Pollution Externalities and Health: A Study of Indian Rivers

Working Paper

Quy-Toan Do (The World Bank)
Shareen Joshi (Georgetown University)
Samuel Stolper (Harvard University)

July 3rd, 2014

ABSTRACT

Water pollution levels in many developing countries remain significantly higher than in the developed world. While such pollution is often a byproduct of economic activity, it also imparts a health burden on the population. We study this health burden in the context of domestic water pollution in India's rivers, focusing on infant mortality as a measure of health outcomes. In particular, we quantify two impacts: The mortality burden of river pollution in the district of its measurement; and the persistence of that burden in neighboring, downstream districts. To avoid endogeneity problems, we construct an instrument for water quality in a given Indian district using water quality *upstream* of that district. Two-stage least squares (2SLS) regression reveals a positive district-level association between one-month infant mortality and the concentration of fecal coliforms in river water. This association strongly holds for both national demographic surveys that we use to compile infant mortality data. We interpret the association to be causal: The average effect of a one-percent increase in fecal coliforms is an additional 3-5 deaths per 100,000 births in a given month. In comparison, the corresponding downstream infant mortality impact is approximately 1-2 deaths per 100,000 births.

Support from International Growth Centre – Central India, as well as Italy's Ministry for Environment, Land and Sea is gratefully acknowledged.

I. INTRODUCTION

River pollution imposes geographically-widespread costs: The burden of reduced water quality is felt not just by the community in the immediate vicinity of its production, but by a continuum of downstream communities as well. The socially-efficient level of water quality at a given point along a river therefore depends on the associated welfare impacts at both *that* point and all affected points downstream (Oates, 2001, Lipscomb and Mobarak, 2007).

We set out to quantify and compare such impacts in the context of Indian rivers. India is currently characterized by alarmingly high levels of domestic water pollution in its surface waters (Central Pollution Control Board, 2013) and expensive clean-up policy whose effects have been ambiguous (Greenstone and Hanna, 2012). This characterization, coupled with extensive data on water quality levels all over the country, creates an opportune setting for a study of water pollution and health. In particular, we focus on infant mortality, given the vulnerability of infants to toxins and the large loss of life represented by that measure.

Since a comparison of overall pollution and health outcomes in Indian districts is susceptible to omitted variable bias, we use the geography of rivers to isolate exogenous variation in water pollution. Our basic strategy is to instrument for water quality in a given district with water quality *upstream* of that district. The intuition for this instrument is that upstream "decisions" to pollute have downstream health consequences due to the flow of rivers but are uncorrelated with all other determinants of health.

We thus construct a two-stage least squares (2SLS) estimator for the relationship between domestic water quality and infant mortality. Our instrumental-variables results quantify the within-district externality – i.e., the association of infant mortality in a district with river water quality in that same district. Meanwhile, our reduced-form results quantify the persistence of water quality impacts in downstream districts. Our findings indicate that the pollution-mortality relationship is strongly positive and significant in the first month of life. Furthermore, this relationship holds even when pollution is compared with mortality in the next district downstream. The predictive, within-district effect of a one-percent increase in fecal coliforms is an additional 3-5 deaths per 100,000 births in a given month, or a 0.09-0.14% rise in the probability of neonatal (first-month) mortality. The corresponding effect in downstream districts

is 1-2 deaths, or a 0.03-0.09% rise the likelihood of neonatal mortality. Strikingly, these results hold for two entirely different samples of Indian children – those with mothers surveyed for the District-Level Household Survey II (DLHS-2; 2002-2005), and those with mothers surveyed for the Demographic and Health Survey II (DHS-2; 1998-2000).

Methodologically, we contribute a novel identification strategy that is replicable in any country where river water quality is monitored. Empirically, we provide three findings that contribute to the general understanding of river pollution's relationship with infant mortality (IM) in India. First, we quantify a mortality impact of acute, one-month shocks to domestic water quality. Second, we show that the first month of life is by far the most risky for infants when it comes to water quality and safety. Third, we confirm the persistence of the mortality burden in downstream communities, suggesting that the downstream IM externality is as much as one-third the size of the within-district externality. Taken together, the size and downstream persistence of water pollution's harm should be strong motivation for discussion of new policies to improve water quality and protect Indian infants.

II. EXISTING LITERATURE

Our research agenda is founded upon three assertions: First, that water pollution exposure is bad for health, and in particular for infant survival; second, that the current level of river water quality in India is socially inefficient; and third, that water pollution persists at dangerous levels even at significant distances downstream of its incidence.

The first assertion is backed by a long literature in epidemiology, starting with John Snow (1854). Snow was the first to correctly suggest that the cause of the "Broad Street cholera outbreak" in London was caused by leakage of sewage (in the form of fecal bacteria) into a public well. Since then, epidemiology research has advanced to produce evidence linking water pollution to a host of pathogens (E. coli, rotavirus, etc.) and illnesses (cholera, diarrhea, etc.) (Fewtrell and Bartram, 2001). Furthermore, it is not just *drinking* of contaminated water that conveys a health burden; Cifuentes et al (2000) identify irrigation to be a link between water pollution and health, while Carr (2001) highlights bathing, food, and person-to-person contact as modes of disease transmission from polluted water.

The recent economics literature zeros in on the causal impacts of water pollution. Program evaluation has linked various public health-type initiatives – such as water filtration and chlorination (Cutler and Miller, 2005), piped water access (Gamper-Rabindran et al, 2010), spring protection (Kremer et al, 2011), deep-water tube wells (Field et al, 2011), privatization of water provision (Galiani et al, 2005), and general sanitation projects (Watson, 2006) – to infant health impacts. Moreover, two recent developing-country studies link water pollution *directly* to health outcomes. Ebenstein (2010) finds that a one-grade deterioration in Chinese river water quality is associated with a 9.7% increase in digestive cancer incidence. Brainerd and Menon (2011) find that a ten-percent increase in agrichemical levels in Indian rivers during the month of conception is associated with an 11% (15%) increase in the likelihood of one-year (one-month) mortality. To provide further context, Appendix Table 1 summarizes the findings of these articles.

Among works focusing on air pollution, Currie and Neidell (2005) rely on a detailed set of fixed effects to uncover a causal link between air quality and infant health outcomes in Los Angeles; Chay and Greenstone (2003) use instrumental variables to determine the infant health impacts of air pollution reductions caused by a U.S. economic recession; and Currie and Walker (2011) employ a difference-in-differences specification to assess whether the introduction of E-ZPass automated tolls affects infant wellbeing. In a developing-country setting, Jayachandran (2009) estimates that Indonesian forest fires in 1997 are associated with 15,600 'missing' children.

The second assertion – that Indian surface-water quality levels are socially inefficient – is legitimized by economics-type research coming directly out of India. In this vein, there are several existing estimates of the impact of water pollution on health and welfare. Dasgupta surveys a sample of Delhi households and their water quality and then estimates a health production function from the data. She finds a per-household cost of diarrheal illness (in terms of treatment cost and man-days lost) of 1,094 Rupees per year (for context, Delhi's population was about twelve million as of 2009). Brandon and Homman (1995) estimate the impact of domestic water pollution on a broader swath of illnesses. They find that providing clean water supply and sanitation to the whole of India would avoid between 3 billion and 8 billion U.S. dollars of foregone earnings. In particular, they suggest that 59% of annual environmental costs in India are

incurred by surface water pollution. In addition to these types of revealed preference studies, there are several contingent valuations of water pollution in the Indian literature. Markyanda and Murty (2000), for instance, survey households within one kilometer of the Ganga River and record an average willingness-to-pay of 582 Rupees for bathing-quality river water.

Finally, our third assertion is backed by the strand of economics research devoted to regional environmental spillovers. This literature focuses primarily on how incentives for pollution abatement depend on the spatial incidence of the associated environmental benefits (Gray and Shadbegian, 2004, Helland and Whitford, 2003). Early work by Sigman (2002, 2005) on interstate and international water quality spillovers finds evidence of free-riding: River pollution monitors upstream of state and country borders have worse water quality than other monitors, all else equal. Lipscomb and Mobarak (2007) more recently exploit county-splitting in Brazil to identify the same free-riding phenomenon, and they find that water pollution increases by 2.3 percent for every kilometer closer a river gets to an exiting border. Our research builds on these studies by connecting the phenomenon of upstream free-riding to downstream water quality and health impacts.

The primary challenge faced in the literature on both the pollution-health nexus and regional environmental spillovers is the endogeneity of pollution exposure. Pollution is not randomly assigned. This means that an ordinary least squares regression of health outcomes on pollution levels across regions and time may be biased due to omitted time-varying factors such as income or infrastructure (in health or sanitation, for instance). It also means that upstream and downstream water quality may be correlated in time and space through more than just river flow. The research detailed above employs a combination of experimental variation, fixed effects, and instrumental variables to try to circumvent this problem of bias. In what follows, we do the same, utilizing panel data and the natural flow of rivers to isolate exogenous variation in water pollution levels in our sample.

III. THE INDIAN CONTEXT

Water pollution in India is overwhelmingly an issue of domestic sewage. While industrial water pollution (Central Pollution Control Board, 1989) and agricultural water pollution (Brainerd and Menon, 2011) are present, it is domestic water pollution that is most prevalent. The Central Pollution Control Board (CPCB, 2013) reports that the total sewage load among Indian cities with at least 50,000 inhabitants is 29,129 million liters per day (MLD), while current installed capacity to treat is a mere 6,190 MLD – a 78.7% shortfall. A World Bank (2006) study on sanitation in India underscores the country's problems. Urban areas are plagued by unreliability: Blocked sewers and non-functioning pumping stations lead to raw sewage overflow into rivers. Meanwhile, rural areas suffer from low access: Only 35% of the rural population had access to basic sanitation as of 2006.

Water quality is monitored and regulated by the Ministry of Environment and Forests (MoEF), the CPCB, and the associated State Pollution Control Boards. Early environmental legislation (beginning in the 1970s) has given way to an extensive water quality monitoring network, which includes 1,019 monitors along rivers, lakes, ponds, and wells in India. *Surface* water quality at these monitors is judged by its "Designated Best Use" (DBU) according to standard pollution metrics, as displayed below in Table 1.

Table 1. Surface Water Quality Classification System, India

Class	Designated Best Use	Water Quality Criteria
A	Drinking water source without conventional treatment but after disinfection	Dissolved Oxygen - 6.0 mg/l or more Biochemical Oxygen Demand - 2.0 mg/l or less Total Coliform - 50 MPN/100 ml
B	Outdoor bathing	Dissolved Oxygen - 5 mg/l or more Biochemical Oxygen Demand - 3 mg/l or less Fecal Coliform - 500 MPN/100 ml or less
C	Drinking water source with conventional treatment followed	Dissolved Oxygen - 4 mg/l or more Biochemical Oxygen Demand - 3 mg/l or less Total Coliform - 5000 MPN/100 ml or less
D	Propagation of wildlife and fisheries	Dissolved Oxygen - 4 mg/l or more Free ammonia - 1.2 mg/l or less
E	Irrigation, industrial cooling, and controlled waste disposal	Electrical Conductivity - 2,250 mhos/cu Sodium Absorption Ratio - 26 or less Boron - 2 mg/l or less

Notes

1. Source: Ministry of Environment and Forests. "Water Quality Criteria for Designated Best Use Classification of CPCB: Factsheet." 11/6/2009. <http://pib.nic.in/newsite/erelease.aspx?relid=53897>

In 1985, the Indian government launched the Ganga Action Plan (GAP) to clean up the Ganga River, India's longest. By 1987, the precise goal of GAP was crystallized: Restore the Ganga River to the 'Bathing Class' standard, as defined by the DBU classification in Table 1 (National River Conservation Directorate, 2013). In the ensuing three decades, GAP was extended first to other rivers in the Ganga Basin, and later to rivers all over India. Currently, 190 towns in twenty states along 41 rivers are regulated under what is now known as the National River Conservation Plan (NRCP), India's flagship water pollution cleanup policy.

NRCP has the stated objective of "interception, diversion, and treatment" of sewage (Government of India, 2003); it is therefore a program focused mainly on domestic pollution. To that end, 4,704 MLD of sewage treatment capacity have been created since its inception in 1985 (MoEF, 2013). New infrastructure for intercepting and diverting sewage towards treatment plants has accompanied these capacity expansions. Furthermore, improvements to riverside bathing ghats, crematoria, toilets have also been a part of NRCP interventions (MoEF, 2013). Unfortunately, the policy has been panned in the media for reasons such as poor inter-agency cooperation, funding imbalances across sites, and inability to keep pace with growing sewage loads (Suresh, 2007). Using data from the water quality monitors described above, Greenstone and Hanna (2012) find that NRCP has had no statistically-significant impact on surface water quality.

IV. DATA

A. Pollution Data

For our own work, we use the same water quality data as Greenstone and Hanna (2012). These data were originally gathered from a combination of CPCB online and print records and are a subset of the universe of data collected under India's national monitoring program. We limit our analysis to the sixteen years between 1986 and 2004; this is because our infant health data only extend to 2004, while we do not have access to pollution data from before 1986. These sixteen years of data cover 472 unique monitors along 139 rivers, yielding 39,731 total monitor-month

observations with non-missing domestic pollution measures[1]. 66 of these rivers have more than one monitor; the greatest number of monitors along a single river is 29, on the Ganga River. All pollution monitors are mapped below in Figure 1.

Figure 1. Pollution Monitors along Indian Rivers

Notes
1. Black points denote river pollution monitors.

As many as 46 different measures of water quality are recorded at these monitoring stations, but only a few measures are consistently recorded over the whole sample timeframe[2]. Among these, we choose to focus primarily on Fecal Coliforms (FCOLI), since it is the best indicator of *domestic* water pollution specifically. FCOLI measures the concentration of water-borne bacteria in units of Most Probable Number per 100 ml (MPN). It is exclusively caused by human and animal waste, so it is not an indicator of industrial pollution. Finally, it is not in and of itself a threat to health, but it is highly correlated with the presence of organisms that *are* a health hazard. In analysis, we use a moving average of FCOLI over the current month and the two months prior, in order to fill in some of the gaps in our panel. Additionally, we take the

[1] Monitors record pollution at either the monthly or quarterly frequency.
[2] These are: Fecal Coliforms; Total Coliforms; Biochemical Oxygen Demand; Chemical Oxygen Demand; Dissolved Solids; pH; Alkalinity; Conductivity; Hardness; Turbidity; Total Dissolved Solids; Calcium; Chlorine; Magnesium; Sodium; Sulfates; and Nitrogen.

logarithm of this moving average, since FCOLI appears to be log-normally distributed in our data (see Appendix Figure 1 for this distribution).

Though our water pollution data are recorded at the monitor level, we use district averages for our main regression analysis. This is because we can only identify the location of infants at the district level; we do not know where, *within* a district, each infant lives. However, in our tests of instrument validity, we are able to utilize monitor-level measurements. Thus, Table 2 below displays summary statistics for river pollution at both the monitor-month and district-month level. Our primary independent variable, the logarithm of FCOLI, takes a mean value of 5.5-5.7 but has a long right tail. Its maximum value of 14.56 (corresponding to 2.1 million organisms per 100 ml) is reached in fourteen months by a monitor along the Sabarmati River in Ahmadabad, Gujarat.

Table 2. Summary Statistics for Environmental Variables

		Monitor-Month (1)	District-Month (2)
FCOLI	Mean	21,557	22,089
(Most Probable Number / 100 ml)	Std. Dev.	127,967	115,843
	[Min, Max]	[1, 2,100,000]	[1, 2,100,000]
	N	51,973	27,221
Log(FCOLI)	Mean	5.53	5.72
(unitless)	Std. Dev.	2.84	2.75
	[Min, Max]	[0, 14.56]	[0, 14.56]
	N	51,973	27,221

Notes
1. Column (1) provides statistics compiled from monitor-month observations, while Column (2) provides statistics compiled from district-month observations.
2. Pollution values are computed as the moving average of the current month and the three previous ones.

B. Health Data

In quantifying the health cost of the inter-district pollution externality, we choose infant mortality as our metric. This choice is motivated by several considerations. First, epidemiology research shows that infants are highly susceptible to water-borne pathogens (Fewtrell and Bartram, 2001). Second, infant mortality is a very significant measure of life lost, especially in India, where infant mortality rates remain high relative to the global average (United Nations, 2011). Third, the measure has the advantage of limiting concerns about prior exposure to

pollution, as noted by Chay and Greenstone (2003). Fourth, we have *two* long, panel datasets of infant survival outcomes in India, whereas we have only a small number of repeated cross-sections of other health variables, such as diarrhea incidence and low birth weight.

Our infant health data come from two national demographic surveys. The first of these is the Reproductive and Child Health II (RCH-2) module of the District-Level Household Survey II (DLHS-2), conducted in two phases from 2002 to 2005. In this module, mothers report age and survival for all of their children, which allows for creation of a "pseudo"-panel of birth-month observations. In all, there are 1,393,431 births from 1967 to 2004 in RCH-2. After merging with our pollution data, we have 264,375 births with matched pollution data from 1986 through 2004. From this sample, we expand out each birth to create an observation for each month of life up to the first of three events: (a) death; (b) one year of survival; or (c) the month of survey. This procedure nets us 2.65 million child-month observations with pollution and survival data.

We similarly create a panel of child-month observations from the National Family Health Survey II (NFHS-2) module of the Demographic and Health Survey II (DHS-2), which was carried out from 1998 to 2000. 268,834 total births from 1961 to 1999 in NFHS-2 are cut to 39,125 with matched pollution data from 1986-1999. After expanding out to child-months in the same fashion as with RCH-2, we are left with 388,301 observations containing both pollution and survival data. Summary statistics for infants from both surveys are provided below in Table 3.

Among RCH-2 children born between 1986 and 2004 with non-missing pollution data, the probability of dying in the first month is 0.039 (0.037 for NFHS-2). This translates to 39 (37) deaths per 1,000 live births. The corresponding one-year infant mortality rate, in both samples, is 57 deaths per 1,000 live births. In comparison, the United Nations (UN) Population Division reports India's IM rate to be 60.6 for the years 2000-2005 (UN, 2011). However, the conventional infant mortality rate is not our precise dependent variable; what we study is instead the acute probability of an infant's death in each month of its first year of life. That variable is summarized in the Table 2 row labeled "1[Died this month]". Its mean in both samples is 0.006, which corresponds to a 0.6% average likelihood of death over all child-months in our merged sample.

Table 3. Summary Statistics for Infants

		RCH-2 (1)	NFHS-2 (2)
1-Month Mortality (0/1)	Mean	0.039	0.037
	Std. Dev.	0.194	0.189
	N	264,375	39,125
1-Year Mortality (0/1)	Mean	0.057	0.057
	Std. Dev.	0.232	0.233
	N	264,375	39,125
1[Died this month] (0/1)	Mean	0.006	0.006
	Std. Dev.	0.077	0.08
	N	2,653,310	388,301

Notes
1. Column (1) provides statistics from the RCH-2 survey module, while Column (2) provides statistics from the NFHS-2 survey module.
2. All statistics are compiled from samples composed only of infants in districts for which there is matching pollution data.
3. Stats for the variable "1[Died this month]" are calculated across child-months; stats for the other two variables are calculated across children (not child-months).

C. Other Data

We include severable types of variables as controls in many of our regression analyses. The main body of these controls are cross-sectional survey answers about birth, mother, and child characteristics taken from RCH-2 and NFHS-2. We add to that climate data graciously provided by Sam Asher and Paul Novosad. These include monthly, gridded rainfall averages from the University of Delaware and air temperature averages from the Indian Meteorological Institute. We use the climate grids to interpolate rainfall and temperature values at each monitor location. Finally, we employ a dummy variable for the incidence of major river cleanup policy (NRCP, described above in Section III). Table 4 below provides a full list of control variables employed in our regression analyses.

Table 4. Full List of Control Variables Used in Regressions

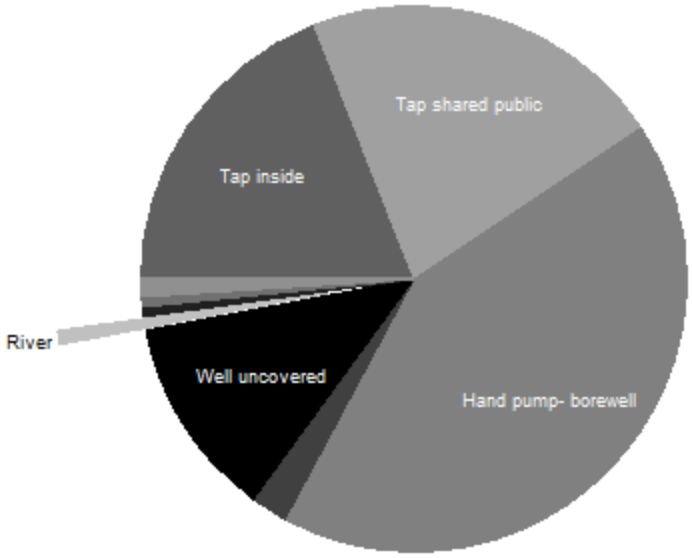
Variable	Data Source
Female (Dummy)	RCH-2 and NFHS-2
Twin (Dummy)	NFHS-2 only
Birth Order (Dummies)	NFHS-2 only
Mother Age (Years)	RCH-2 and NFHS-2
Mother Schooling (Years)	RCH-2 and NFHS-2
Mother Literacy (Dummy)	RCH-2 only
Mother Religion (Dummies)	RCH-2 and NFHS-2
Mother Caste/Tribe (Dummies)	RCH-2 and NFHS-2
Primary Drinking Water Source (Dummies)	RCH-2 and NFHS-2
Type of Toilet Facility (Dummies)	RCH-2 only
Standard of Living Index (Dummies)	RCH-2 only
Electricity (Dummy)	NFHS-2 only
Sanitation (Dummy)	NFHS-2 only
Total Rainfall (mm)	IMD
Average Temperature (degrees Celsius)	University of Delaware
National River Conservation Plan (Dummy)	CPCB

D. A Note on Water Source

Ideally, we would augment our pollution and health data with detailed, time-varying information on individuals' usage of all the types of water available to them. This information would greatly aid us in explaining the mechanism underlying any relationship identified between river pollution and infant mortality. Unfortunately, the only data we possess on water usage at the individual level is a cross-sectional answer to the question "What is your primary source of drinking water?", available in both RCH-2 and NFHS-2. The data provided by this question do not say anything about drinking water choices made before the date of survey; they do not speak to *secondary* (i.e., alternative) drinking water choices; and they do not reveal anything about non-drinking water usage, such as bathing, irrigation, and food preparation.

We nonetheless display 'primary drinking water source' statistics (for RCH-2; the corresponding statistics for NFHS-2 are qualitatively similar) below in Figure 2. Notably, only 1.2 percent of infants are from households citing 'Rivers' as their primary source of drinking water. On its own, this statistic suggests that drinking river water is not likely to be the primary mechanism through which river pollution is related to infant health. However, the aforementioned caveats about these data mean that we cannot rule out the mechanism of drinking river water.

Figure 2. Primary Drinking Water Source, RCH-2



Notes
1. The counts above are taken from all births, rather than all households.

IV. METHODS

We seek to causally link domestic river pollution to infant mortality. We are interested in both the "within-district" relationship – i.e. how river pollution in a given district affects infant mortality in that *same* district – and the downstream spillover – i.e. how much the mortality impact persists in the next district downstream of pollution incidence. For both of these relationships, we require a dataset that links upstream and downstream observations on water quality. Thus, the first step in our analysis is to pair every pollution monitor in our dataset with an "upstream" counterpart. Since there are frequently several monitors along a given river, we have some flexibility with respect to the choice of upstream monitor. Our goal in choosing is to strengthen both the relevance and the validity of the instrument. *Relevance* is an issue because water pollution decays over time; water quality 500 km upstream, for instance, is not a good instrument for downstream pollution because little to no pollution will persist after a distance of 500 km traveled. *Validity* is an issue because water pollution exhibits spatial correlation: Two

monitors that are very near each other may be highly correlated for reasons having nothing to do with river flow.

These concerns prompt us to bound the distance between upstream-downstream pairs. The lower bound mitigates the spatial correlation issue, and the upper bound ensures monitor pairings for which it is reasonable to expect upstream pollution to persist downstream. In practice, we use the following algorithm, for each monitor in our dataset:

1. Identify the river on which the monitor resides.
2. Follow that river upstream until it enters a new district[3].
3. Find the next monitor upstream. If its distance from the original monitor falls within $[X, Y]$, assign it as the upstream monitor. If it is not, continue upstream until a monitor is found that *is*.[4] If no monitor is found within $[X, Y]$, drop that monitor from the sample.

In the above algorithm, we vary the value of X between 20 and 100 km and the value of Y between 200 and 300 km. We then test the relevance and validity of the instrument with each range permutation and choose the range that is empirically most promising.

The first stage of our two-stage least squares (2SLS) estimation strategy, which captures regional water pollution spillovers, is described by Equation 1:

$$P_{kym} = \alpha_0 + \alpha_1 P_{(k-1)ym} + \alpha_X X_{kym} + \phi_k + \pi_{ym} + \tau_{sy} + \varepsilon_{kym} \qquad (1)$$

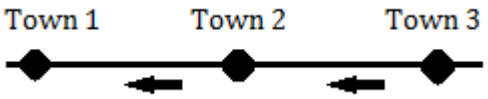
Here, P_{kym} denotes average pollution at monitor k in year y and month m. $P_{(k-1)ym}$ thus denotes the average pollution in the same month at the next monitor upstream. X_{kym} is a vector of monitor-level time-varying covariates – rainfall and air temperature interpolated at each monitor location, in practice. Finally, ϕ_k is a vector of monitor-pair fixed effects, π_{ym} is a vector

[3] Without this step, district-aggregated pollution may be mechanically correlated with its upstream analog due to some monitors being used in both the downstream *and* upstream calculations.
[4] If a river splits during this algorithm, then the original monitor is given two upstream assignments – one for each river arm – and the values at these two upstream locations are averaged.

of year-month fixed effects, and τ_{sy} is a vector of state-year fixed effects. The first vector of fixed effects controls for the time-invariant components of monitor-pair-specific water-quality determinants – such as local soil quality and natural resource endowment. The second controls for those determinants of water quality which vary from month to month but affect the whole country in the same way – such as national trends in economic output and technological development. The third controls for annual determinants of water quality that are specific to states – such as state-year shocks to the economy.

The model in Equation 1 will capture the average pollution spillover at downstream locations in our sample, as long as we control for all other (non-river) factors that jointly determine upstream and downstream water quality. Instrument *relevance* requires that α_1 in Equation 1 is positive and statistically significant. Importantly, we can estimate Equation 1 at the monitor level as shown and at the district level as well.

To assess the *validity* of our upstream instrument, we construct a test that takes advantage of the logic of river flow. Consider the simplistic diagram below, with three towns situated along a single river:



According to our identification strategy, water quality in Town 1 can be predicted by water quality in Town 2 or in Town 3; they are both upstream of Town 1. If our upstream assignment is valid, then, conditional on control covariates, Town 3 water quality is only correlated with Town 1 water quality through the flow of the river. But in that case, the entire pollution spillover from Town 3 to Town 1 is captured by Town 2 water quality levels. Thus, instrument validity requires that both (a) Town 1 water quality is correlated with Town 3 water quality, and (b) *conditional on Town 2 water quality*, Town 1 water quality is *un*correlated with Town 3 water quality. In mathematical terms, the dual condition is

$$\mathrm{Cov}\left(P_{kym} * P_{(k-2)ym}\right) > 0 \quad (2.1)$$

$$\mathrm{Cov}\left(P_{kym} * P_{(k-2)ym} | P_{(k-1)ym}\right) = 0 \quad (2.2)$$

The advantage of using this condition as a test of validity is that it follows logically from the physical laws of river flow. Its disadvantage is that it is a difficult condition to satisfy – as soon as measurement error is introduced into the data, it becomes likely that water quality in Town 2 will not purge all of the correlation between water quality in Towns 1 and 3.

We test the degree to which this condition holds by estimating two regression equations[5]:

$$P_{kym} = \theta_0 + \theta_2 P_{(k-2)ym} + \theta_X X_{kym} + \phi_k + \pi_{ym} + \tau_{sy} + \varepsilon_{kym} \quad (3.1)$$

$$P_{kym} = \beta_0 + \beta_1 P_{(k-1)ym} + \beta_2 P_{(k-2)ym} + \beta_X X_{kym} + \phi_k + \pi_{ym} + \tau_{sy} + \varepsilon_{kym} \quad (3.2)$$

The coefficient θ_2 should be positive and statistically significant, while the coefficient β_2 should be a statistical zero. If this test is passed, then our estimate of α_1 in Equation 1 can be interpreted as the average magnitude of the per-unit river pollution externality.

With a valid instrument, we can focus on quantifying and comparing the health impacts of river pollution at different relative locations. It is, however, important to note households may adapt to variations in river water quality, depending on the visibility of such variation. They may switch their primary water source, or change their investment in water treatment, or even possibly migrate away. Our estimates should thus be interpreted as lower bounds on the dose-response mortality effect of water quality shocks.

Because we do not know the location of infants within their listed district, each infant can only be assigned its district-aggregated measures of pollution and weather. The within-district mortality impact of river pollution is thus captured by the following 2SLS model:

[5] As with Equation 1, estimation of Equations 3.1 and 3.2 can be done at both the monitor- and district-level.

$$P_{diym} = \delta_0 + \delta_1 P_{(d-1)iym} + \delta_X X_{diym} + \alpha_Z Z_{di} + \phi_d + \pi_{ym} + \tau_{sy} + \varepsilon_{diym} \quad (4.1)$$

$$IM_{diym} = \varphi_0 + \varphi_1 \widehat{P_{diym}} + \varphi_X X_{diym} + \gamma_Z Z_{di} + \phi_d + \pi_{ym} + \tau_{sy} + \eta_{diym} \quad (4.2)$$

Above, IM is a dummy variable for infant mortality which varies with district d, infant i, year y, and month m. X remains a vector of climate variables, varying by district and month. Z is a vector of demographic characteristics at the infant, mother, and household levels (the precise characteristics are listed in Table 2 above). Meanwhile, the reduced-form analog of the IV model above captures the downstream impact of river pollution:

$$IM_{diym} = \gamma_0 + \gamma_1 P_{(d-1)iym} + \gamma_X X_{dym} + \gamma_Z Z_{di} + \phi_d + \pi_{ym} + \eta_{diym} \quad (5)$$

If the mortality risk imparted by water pollution varies with age, such variation will be obscured by the models described by Equations 4.1/4.2 and Equation 5. We are particularly interested in modeling age-specific mortality risk in light of the numbers in Table 3 of Section IV above: In both of our survey samples, the majority of infant mortality in the first year comes in the first month of life (approximately 65%). To examine the possibility of a pollution-health relationship that varies with age, we add an interaction term to the IV and reduced-form systems above. This interaction takes the form $P_{diym} \times 1[Age = 1\ month]_{diym}$ – a single term that is non-zero only for neonates (infants in their first month of life). Age (in months) fixed effects are added to complete the specification.

V. RESULTS

A. Pollution regressions

The first results we show are basic estimates of the first stage. Table 5 below displays point estimates of Equation 1 at both the monitor level and district level. We show only the results using upstream bounds of [50 km, 300 km] (Panel A) and [75 km, 300 km] (Panel B), for which our empirical results are soundest. Results with the other ranges considered are, however, qualitatively similar. All regressions employ year-month, monitor-pair, and state-year fixed effects, controls for rainfall and air temperature, and standard errors clustered at the monitor-pair level. Columns 1 and 3 reflect Equation 1 as is, at the monitor and district levels, respectively.

Column 2 adds an interaction term to allow the association of upstream and downstream to vary with distance. The un-interacted distance term is subsumed by monitor-pair fixed effects; that is why it is not shown in Table 5.

Table 5. First-Stage Pollution Regressions

	MONITOR		DISTRICT
	(1)	(2)	(3)
Panel A. Bound of [50 km, 300 km]			
Upstream Log(FCOLI)	0.259***	0.358***	0.309***
	(0.051)	'(0.078)	(0.049)
Upstream Log(FCOLI)*Distance (km)		-0.0009*	
		(0.0005)	
R^2	0.81	0.81	0.83
N	55,700	55,700	19,900
Panel B. Bound of [75 km, 300 km]			
Log(FCOLI)	0.195***	0.387***	0.229***
	(0.053)	(0.097)	(0.047)
Upstream Log(FCOLI)*Distance (km)		-0.0014**	
		(0.0006)	
R^2	0.79	0.80	0.82
N	57,200	57,200	19,000

Notes
1. The dependent variable in all regressions is (downstream) Log(FCOLI).
2. All regressions include total rainfall and average air temperature as controls, as well as fixed effects by monitor-pair (or district) , year-month, and state-year.
3. All regressions cluster standard errors by monitor-pair.
4. Pollution values are computed as the moving average of the current month and the three previous ones.

The results in Table 5 show that our first stage is very relevant: The magnitude of the upstream coefficient is large and its significance is high. According to column 1, a one-percent rise in upstream FCOLI is associated with a 0.259-percent rise in downstream FCOLI. Meanwhile, column 2 shows that this association weakens as the distance between two monitors increases: The sign on the interaction term is negative and significant. Together, the pollution coefficients above are a first check on the validity of our identification strategy. If our identification is coming from actual pollution flow (as we desire), then increasing distance should be found to weaken the upstream-downstream relationship[6]. Indeed it does, according to Table 5.

[6] However, off-river spatial correlation might be expected to have a similar relationship with distance.

The next results we show are from the test of validity proposed in the preceding section. Table 6 below displays point estimates for Equations 1, 3.1, and 3.2 in columns 1, 2, and 3, respectively. For this table, the pollution sample is restricted to all monitor-months for which there is non-missing data at downstream, 1st-upstream, and 2nd-upstream locations. Control variables, fixed effects, and clustering choices are exactly the same as in Table 5 above.

Table 6. First-Stage Validity Test Results

	(1)	(2)	(3)
Panel A. Bound of [50 km, 300 km]			
1st-Upstream Log(FCOLI)	0.200***		0.197***
	(0.048)		(0.048)
2nd-Upstream Log(FCOLI)		0.045**	0.019
		(0.020)	(0.015)
R^2	0.84	0.84	0.84
N	30,400	30,400	30,400
Panel B. Bound of [75 km, 300 km]			
1st-Upstream Log(FCOLI)	0.150***		0.149***
	(0.051)		(0.051)
2nd-Upstream Log(FCOLI)		0.020*	0.010
		(0.012)	(0.012)
R^2	0.83	0.83	0.83
N	28,900	28,900	28,900

Notes
1. The sample is restricted to all monitor-months for which there is pollution data at downstream, 1st-upstream, and 2nd-upstream locations.
2. The dependent variable in all regressions is (downstream) Log(FCOLI).
3. All regressions include total rainfall and average air temperature as controls, as well as fixed effects by monitor-pair (or district) , year-month, and state-year.
4. All regressions cluster standard errors at the monitor-pair level.
5. Pollution values are computed as the moving average of the current month and the three previous ones.

Recall that we are looking for evidence that (a) both first-upstream pollution and second-upstream pollution are predictors of downstream pollution; and (b) second-upstream pollution loses its predictive strength once first-upstream pollution is included in the same regression. Table 6 provides just this evidence. Columns 1 and 2 indicate that both measures of upstream pollution are significant predictors of downstream levels, regardless of whether the lower bound is 50 km (Panel A) or 75 km (Panel B). Further, the second-upstream coefficient has a smaller magnitude than the first-upstream coefficient. This makes sense, since second-upstream monitors are significantly further away from downstream monitors than their first-upstream counterparts.

Column 3 depicts the results of regressions with both upstream monitors as explanatory variables, yielding exactly the result we would like to see. The first-upstream coefficient remains highly significant and decreases only negligibly in magnitude, while the second-upstream coefficient loses both its size and significance. Put another way, second-upstream water quality is not correlated with downstream water quality conditional on first-upstream water quality. This result suggests that our instrument does what we want: It isolates upstream pollution that drives downstream pollution *only* via the transmission of pollutants along rivers.

More than simply validating our instrument, the results of Table 6 contribute to the economics literature on regional pollution spillovers. Column 6 provides hard evidence that upstream pollution spills over into downstream districts, suggesting a very real cost to the free-riding pollution behavior identified by Sigman (2002, 2005) and Lipscomb and Mobarak (2007). As yet, however, this cost is expressed in terms of changes in water quality, which are difficult to value. This, of course, is exactly why we estimate 2SLS and reduced-form models of infant health – such analysis translates water quality costs into human health terms. Admittedly, infant mortality is only a partial measure of human health costs, and human health is in turn merely one component of the total costs of water pollution – ecosystem health, recreation, and non-use values being some of the others. Nonetheless, infant mortality carries with it a huge swath of life-loss for a particularly vulnerable sub-population.

B. Health regressions

Table 7 displays our main estimation results for econometric models of infant mortality. For this table, we use RCH-2 as its infant mortality sample, since that survey is by far the larger one. Columns 1 and 2 correspond to an OLS model where downstream infant mortality is regressed on *downstream* pollution. Thus, upstream pollution is omitted from consideration entirely in these regressions; this is our "naïve" estimator. Columns 3 and 4 correspond to our IV model and thus depict estimates of Equations 4.1 and 4.2. The IV estimator is key for identifying the within-district negative externality of river pollution. Finally, columns 5 and 6 correspond to our reduced-form (RF) model and thus show estimates of Equation 5, which captures the spillover of pollution-induced harm at downstream districts.

An observation here is a child-month; however, pollution and weather values reflect district-averages as mentioned earlier, since children are identified only by their district of origin and not their village. All regressions include year-month, state-district, and state-year fixed effects, standard errors clustered at the district level, and weights provided by DLHS-2. Columns 1, 3, and 5 represent estimation results from models that do not allow the pollution-health relationship to vary by age. In contrast, columns 2, 4, and 6 represent models that allow for a different pollution-health relationship among neonatal infants.

Table 7. Health Regression Results, RCH-2

	OLS		IV		RF	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Bound of [50 km, 300 km]						
Log(FCOLI)	-0.065	-0.07	0.041	-0.291	0.052	-0.057
	(.)	(.)	(0.194)	(0.248)	(0.051)	(0.055)
Log(FCOLI)*1[Age=1 month]		0.051		3.676***		1.272***
		(.)		(1.158)		(0.376)
Implied elasticity for one-month-olds				*0.09*		*0.03*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	1,870,000	1,870,000	1,790,000	1,790,000	2,190,000	2,190,000
Panel B. Bound of [75 km, 300 km]						
Log(FCOLI)	-0.08	-0.102	0.17	-0.088	0.062	-0.045
	(.)	(.)	(0.280)	(0.281)	(0.052)	(0.056)
Log(FCOLI)*1[Age=1 month]		0.254		3.557***		1.256***
		(.)		(1.143)		(0.368)
Implied elasticity for one-month-olds				*0.09*		*0.03*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	1,780,000	1,780,000	1,700,000	1,700,000	2,100,000	2,100,000

Notes

1. An observation is a child-month.
2. Pollution and weather values are aggregated to the district-level for matching with infants.
3. The dependent variable in all regressions is 1[Died this month].
4. All regressions include fixed effects at the year-month, state-district, and state-year levels and the full set of controls.
5. All regressions cluster standard errors at the district level and use sampling weights provided by DLHS-2.
6. Pollution values are computed as the moving average of the current month and the three previous ones.

The first observation we make about the results in Table 7 is that the point estimates in columns 1, 3, and 5 are very small and statistically insignificant[7]. That is, the average association between fecal coliforms concentration and infant mortality over our whole sample is very weak. Omitted variable bias does not explain this finding, since the point estimates are insignificant even with quasi-random variation in water quality and a large set of controls and fixed effects. However,

[7] In fact, we are unable to estimate standard errors (SEs) for our OLS point estimates. Removing state-year fixed effects allows SE computation and confirms that none of the pollution coefficients are statistically significant.

columns 2, 4, and 6 separately identify a pollution-health association for neonates, and the results *here* suggest a quite strong and statistically significant relationship. Column 2 point estimates still tell us nothing, but this is not surprising, given our endogeneity concerns. The IV serves the precise purpose of removing this endogeneity, and when we look at its results (in Column 4), we see that domestic river pollution imparts a significant mortality burden in the first month of life. The RF (in column 6) utilizes the same exogenous variation as the IV and shows that the mortality burden of river pollution persists well-downstream of the pollution's measurement.

The second observation we make about Table 7 pertains to magnitudes: Taken together, the OLS, IV, and RF results in columns 2, 4, and 6, respectively, exactly match our *ex ante* hypotheses. Consider identification concerns: The conventional wisdom about the non-random assignment of pollution says that the dose-response impact of pollution on health is likely to be biased downwards, because pollution may be positively correlated with economic activity and urbanization. Our empirical results are consistent with this reasoning, yielding much larger health impacts with the IV specification than with OLS. Also consider the logic of river flow: Since water pollution decays with river flow and time, one would predict that the downstream impact is smaller in magnitude than the within-district health impact. Indeed, we find exactly that, with IV point estimates (3.676 and 3.557 in Table 7) being relatively much larger than corresponding RF point estimates (1.272 and 1.256).

Because of the details of our econometric specification[8], these point estimates have the interpretation of "average number of neonatal deaths, out of 100,000 births, associated with a one-percent rise in fecal coliform concentration." Thus, the IV results in Table 7 imply that 3.5 additional children die within one month of birth when FCOLI rises by one percent, and the RF results imply a corresponding marginal mortality burden of about 1.2. Another way to describe these results is in terms of elasticities, which we provide in the italicized rows of Table 7. The implied within-district elasticity of neonatal mortality with respect to FCOLI is 0.09, and the downstream-district analog is 0.03. These may seem small as absolute numbers, but they only reflect small changes in domestic river pollution. In fact, the logarithm of fecal coliforms (our

[8] In particular: We define the infant mortality dummy as a 0/1000 variable for numerical tractability in our tables.

key explanatory variable) varies tremendously from month to month in our data: The sample-wide standard deviation of Log(FCOLI) is 2.84, as Table 2 shows.

We are lucky to have not one but two large samples of infant survival outcomes with which to work, so we can test the robustness of our main results by repeating our analysis with NFHS-2 substituted for RCH-2. Everything in this second analysis is the same as in our first analysis, with the exception of a small number of control variables which are only available from one survey or another (see Table 4 for a list of these variables). NFHS-2 results are displayed below in Table 8. The qualitative results are exactly the same: Increases in domestic river pollution are associated with a very statistically-significant rise in neonatal (one-month) mortality, and this association persists at districts downstream of pollution measurement. The implied elasticities actually rise, from 0.09 to 0.13 in the IV model, and from 0.03 to 0.05 in the RF model.

Table 8. Health Regression Results, NFHS-2

	OLS		IV		RF	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Bound of [50 km, 300 km]						
Log(FCOLI)	-0.066	-0.174	-1.802	-0.345	-0.042	-0.193
	(.)	(.)	(158.219)	(0.464)	(0.126)	(0.133)
Log(FCOLI)*1[Age=1 month]		1.362		4.755***		1.755***
		(.)		(1.616)		(0.586)
Implied elasticity for one-month-olds				*0.13*		*0.05*
R^2	0.03	0.03	0.01	0.03	0.03	0.03
N	280,000	280,000	271,000	271,000	325,000	325,000
Panel B. Bound of [75 km, 300 km]						
Log(FCOLI)	-0.028	-0.152	0.113	-0.315	-0.006	-0.163
	(0.149)	(0.160)	(0.650)	(0.646)	(0.126)	(0.125)
Log(FCOLI)*1[Age=1 month]		1.485**		5.057***		1.822***
		(0.659)		(1.752)		(0.589)
Implied elasticity for one-month-olds				*0.14*		*0.05*
R^2	0.03	0.03	0.03	0.03	0.03	0.03
N	271,000	271,000	259,000	259,000	313,000	313,000

Notes

1. An observation is a child-month.
2. Pollution and weather values are aggregated to the district-level for matching with infants.
3. The dependent variable in all regressions is 1[Died this month].
4. All regressions include fixed effects at the year-month, state-district, and state-year levels and the full set of controls.
5. All regressions cluster standard errors at the district level.
6. Pollution values are computed as the moving average of the current month and the three previous ones.

VI. DISCUSSION

Our instrumental variables analysis (Tables 7 and 8) identifies a strong correlation between domestic river pollution and infant mortality. The logic of our "upstream" instrument for pollution and the empirical evidence provided by our first-stage tests (Tables 5 and 6) suggest that this correlation is causal: Rises in domestic water pollution impart a mortality burden to infants. This is important because the academic literature contains few causal estimates of the water pollution–health relationship directly. While there are many studies of interventions that identify ways to improve health via the channel of water quality, there is a dearth of knowledge on *how* harmful water pollution actually is, particularly in developing countries. We think our research advances the frontiers explored by the two most similar studies to ours. Ebenstein (2012) finds river water quality to be associated with adult cancer mortality in China but does so cross-sectionally, with pollution data from 2004 and mortality data from 1991-2000. Brainerd and Menon (2012) find river pollution to be associated with infant mortality in India – our exact setting – but focus specifically on *agricultural* pollution and can only match infants to *state*-average pollution (we are able to match at the district level).

One particular finding of note from our analysis is that the first month of birth is by far the most dangerous when it comes to river pollution. In fact, we find no evidence of a strong relationship existing between mortality and pollution in months two through twelve of a child's life. Our neonatal (first-month) mortality impact, however, is large and robust. This should not be surprising, given both our summary statistics on mortality and the widespread belief that the first month is the most fragile period of an infant's life. This belief is bolstered by empirical evidence going back to at least Chay and Greenstone (2003), who find a similarly strong impact on neonatal mortality in the context of air pollution in the United States. Meanwhile, two recent studies focus on Indian neonatal mortality in the context of water pollution. Brainerd and Menon (2011) find a significant impact of agricultural pollution; Spears (2012), studying the impacts of rural sanitation (which is intimately related with domestic water pollution), does not.

Though we hope our numerical estimates are clear of bias, we acknowledge that measurement error prevents us from precisely calibrating the water pollution–health dose-response function. We do not know where infants live, nor do we know how infants interact with water; this means

we cannot know their precise exposure to water pollution. Additionally, the magnitudes we find are net of behavioral adjustment, such as water-source switching. We believe our results are likely to underestimate the true dose-response function, because (1) some (if not most) of our sample does not actually consume river water, and (2) some households are liable to switch away from rivers to alternative water sources when river pollution rises.

A further caveat of our results is that we cannot definitively make a case for a particular mechanism by which river pollution imparts a neonatal mortality burden. Drinking water is the most likely channel *ex ante*, but, as we noted in Section IV, only 1.2% of RCH-2 infants come from households whose primary source of drinking water is rivers at the time of survey. As explained in Section II, the epidemiology literature provides evidence that bathing, eating, irrigation, and person-to-person contact can all transmit disease. However, we do not have the data to study these modes of transmission directly. Understanding the mechanism of the river pollution–health relationship is a top priority for our future research.

One final aspect of our research that we wish to highlight is the finding of spillovers of both pollution and mortality in downstream districts. Our discussion of pollution spillovers in Section II highlights the presence of both theoretical and empirical evidence of free-riding in the context of river water quality, but no attempt has ever been made to quantify the costs of such free-riding. The results of our reduced-form regression analysis do just that, albeit in partial fashion (the costs of water pollution are not solely health-related, and infant mortality is an incomplete measure of health costs). The fact that pollution measured in one district kills infants living 50 – 300 km further downstream should be strong incentive for policy dialogue across district borders. Even if zero water pollution is not the optimal policy, the incentive to free-ride on pollution cleanup and the associated health costs identified here make it highly unlikely that the current levels of water quality are socially efficient.

VII. CONCLUSION

Water pollution remains a public health problem today in spite of major policy, spending, and improvements to water supply and sanitation. This holds true especially in India: The National

River Conservation Plan has engendered billions of dollars in expenditures to reduce domestic pollution into India's rivers, yet we nonetheless find that a one-percent rise in domestic river pollution in a given district-month has a causal impact of 3-5 additional neonatal deaths per 100,000 births. This finding is convincingly robust – it holds for each of the two major national surveys of Indian health that we use in our analysis. Furthermore, pollution has health impacts that spill over into downstream districts: We find an additional 1-2 neonatal deaths per 100,000 births in the district downstream of pollution's measurement.

These findings only begin to uncover the details of the water pollution–health relationship, but they shed light on several important aspects to consider for policymaking to improve public health. First, *domestic* (and not just industrial) pollution into *rivers* (and not just groundwater) has a real health cost. Second, infants appear to be very vulnerable to water pollution in their first month of life, and not very much at all in the remainder of their first year. Third, there is a geographic mismatch between the production of pollution and the incidence of its external health costs, which implies a need for cooperation and bargaining across jurisdictional borders.

REFERENCES

Brainerd, E. and N. Menon (2011). "Seasonal Effects of Water Quality on Infant and Child Health in India." Mimeo.

Brandon C. and K. Homman (1995). "The Cost of Inaction: Valuing the Economy-Wide Cost of Environmental Degradation in India." Asia Environment Division, World Bank. Mimeo.

Carr, R. (2001). "Excreta-related Infections and the Role of Sanitation in the Control of Transmission." In L Fewtrell and J. Bartram (eds.), Water Quality Guidelines, Standards, and Health: Assessment of Risk and Risk-management for Water-related Infectious Disease. Pp. 89-114.

Chay, K.Y. and M. Greenstone (2003). "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *Quarterly Journal of Economics* 118(3): 1121-1167.

Central Pollution Control Board (2013). "Industry Specific Standards." <http://cpcb.nic.in/Industry_Specific_Standards.php>

Central Pollution Control Board (2013). "Status of STPs." <http://cpcb.nic.in/statusSTP.php>.

Cifuentes, E., M. Gomez, U. Blumenthal, M.M. Tellez-Rojo, I. Romieu, G. Ruiz-Palacios, and S. Ruiz-Velazco (2000). "Risk Factors for *Giardia Intestinalis* Infection in Agricultural Villages Practicing Wastewater Irrigation in Mexico." *American Journal of Tropical Medicine and Hygiene* 62(3): 388-391.

Currie, J. and M. Neidell (2005). "Air Pollution and Infant Health: What Can We Learn from California's Recent Experience?" *Quarterly Journal of Economics* 120(3): 1003-1030.

Currie, J. and R. Walker (2011). "Traffic Congestion and Infant Health: Evidence from E-ZPass." *American Economic Journal: Applied Economics* 3: 65-90.

Cutler, D. and G. Miller (2005). "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States." *Demography* 42.1: 1-22.

Dasgupta, P. "Valuing Health Damages from Water Pollution in Urban Delhi, India: A Health Production Function Approach." Institute for Economic Growth: Delhi, India.

Deschenes, O., M. Greenstone, and J.S. Shapiro. "Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program and Ozone Reductions." MIT Department of Economics Working Paper No. 12-18.

Ebenstein, A. (2012). "The Consequences of Industrialization: Evidence from Water Pollution and Digestive Cancers in China." *Review of Economics and Statistics* 94(1): 186-201.

Fewtrell, L. and J. Bartram (2001). Water Quality Guidelines, Standards, and Health: Assessment of Risk and Risk-management for Water-related Infectious Disease. IWA Publishing on behalf of the World Health Organization: London, 2001.

Field, E., R. Glennerster, and R. Hussam. (2011). "Throwing the Baby out with the Drinking Water: Unintended Consequences of Arsenic Mitigation Efforts in Bangladesh." Mimeo.

Galiani, S., P.J. Gertler, and E. Schargrodsky (2005). "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy* 113: 83-120.

Gamper-Rabindran, S., S. Khan, and C. Timmins (2010). "The Impact of Piped Water Provision on Infant Mortality in Brazil: A Quantile Panel Data Approach." *Journal of Development Economics* 92: 188-200.

Government of India (2003-2004). "62nd Report of the Public Accounts Committee." 13th Lok Sabha, Union Government.

Gray, W. and R.J. Shadbegian (2004). "'Optimal' pollution abatement – Whose benefits matter, and how much?" *Journal of Environmental Economics and Management* 47: 510-534.

Greenstone, M. and R. Hanna (2011). "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." HKS Working Paper No. RWP11-034.

Helland, E. and A.B. Whitford (2003). "Pollution incidence and political jurisdiction: evidence from the TRI." *Journal of Environmental Economics and Management* 46: 403-424.

Jayachandran, S. (2009). "Air Quality and Early-Life Mortality: Evidence from Indonesia's Wildfires." *Journal of Human Resources* 44(4): 916-954.

Kremer, M., J. Leino, E. Miguel, and A.P. Zwane (2011). "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics* 126(1): 145-205.

Lipscomb, M. and A.M. Mubarak (2007). "Decentralization and Water Pollution Spillovers: Evidence from the Redrawing of County Boundaries in Brazil." International Policy Center Working Paper Series Number 67.

Maria, A. (2003). "The Cost of Water Pollution in India." Paper presented at the conference on Market Development of Water & Waste Technologies through Environmental Economics: Delhi, Oct 2003.

Markyanda, A. and M.N. Murty (2000). "Cleaning Up the Ganga: A Cost-Benefit Analysis of the Ganga Action Plan." Oxford University Press, New Delhi.

National River Conservation Directorate (2013). "Ganga Action Plan Phase I." Government of India. <http://envfor.nic.in/modules/recent-initiatives/NRCD/GAP-1.html>.

Oates, W. (2001). "A Reconsideration of Environmental Federalism". Resources For the Future Discussion Paper 01-54.

Planning Commission (2002). "Tenth Five Year Plan". Government of India.

Sigman, H. (2002). "International Spillovers and Water Quality Within Rivers: Do Countries Free Ride?" *American Economic Review* 92(4): 1152-1159.

Sigman, H. (2005). "Transboundary Spillovers and Decentralization of Environmental Policies." *Journal of Environmental Economics and Management* 50: 82-101.

Smith, V.K. and C.V. Wolloh (2012). "Has Surface Water Quality Improved Since the Clean Water Act?" NBER Working Paper 18192.

Suresh, B.S.V., B.L. Seth, A. Chak, V.K. Dixit, and S. Narain (2007). Sewage Canal: How to Clean the Yamuna. New Delhi: Centre for Science & Environment.

United Nations (2011). "World Population Prospects – The 2010 Revision." United Nations, Department of Economic and Social Affairs, Population Division, Population Estimates and Projections Section.

Watson, T. (2006). "Public Health Investments and the Infant Mortality Gap: Evidence from Federal Sanitation Interventions on U.S. Indian Reservations." *Journal of Public Economics* 90(8-9): 1537-1560.

The World Bank (2006). "India: Water Supply and Sanitation: Bridging the Gap between Infrastructure and Service." The World Bank: p. 6.

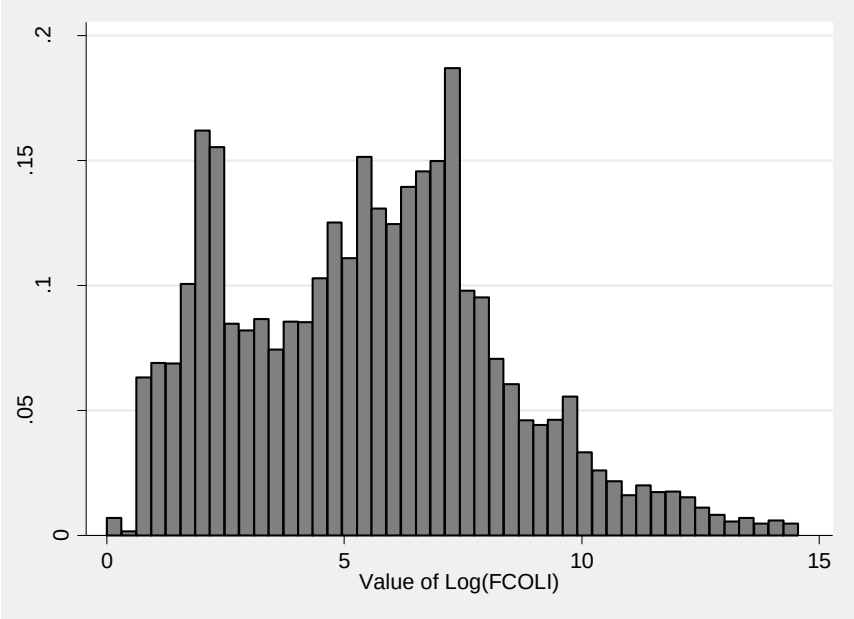
Appendix Table 1. A survey of results from recent studies of water quality and infant health

Author	Year	Locale	Proxy for Pollution	Outcome Variables	Findings
Cutler and Miller	2005	U.S.A.	Clean water technology	1-year mortality	Introduction of water filtration and chlorination systems is associated with infant mortality reductions that explain 75% of overall infant mortality changes between 1900 and 1936.
Galiani, Gertler, and Schargrodsky	2005	Argentina	Privatization of water services	child mortality, aged 0-5	Privatization of water provision is associated with an 8% decrease in child mortality.
Watson	2006	U.S.A. Indian reservations	Sanitation improvements	1-month, 1-year mortality	A 10% increase in the number of sanitation projects is associated with 50 fewer infant deaths per 100,000 births three years after project completion. The effect is driven by non-neonatal mortality.
Ebenstein	2010	China	Water quality grade	digestive cancers	A 1-grade deterioration in water quality is associated with a 9.7% increase in digestive cancer incidence.
Gamper-Rabindran, Khan, and Timmins	2010	Brazil	Piped water provision	1-year mortality	Provision of piped water is associated with 1.25 fewer deaths per 1,000 live births at the 90th percentile of baseline 1-year mortality and only 0.55 fewer deaths at the 10th percentile.
Kremer, Leino, Miguel, Zwane	2011	Kenya	Spring protection	diarrhea among children aged 0-3	Spring protection is associated with a 25% drop in diarrhea incidence among children aged 0-3.
Field, Glennerster, and Hussam	2011	Bangladesh	Fecal contamination	1-year, 2-year, 5-year mortality	An additional year of exposure to high-risk water sources is associated with a 27% increase in the probability of 1-year infant mortality.
Brainerd and Menon	2011	India	Fertilizer agrichemicals	1-month, 1-year mortality	A 10% increase in agrichemical levels in the month of conception is associated with an 11% (15%) increase in the likelihood of 1-year (1-month) mortality.

Notes

1. The studies shown here are merely a subjective sample of the greater economics literature on water quality and health.

Appendix Figure 1. Histogram for Log(FCOLI)



Notes

1. FCOLI values are computed as the moving average of the current month and the three previous ones. Logarithms are applied afterwards.