., Lavarello, R., Dahodwala, N., Bazán, A., Rodríguez, V., Cama, R.I., Tovar. M., and Lescano, A. 2002. "Risk Factors for Burns in Children: Crowding, Poverty, and Poor Maternal Education." *Injury Prevention*, 8(1), 38-41.
- Duflo, E. and Udry, C. 2004. "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts, and Consumption Choices." *NBER Working Paper*, No.10498.
- Ebensitein, A.Y. 2012. "The Consequences of Industrialization: Evidence from Water Pollution and Digestive Cancers in China." *The Review of Economics and Statistics*, 94(1): 186–201.

- Freedman, D. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology, American Sociological Association*, vol. (21), pp. 291–313.
- Frost, M.B., Forste, R., and Haas, D.W. 2005. "Maternal Education and Child Nutritional Status in Bolivia: Finding the Links." *Social Science & Medicine*, 60(2): 395–407.
- Galiani, S., Gertler, P., and Schargrodsky, E. 2005. "Water for Life: the Impact of privatization of Water Services on Child Mortality." *Journal of Political Economy*, 113(1): 83–120
- Greene, W.H. 2002. "LIMDEP Version 8.0 Econometric Modeling Guide." Plain-view, NY: Econometric Software.
- Greenstone, M., and Hanna, R. 2014. "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." *The American Economic Review*, 104(10), 3038-3072.
- Heckman, J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47: 153–162.
- Jalan, J., and Ravallion, J. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity-Score Matching." *Journal of Business & Economic Statistics, American Statistical Association*, 21(1): 19–30.
- Lavy, V., J. Strauss, D. Thomas, and P. de Vreyer. 1996. "Quality of Health Care, Survival and Health Outcomes in Ghana." *Journal of Health Economics*, 15: 33–57.
- Lee, L., Rosenzweig, M. R., and Pitt, M. M. 1997. "The Effects of Improved Nutrition, Sanitation, and Water Quality on Child Health in High-Mortality Populations." *Journal of Econometrics*, 77: 209–35.
- Levine, D. and D. Yang. 2006. "A Note on the Impact of Local Rainfall on Rice Output in Indonesian Districts." mimeo.

- Mansfield, C., Johnson, F. R., and Van Houtven, G. 2006. "The Missing Piece: Valuing Averting Behavior for Children's Ozone Exposures." *Resource and Energy Economics*, 28(3), 215-228.
- Martin, A.E. 1967. "Environment, Housing and Health." *Urban Studies*, 4(1): 1–21.
- Merrick, T. W. 1985. "The Effects of Piped Water on Early Childhood Mortality in Urban Brazil, 1970 to 1976", *Demography*, 22(1): 1–24.
- Miguel E. 2005. "Health, Education, and Economic Development." In: Lopez-Casasnovas G, Rivera B, Currais L, eds, *Health and Economic Growth*. Cambridge, MA, MIT University Press.
- Moretti, E., and Neidell, M. 2011. "Pollution, Health, and Avoidance Behavior Evidence from the Ports of Los Angeles." *Journal of Human Resources*, 46(1), 154-175.
- Neidell, M. 2009. "Information, Avoidance Behavior, and Health: The Effect of Ozone on Asthma Hospitalizations." *Journal of Human Resources*, 44(2): 450–78.
- Puhani, P. A. 2000. "The Heckman Correction for Sample Selection and its Critique." *Journal of Economic Surveys*, 14: 53–68.
- Shimshack, J., Ward, M., and Beatty, T. 2007. "Mercury Advisories: Information, Education, and Fish Consumption." *Journal of Environmental Economics and Management*, 53(2): 158-179.
- Van de Poel, E., O'Donnell, O., and Van Doorslaer, E. 2009. "Urbanization and the Spread of Diseases of Affluence in China." *Economics & Human Biology*, 7(2): 200–216.
- Victora et al. 1988. "Water Supply, Sanitation and Housing in Relation to the Risk of Infant Mortality From Diarrhea." *International Journal of Epidemiology*, 17(3): 651–654.

World Bank. 2006. "Water Quality Management Policy and Institutional Considerations."

*Discussion Paper*, The World Bank, Washington, D.C.

Chenghua, Z., Zhiguo, X., Wenqian, Z., Derui, W., Chunguang, D., Yongjian, L., and Mei, X.

2005. "Eutrophication Investigation and Assessment of the Daning River after Water Storage of the Three Gorges Reservoir." *Chinese Journal of Geochemistry*, 24(2), 149-154.

Zivin, J. G., M. Neidell, and W. Schlenker. 2011. "Water Quality Violations and Avoidance

Behavior: Evidence from Bottled Water Consumption." *American Economic Review: Papers and Proceedings*, 101:3, 448–453.

**Table 1. Water Quality Distribution in the Sample**

Water Quality	Type I	Type II	Type III	Type IV	Type V	Type VI	Total
Frequency	23	132	90	60	59	97	461
Percent	4.99	28.63	19.52	13.02	12.8	21.04	100

Sources.-China Water Quality Yearbooks(2000).



**Table 2. Chinese Infant Mortality Rates (per thousand)**

	Mean	Std. Dev.	25% Quantile	50% Quantile	75% Quantile
All	19.23	15.82	8.64	14.62	25.31
Male	16.97	13.44	7.89	13.63	21.81
Female	21.97	20.42	8.61	15.17	29.47

Source.-China 2000 Census.

**Table 3. Summary Statistics of the Explanatory Variables**

Variable	Mean	Std. Dev.	Min	Max
Precipitation (100 mm)	11.08	5.07	1.15	24.60
Wastewater Dumping	19.29	11.34	0.42	47.38
Non-Agricultural Population (%)	30.16	26.55	4.38	97.40
Illiteracy Rate	9.52	6.15	1.57	48.34
Rooms per Household	2.60	0.58	1.50	4.92
Housing Area per capita (sq.m.)	23.47	6.48	9.47	45.64
GDP per capita (Yuan)	9.25	8.20	0.82	55.84
Government Expenditure per capita	0.78	0.93	0.02	9.49
Hospital Beds per 10,000 people	32.32	23.39	5.98	161.58

Sources.-China 2000 Census and China Provincial Statistical Yearbooks (2000).

**Table 4. Infant Mortality Rate Ordered-Probit Selection Model**

First Step (Water Quality)		Second Step (Infant Mortality Rate by Water Type)				
		I or II	III	IV	V	VI
Precipitation (100 mm)	-0.271*** (0.05)	-	-	-	-	-
Precipitation Sq.	0.006*** (0.00)	-	-	-	-	-
Dumping	0.033** (0.02)	-	-	-	-	-
Dumping Sq.	-0.001*** (0.00)	-	-	-	-	-
Non-Ag Pop (%)	0.006** (0.00)	-0.185*** (0.07)	-0.007 (0.11)	-0.164* (0.09)	-0.243*** (0.07)	-0.023 (0.06)
Illiteracy (%)	0.022** (0.01)	0.925*** (0.20)	1.330*** (0.35)	1.053*** (0.28)	1.296*** (0.37)	1.000*** (0.18)
Rooms	0.236** (0.11)	-3.813* (2.31)	1.005 (3.12)	0.251 (3.93)	-5.964** (2.60)	-0.734 (2.06)
Housing Area	-0.008 (0.01)	-0.401*** (0.15)	-0.597** (0.28)	-0.580* (0.31)	-0.586** (0.26)	-0.33 (0.31)
GDP pc (1,000 Yuan)	0.021* (0.01)	-0.323 (0.21)	-0.644* (0.37)	-0.26 (0.33)	-0.169 (0.14)	-0.463* (0.28)
Gov. Exp. pc	0.004 (0.08)	-2.223 (2.45)	2.134 (5.50)	-1.585 (2.64)	-0.934 (0.97)	-0.872 (2.07)
Hospital Beds (per 10,000)	-0.002 (0.00)	0.04 (0.05)	-0.025 (0.10)	0.204 (0.14)	0.224** (0.11)	0.008 (0.07)
$\lambda_i$	-	-0.994 (3.04)	-0.106 (2.94)	9.552*** (3.40)	3.309 (2.06)	10.20*** (2.78)
Observations	460	$\rho_0=-0.09$	$\rho_1=-0.01$	$\rho_2=0.66$	$\rho_4=0.37$	$\rho_5=0.79$

Note.-The relationship between infant mortality and surface water quality is estimated by an ordered-probit sample-selection process using a two-step consistent estimator. Standard errors are reported in parentheses.

\*\*\* Significant at 1% level.

\*\* Significant at 5% level.

\* Significant at 10% level.

**Table 5. The Effects of Water Quality on the Infant Mortality Rate**

		Type I or II	Type III	Type IV	Type V	Type VI
Overall	(1) Counterfactual Aver.	18.73	24.76	16.89	15.09	4.57
	(2) Differences	-6.03		-7.86	-9.67	-20.19
	(3) OLS Estimates	-3.96		-5.68	-7.68	-7.96
Male	(4) Counterfactual Aver.	12.79	22.46	15.36	14.40	7.38
Infant	(5) Differences	-9.67		-7.10	-8.06	-15.08
	(6) OLS Estimates	-4.55		-5.30	-6.70	-6.02
Female	(7) Counterfactual Aver.	26.43	27.55	18.67	15.96	1.28
Infant	(8) Differences	-1.12		-8.87	-11.59	-26.26
	(9) OLS Estimates	-3.10		-6.19	-8.90	-10.28

Note.-This table compares the differences in infant mortality for each water quality type using estimates from both ordered probit selection model and OLS. Type III water is used as the reference group.

**Table 6. OLS Regression Results on the Infant Mortality Rate**

	Infant Mortality Rate		
	Overall	Male	Female
Type I or II	-3.964** (1.74)	-4.552*** (1.56)	-3.102 (2.26)
Type IV	-5.682*** (2.19)	-5.304*** (1.99)	-6.187** (2.73)
Type V	-7.683*** (1.98)	-6.699*** (1.66)	-8.895*** (2.69)
Type VI	-7.956*** (1.88)	-6.018*** (1.74)	-10.28*** (2.27)
Non-Agricultural Population (%)	-0.128*** (0.03)	-0.081*** (0.02)	-0.187*** (0.04)
Illiteracy Rate (%)	1.054*** (0.13)	1.026*** (0.11)	1.066*** (0.17)
Ave. Rooms per Household	-2.678** (1.02)	-1.883** (0.91)	-3.586*** (1.39)
Housing Area per capita	-0.356*** (0.08)	-0.166** (0.07)	-0.587*** (0.12)
GDP per capita (1000 Yuan)	-0.356*** (0.09)	-0.352*** (0.08)	-0.367*** (0.11)
Gov. Exp. per capita (1000 Yuan)	-0.339 (0.56)	-0.226 (0.51)	-0.439 (0.71)
Hospital Beds per 10,000 people	0.007 (0.03)	0.033 (0.02)	-0.023 (0.04)
F-Statistics	36.81	31.76	35.28
R <sup>2</sup>	0.49	0.48	0.43
Number of Observations	460	460	460

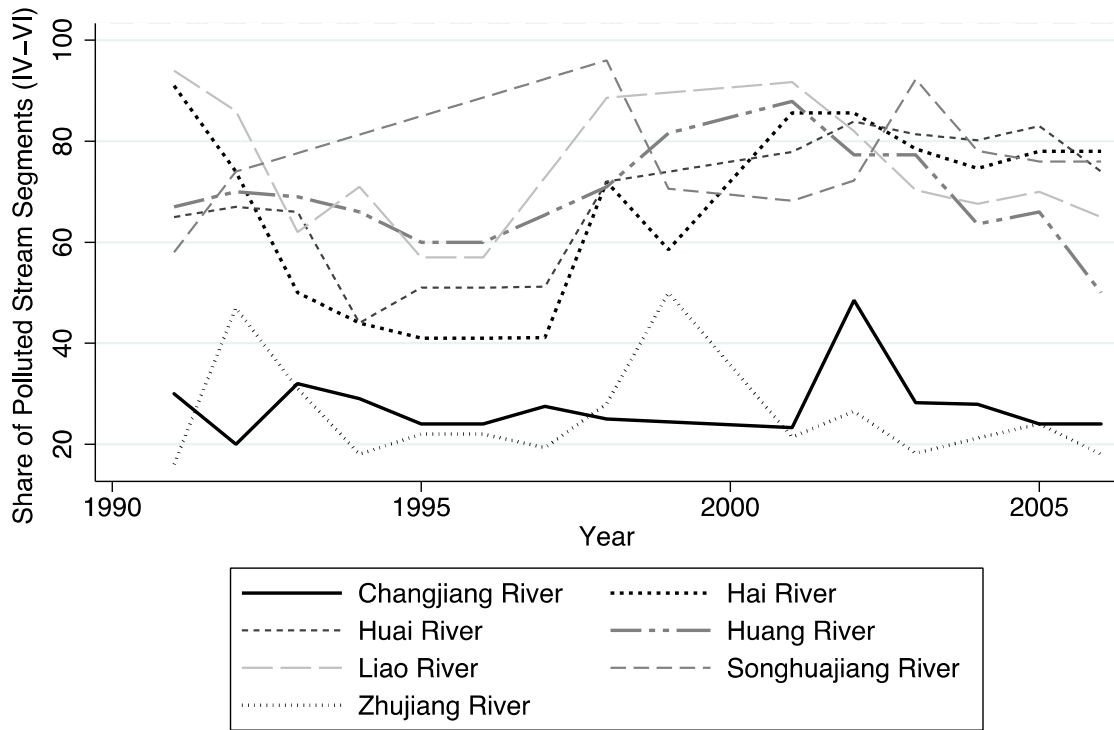
Note.-The relationship between infant mortality and surface water quality is estimated by a set of ordinary least square (OLS) regressions. Huber-White robust standard errors are reported in the parenthesis.

\*\*\* Significant at 1% level.

\*\* Significant at 5% level.

\* Significant at 10% level.

**Figure 1. Polluted Segment of the Main Water System in China**



## Online Appendix

**Table A1. Surface Water Environmental Quality Standard: Limit Values**

Classifications	I	II	III	IV	V
PH values			6-9		
Dissolved oxygen $\geq$	7.5	6	5	3	2
Hypermanganate index $\leq$	2	4	6	10	15
COD $\leq$	15	15	20	30	40
BOD5 $\leq$	3	3	4	6	10
NH3-N $\leq$	0.15	0.5	1	1.5	2
	0.02	0.1	0.2	0.3	0.4
Total phosphorus	(Lakes 0.01)	(Lakes 0.025)	(Lakes 0.05)	(Lakes 0.1)	(Lakes 0.2)
Total nitrogen $\leq$	0.2	0.5	1	1.5	2
Cu $\leq$	0.01	1	1	1	1
Zn $\leq$	0.05	1	1	2	2
Fluoride $\leq$	1	1	1	1.5	1.5
Se $\leq$	0.01	0.01	0.01	0.02	0.02
As $\leq$	0.05	0.05	0.05	0.1	0.1
Hg $\leq$	0.00005	0.00005	0.0001	0.001	0.001
Cd $\leq$	0.001	0.005	0.005	0.005	0.01
Hexavalent chrome $\leq$	0.01	0.05	0.05	0.05	0.1
Pb $\leq$	0.01	0.01	0.05	0.05	0.1
Cyanide $\leq$	0.005	0.05	0.2	0.2	0.2
Volatile hydroxybenzene $\leq$	0.002	0.002	0.005	0.01	0.1
Petroleum $\leq$	0.05	0.05	0.05	0.5	1
Anionic surface-active agent $\leq$	0.2	0.2	0.2	0.3	0.3
Sulfide $\leq$	0.05	0.1	0.05	0.5	1
Coliform group (pieces/L) $\leq$	200	2000	10000	20000	40000

Sources.-Standard GB3838-2002, State Environment Protection Administration of PR China and General Administration for Quality Supervision, Inspection and Quarantine.

**Table A2. The Effects of Rainfalls on the Infant Mortality Rate**

Study	Outcome Variables	Explanatory Variables	Data	Conclusions
Aguilar and Vicarelli (2011)	Cognitive tests, anthropometric variables, health indicators	Exogenous excessive rainfall shocks.	Individual survey data from 506 rural communities in Mexico	Children born in years and regions affected by excessive rain experience have slower anthropometric growth and cognitive development.
Baird et al. (2007)	Infant Mortality Rate	Rainfalls, per capita GDP, characteristics of women, conflict, quality of institution, etc.	Birth data in 59 developing countries	Rainfall shocks have no significant effect on infant mortality.
Friedman (2010)	Infant Mortality Rate	Daily temperature and rainfalls	Birth data in 47 South African countries	Excess rainfall is both detrimental and protective depending on the timing in which it occurs.
Kovats and Wilkinson (2004)	Mortality by Age and by Cause of Death	Daily rainfalls, temperature, and season, trend, holidays and air pollution.	Birth data in New Delhi	Any increase in rainfall increases the risk of infectious disease mortality in the near term.



**Table A2 (continued). The Effects of Rainfalls on the Infant Mortality Rate**

Study	Outcome Variables	Explanatory Variables	Data	Conclusions
Kim (2009)	Infant Mortality Rate	Monthly Rainfalls, mother's characteristics, child's characteristics, religion, region, etc.	Birth data in 9 West African countries	On average, rainfall and infant mortality are unrelated. But rainfall shocks have an adverse effect on the survival of young children who were born in the rainy season.
Kudamastu et al. (2012)	Infant Mortality Rate by Area, by Household Type	Monthly rainfalls, season, drought indicator, household type, area (endemic, epidemic, non-malarious, rainy and arid)	Birth data in 28 African countries	Increased rainfall is associated with higher mortality by malaria in epidemic areas, but not in endemic areas.
Rocha and Soares (2015)	Gestation, Birth Weight and Infant Mortality Rate	Monthly rainfall, drought indicator, temperature, trend, etc.	Birth data in Semiarid Northeast Brazil	Negative rainfall fluctuations lead to higher incidences of low birth weight, preterm gestation and infant mortality rates, in particular due to intestinal infections and malnutrition.
Skoufias et al. (2011)	Child height-for-age	Rainfall, household and individual characteristics.	Mexico Family Life Survey and National Nutrition Survey of Mexico	The effects of rainfall on height-for-age is heterogeneous. A positive rainfall shock during the wet season is associated with shorter children in the North, but not in the Centre/South regions.

**Table A3. Per capita Agricultural Production and Rainfalls**

	Agricultural Production per capita		
Precipitation	-24.5	56.9	72.8
(100 mm)	(16.5)	(49.6)	(74.7)
Precip. Sq.		-3.4	-4.6
		(2.4)	(3.6)
Non-Ag Pop.			13.3
(%)			(13.5)
Illiteracy			27.9
(%)			(18.9)
Rooms			1,340.6***
			(475.7)
Housing Area			-39.5
			(26.7)
GDP pc			69.5**
(1000 Yuan)			(32.7)
Gov. Exp. pc			-393.5**
			(195.8)
Hospital Beds			-17.1**
(per 10,000)			(8.5)
Observations	461	461	460
R-squared	0.0	0.0	0.1

Note.-The relationship between per capita agricultural production and precipitation is estimated by a set of ordinary least square (OLS) regressions. Huber-White robust standard errors are reported in the parenthesis.

\*\*\* Significant at 1% level.

\*\* Significant at 5% level.

\* Significant at 10% level.

**Table A4. Water Quality and IVs: OLS Estimates**

	Water Quality Type (Type I-VI)			
Precipitation (100 mm)	-0.359*** (0.044)		-0.339*** (0.045)	-0.332*** (0.051)
Precip. Sq.	0.009*** (0.002)		0.008*** (0.002)	0.008*** (0.002)
Dumping		0.020 (0.020)	0.023 (0.015)	0.030* (0.016)
Dumping Sq.		-0.001** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Non-Ag Pop. (%)				0.008** (0.004)
Illiteracy (%)				0.017 (0.017)
Rooms				0.239* (0.143)
Housing Area				-0.012 (0.012)
GDP pc (1000 Yuan)				0.024** (0.011)
Gov. Exp. pc				0.034 (0.092)
Hospital Beds (per 10,000)				-0.005* (0.003)
F-stats	103.57	12.49	62.63	24.7
Observations	461	461	461	460
R-squared	0.261	0.028	0.289	0.310

Note.-Huber-White robust standard errors are reported in the parenthesis.

\*\*\* Significant at 1% level.

\*\* Significant at 5% level.

\* Significant at 10% level.

## Appendix References

- Aguilar, A. and Vicarelli, M. 2011. “El Nino and Mexican Children: Medium-term Effects of Early-Life Weather Shocks on Cognitive and Health Outcomes.” Unpublished Paper,
- Baird, S., J. Friedman, and N. Schady. 2007. “Aggregate Income Shocks and Infant Mortality in the Developing World.” *Policy Research Working Paper*, World Bank, No. 4346.
- Friedman, J. 2011. “Climate Variability and Infant Mortality in Africa.” PopPov Research.
- Kovats, S., and Wilkinson, P. 2004. “Rainfall, Temperature and Mortality in New Delhi, India.” *Epidemiology*, 15(4), S96.
- Kim, Y. S. 2009. “The Impact of Rainfall on Early Child Health.” Unpublished paper.
- Kudamatsu, M., Persson, T., and Strömberg, D. 2012. “Weather and Infant mortality in Africa.” Working Paper.
- Rocha, R., and Soares, R. R. 2015. “Water Scarcity and Birth Outcomes in the Brazilian Semiarid.” *Journal of Development Economics*, 112, 72-91.
- Skoufias, E., Vinha, K., and Conroy, H. 2011. “The Impacts of Climate Variability on Welfare in Rural Mexico.” *World Bank Policy Research Working Paper*, No. 5555.
- Masayuki K., T. Persson, and D. Strömberg. “Weather and Infant Mortality in Africa”, Unpublished Paper, 2010.