(Not So) Gently Down The Stream:
River Pollution and Health in Indonesia *

Teevrat Garg[†]
Stuart E. Hamilton [‡]
Jacob P. Hochard[§]
Evan P. Kresch[¶]
John Talbot [‖]

Abstract

Waterborne diseases are the leading cause of mortality in developing countries. We emphasize a previously ignored cause of diarrhea - upstream river bathing. Using newly constructed data on upstream-downstream hydrological linkages along with village census panel data in Indonesia, we find that upstream river bathing can explain as many as 7.5% of all diarrheal deaths. Our results, which are net of avoidance behavior, show no effect of trash disposal on diarrheal infections. Furthermore we find that individuals engage in avoidance behavior in response to trash disposal (visible pollutants) but not river bathing (invisible pollutants). We conduct policy simulations to show that targeting upstream individuals could generate substantial environmental and health savings relative to targeting downstream individuals. This provides a potential roadmap for low- and middle-income countries with limited resources for enforcement of water pollution.

JEL Codes: Q56, D62, H23, Q53, I18
Keywords: Bathing, Water Pollution, Diarrhea, Indonesia, Avoidance Behavior

[†]Corresponding Author. School of Global Policy and Strategy, University of California - San Diego. E: teevrat@ucsd.edu

[‡]Department of Geography and Geosciences, Salisbury University

[§]Department of Economics & Institute for Coastal Science and Policy, East Carolina University

[¶]Department of Economics, Oberlin College

[‖]Eastern Shore Regional GIS Cooperative

1 Introduction

Each day, economic agents face a number of harms and irritants from which they seek to protect themselves. These threats are various, including environmental (e.g., air pollution, water pollution), social (e.g., crime, conflict) and incidental (e.g., accidents) in nature. Against each of these potential harms, individuals can self-protect (e.g. drinking bottled water to protect against water pollution, seat belts to protect against car accidents). To the extent that self-protection is costly (Bartik, 1988; Deschenes, Greenstone and Shapiro, 2012), individuals optimize self-protection and protect themselves to varying extents from each pollutant such that the marginal cost of avoidance is less than the marginal benefit from reduced exposure to pollutants. However, this duality between observed choice and experienced utility can break down when exposure to certain pollutants is not salient, resulting in low avoidance behavior or defensive underinvestment - even when such behavior can be welfare improving. This is of particular importance in developing countries, where individuals routinely underinvest in profitable and health saving technologies (Ashraf, Berry and Shapiro, 2010; Bryan, Chowdhury and Mobarak, 2014; Greenstone and Jack, 2015; Barrett, Garg and McBride, 2016). In such a context, an understanding of the nature and extent of externalities as a result of these "silent killers" can increase the efficacy of regulatory interventions designed to promote public health and environmental improvements.

In this paper, we consider the case of river pollution and resulting waterborne diseases in Indonesia. We focus on diarrhea, which globally accounts for more than 1.5 million deaths each year (WHO, 2014). Freshwater pollution is of particular importance in low- and middle-income countries where untreated river water is routinely consumed, in part due to the low

enforcement of policies intended to prevent contamination.[1] Earlier work in the area of freshwater pollution has focused primarily on industrial waste disposal in rivers (Ebenstein, 2012). We emphasize a previously ignored and seemingly benign source of water pollution, in-river bathing, and show that it has a large welfare cost. We present three major findings: (1) upstream river bathing in Indonesia can explain as many as 7.5% of all diarrheal deaths in a given year, which over our four year sample translates to 865 diarrheal fatalities, (2) individuals exhibit avoidance behavior in response to upstream trash disposal (visible) but not to upstream river bathing (invisible) and (3) targeting upstream villages can reduce diarrheal mortality by 57% more than would targeting downstream villages.

The challenge in causally estimating the impacts of river pollution on public health is finding variation in river pollution that is exogenous to local health outcomes, and also large enough to be economically meaningful. While researchers have previously employed randomized designs in subsidies for provision of clean water (Ahuja, Kremer and Zwane, 2010; Kremer et al., 2011), to the best of our knowledge no one has used experimental or quasi-experimental variation to study the impact of specific river pollutants on local health outcomes.[2] We fill this void in the literature by constructing a novel data set of drainage basins in Indonesia to assign to each of the villages in our sample their respective set of upstream and downstream villages from approximately 5.8 billion possible upstream-downstream hydrological linkages.

Using biennial village census data from 2000 to 2008 and employing village fixed effects,

---

[1] Greenstone and Hanna (2014) find that while air pollution regulations had a measurable impact on infant mortality in India, water pollution laws had no measurable effect.

[2] The closest study to our work is Ebenstein (2012) who uses rainfall as an instrument for water quality (and not polluting behavior or actual pollutant concentrations) at certain sites in China to study the impact of poor water quality on digestive cancers.

we rely on the identifying assumption that year-to-year changes in upstream polluting behavior are exogenous to downstream health outcomes, which is plausible for two reasons. First, household level in-river polluting behavior in Indonesia is *de facto* unregulated, minimizing the regulatory factors that could drive household pollution selectively towards low-income villages. Second, we use data on *upstream* polluting behavior rather than data on *local* pollution levels (Lipscomb and Mobarak, 2016) or official quality grades (Ebenstein, 2012). As a result, we are not relying on correlating local pollution with local health outcomes, which could be spurious for many reasons including but not limited to geographic (Tiebout) sorting. Instead, we rely on upstream behavior that is plausibly exogenous to downstream health outcomes.

We test the validity of our identifying assumption and rule out geographic sorting through a battery of placebo and falsification tests and find that for a given village, bathing by upstream villages increases diarrheal incidence, while bathing by downstream villages has no effect. Furthermore, we show that the effect is specific to diarrhea (consistent with its waterborne nature), with no measurable effect on other non-waterbourne diseases. The absence of an effect on other diseases is inconsistent with the existence of a spurious correlation between upstream polluting and downstream health, as we would expect to see that manifest as a significant impact between upstream bathing and at least one of the other major diseases as well.[3] Importantly, while alternative explanations could exist for each of our results, we believe there is no plausible alternative explanation that would rationalize all of our empirical findings.

---

[3]The remaining threat to identification would be if individuals geographically sort across upstream and downstream villages within a province in our time frame in response to time-varying omitted variables correlated with upstream polluting behavior and diarrheal outbreaks, but not upstream polluting behavior and outbreaks of malaria, measles, respiratory infections or dengue. Given that in Indonesia, downstream (coastal) villages tend to be economically better off than upstream villages, this seems very unlikely.

Our results are relevant for policymakers interested in reducing mortality from waterborne diseases. Given the limited enforcement of water pollution in most low- and middle-income nations, we identify sources of pollution that individuals fail to avoid and crucially, where enforcement resources will be most effective. When large-scale government programs aimed at river basin cleanups are financially or technically infeasible, policymakers could enact cost-effective prevention policies that could substantially reduce diarrheal incidence; we show that a 1% reduction in the most upstream decile's river bathing activity reduces downstream diarrheal incidence by 2.54%.

The policy implications of our work are underscored by contributions to several relevant research areas in economics. First, we build on the literature emphasizing the estimation of causal impacts of environmental quality on human health, particularly in developing countries (see Graff Zivin and Neidell (2013) for an exhaustive review). The construction of all riparian linkages between villages using a combination of geospatial and hydrological techniques enables us to overcome previously identified data limitations (Currie et al., 2014), and extend the research on the linkage between environmental quality and human health to freshwater pollution. Our data construction and identification method can generalize to other settings to understand downstream externalities when the path of the pollutant determines the marginal social cost of pollution.

Second, we contribute to the literature on consequences of water pollution (Leggett and Bockstael, 2000; Resosudarmo, 2003; Dasgupta, 2004; Carson, Koundouri and Nauges, 2011; Ebenstein, 2012; Currie et al., 2013), as well as strategies to reduce associated health consequences (Jalan and Ravallion, 2003; Cutler and Miller, 2005; Galiani, Gertler and Schargrodsky, 2005; Ahuja, Kremer and Zwane, 2010; Graff Zivin, Neidell and Schlenker, 2011).

We are unaware of any paper that isolates pollutants differentially on the basis of avoidance behavior, even though such compensatory behavior has been documented in several instances (Deschenes, Greenstone and Shapiro, 2012).

Third, we relate to a growing literature on the political economy of environmental and natural resources (Burgess et al., 2012; Brollo et al., 2013). Our finding that targeting individuals in upstream areas can generate much larger health savings than targeting individuals in downstream areas is consistent with other work in this literature, such as Lipscomb and Mobarak (2016), who show that within governing jurisdictions, water quality is lowest in the downstream areas.

The rest of the paper is organized as follows. In section 2 we provide an overview of water pollution regulation in Indonesia and the epidemiological evidence on the link between river bathing and diarrhea. In section 3 we describe the health and demographic data used in this paper, as well as the construction of the upstream-downstream village networks. Section 4 details the econometric strategy that we use and in section 5 we discuss the empirical results. In section 6, we simulate different targeted moratoriums on river bathing and associated impacts on health outcomes. Section 7 provides our concluding notes.

2   Background

In this section, we provide a brief overview of the state of water pollution and associated regulations in Indonesia to demonstrate that river pollution, particularly originating from households, is *de facto* unregulated. We also provide evidence of the epidemiological foundations on the impact of river bathing on water quality.

2.1 Water Pollution Regulation in Indonesia

Indonesia has made recent advances in environmental regulation, including the 2009 Environmental Protection and Management Law that recognizes the "serious problem" of decreasing environmental quality, as well as executive actions designed to reduce emissions and other forms of pollution (Nachmany et al., 2014). Yet, the regulation of water pollution in Indonesia can be characterized as nominally mandated but not regulated for some industries, and fully non-existent for others. Ostensibly, any individual or business that purposely pollutes or otherwise damages water sources can face imprisonment for up to 9 years and a maximum fine of 1.5 billion rupiah (USD 115,000), in accordance with Article 94 of Indonesia's Law No. 7/2004 concerning water resources. The provincial governments are responsible for the regulation and supervision of all water resources, including rivers, that fall within their jurisdictions. Despite the steep penalties for polluters, the quality of Indonesia's water sources remains low. Several studies including those by the Ministry of the Environment (MoE) found that all 35 rivers that were tested across Indonesia were unsafe sources for drinking water (AECEN, 2008).[4]

The most unregulated source of water pollution in Indonesia is household and municipal discarding of sewage. Households routinely dispose of waste directly into rivers, while the improper construction of municipal wastewater facilities leads to the disposal of untreated sewage into river waters (Kerstens et al., 2013). Nearly two-thirds of the Citarum River's biological oxygen demand (BOD) comes from household pollution, as compared to one-third from all industrial and agricultural activities combined (Kerstens et al., 2013). Regulation of water pollution at the household level is non-existent, with households polluting into lakes

---

[4]In fact, the Citraum River has been found to be one of the most polluted places on Earth (Bernhardt and Gysi, 2013).

and rivers with *de facto* impunity.

Water pollution in Indonesia is also generated from industrial waste and agricultural run-off (GWP, 2013). Industrial polluting causes toxic materials such as heavy metals and mercury to enter and poison drinking water sources. The 1989 Clean Water Program (CWP), a government initiative to curb water pollution, achieved spotty reductions in industrial pollutants with disproportionate success in East Java (Lucas and Djati, 2000). The mixed success of the CWP may be attributable to its enforceability, as the program was designed to be voluntary (Bedner, 2010) and water pollution regulations across Indonesia generally do not apply to small firms and home industries (Braadbaart, 1995).[5]

2.2   Implications of River Bathing

River bathing poses two major risks to human health, both of which are symptomatic of diarrhea. First, riparian bathing increases the amount of free carbon dioxide (CO_2) and decreases the amount of dissolved oxygen (DO) in rivers (Bhatnagar and Sangwan, 2009; Sharma, Bhadula and Joshi, 2012). Organic and biodegradable waste from the bathers is decomposed by microbes that use oxygen and release carbon dioxide back into river water. This effect is amplified by the use of soaps and detergents that are absorbed by aquatic flora. Higher CO_2 levels drive phosphate and alkalinity concentrations, which lead to river eutrophication. Consumption of water from eutrophic rivers has been linked to gastroenteritis (WHO, 2002) and cyanobacterial toxins (Scott et al., 1985; Wu, 1999), both of which can cause symptoms of intestinal pain, nausea, and diarrhea. Second, river bathing can lead

---

[5]This is true for small scale industrial enterprises, as well as small and family-run farms, which account for nearly a fifth of national GDP. In general, agriculture has been relatively less regulated than industrial activity. For example, the Clean Water Program only applies to industrial firms and no counterpart has been created for the agricultural sector.

to an increased presence of coliform bacteria, which are associated with harmful pathogens known to cause nausea, vomiting, and bloody diarrhea, especially among infants and those with compromised immune systems (Joshi and Sati, 2011; Tyagi et al., 2013). While fecal coliforms do not necessarily cause diarrhea, their presence is correlated with diarrhea pathogens that may arise from removing trace amounts of fecal matter from the body during bathing, the cleaning of infants directly in the river after defecation, and from bathers (particularly infants) defecating while bathing.

3 Data

In this section, we describe the health and demographic data, the geospatial methods used to conduct the upstream and downstream assignment of villages along Indonesia's hydrological network, and the classification of each village into an identifiable drainage basin.

Health and Demographic Data

The Indonesian statistical agency, *Badan Pusat Statistik* (BPS), conducts a biennial census of all Indonesian villages known as *Podes*. Our sample consists of an unbalanced panel of 32,107 villages[6] over years (2000, 2003, 2006 and 2008) spread across all major Indonesian islands with the exception of Java.[7] The census is conducted in a short span of 4-6 weeks in October and November and consists of an exhaustive questionnaire to which village heads

---

[6] We have 25,894 villages in 2000, 22,952 villages in 2003, 28,041 villages in 2006 and 32,107 villages in 2008. We construct our key independent variable - upstream bathing - separately for each year, reducing concerns over an unbalanced panel.

[7] Java, where the capital city of Jakarta is located, is the most densely populated island with an unusually high level of urbanization relative to the other islands. Additionally, Java has differential level of enforcement of water pollution that we do not have the data to capture. We leave it as a future follow-up exercise, and in this paper we focus on all other major Indonesian islands.

respond.[8] The census contains village-level information on health, population, river location and other demographic variables.

Disease Data

The *Podes* is the most spatially and temporally expansive dataset on outbreaks and resulting deaths from five major diseases: dengue, diarrhea, malaria, measles and respiratory infections. In each year of the census, the village head is asked to report if there was an outbreak of each of the different infections in that year. To the best of our knowledge, the village head is not provided any instructions - such as a cut-off or a point of comparison with respect to deaths or infections - in determining whether an outbreak has taken place. As such it is feasible that there were disease-driven deaths in a given year that were unreported in *Podes* because the degree of spread was not determined "high" enough to be considered an outbreak.

Therefore, we validate the accuracy of this 'outbreak' measure. We use actual mortality data from *Podes* that is only available for 4681 village-years.[9] In the appendix table A.1, we show the outcomes using log mean death rate per epidemic using the 'outbreak' variable and the log of *actual* death rate and show that the results are qualitatively and quantitatively similar. With this validation, we defer to using the outbreak variable since we have a larger number of observations over a more diverse area geographically. Use of the outbreak variable generates increased power for better and more nuanced estimates of the effect of upstream bathing on downstream health outcomes.

---

[8]We are aware that some components of this information are verified at the sub-district or district offices but we do not have information of the precise sections of the survey that are verified.

[9]We have 19,933 village-years, of which only 4681 village-years have more than one year of data and provide useful variation for our panel data approach.

Table 1 provides summary statistics on the probability of village-level outbreaks for different diseases. Diarrhea, following only malaria, is the second most prevalent disease in Indonesia, followed closely by respiratory infections. Diarrhea is slightly more prevalent in hilly areas relative to flatter ones, and in rural areas relative to urban settings. This geographic pattern of disease prevalence is consistent across all diseases with the exception of dengue, which is prone to urban areas. The differences between these groups are statistically significant but not meaningfully large (columns 2 and 3).

Population and Demographic Data

We also obtain population and demographic data from *Podes*. In addition to village population information, the census also contains information on whether a river passes through a village, which we use to ground truth the hydrological river network data. *Podes* also contains information on a range of socio-economic variables that we use for robustness in our econometric strategy: dominant source of income in village, geography of village, quality of governance (e.g. education of village head), access to medical facilities in the village and political status of the village.

The dominant form of trash disposal (e.g. carry away, burnt, river polluting, other) and bathing activity (e.g. in-river, other) is reported in *Podes* for each village. These are binary variables and the census does not contain information on the number of individuals in each category. As an alternative, we use village populations to construct our key explanatory variables: the number of individuals polluting in the river through trash disposal and in-river bathing.

To create these variables, we replicate the following exercise for both trash disposal and

river bathing for both upstream and downstream aggregate measures, which we illustrate by considering our key independent variable - upstream river bathing. For a given year t, village v, with n_{vt} villages upstream along the river that passes through village v, we define our key independent variable - the number of individuals engaging in upstream river bathing as,

$$Upstream_{vt} = \sum_{n_{vt}} [population_{nt} * bathing_{nt}] \tag{1}$$

where $population_{nt}$ is the population of the n_t^{th} upstream village and $bathing_{nt}$ is a binary variable, which is equal to 1 if the majority of households in the village bathe in the river in year t, and 0 otherwise. We repeat this exercise for all downstream villages.

It is important to note that our independent variables could be either over- or understated. Including all individuals as river bathers, where only the majority of individuals engage in river bathing, likely overstates the number of individuals bathing in rivers. However, excluding any individuals as river bathers in villages where less than a majority of individuals bathe in rivers understates the number of individuals bathing in rivers. The concern of unpredictable measurement error in our key independent variables is addressed in our placebo tests in table 7.[10] Since we construct our upstream and downstream variables in the same way, any bias should be present in both, and given that the downstream effects are negligible we are confident that this is not a source of bias in our results. Cautiously, we may interpret our results as the differential impact of upstream bathing health effects net of downstream

---

[10]Another concern may be turnover in village heads significantly altering the responses in *Podes*. *Podes* does not contain information on the village head, so it is not possible to control for this using village head fixed effects. However, in order for this concern to jeopardize our identification strategy, we would need the village head turnover to be correlated with the changes in upstream behaviors, which seems unlikely.

bathing health effects. We find equivalence in the two interpretations due to the approximate null effects of downstream river bathing.[11]

Construction of Drainage Basins Data and Assignment of Upstream and Downstream Villages

Linking villages along a hydrological network enables us to track the impact of upstream river bathers on downstream river users. Indonesia contains seven major islands with relatively mountainous and high-elevation interiors that create a complex hydrological network of streams and rivers. The official river network for the country fails to identify minor waterways that are being used by villages for bathing, drinking and trash disposal. Conducting a classic hydrological network analysis poses the risk of failing to assign villages located along minor rivers to the river network, which may understate upstream pollutant runoff (figure 2).

Instead of tracing the hydrological network directly, we adopt a watershed approach that identifies all upstream-downstream relationships within each basin using a high resolution digital elevation model (DEM), and then determine on-river status using survey data. Proprietary approaches to processing such a DEM are less adept to managing canopy interference - where the presence of tree canopy is mistaken for terrain - which could also render an underestimation of the number of upstream villages connected topographically to a given downstream village. The problem of canopy interference is compounded by the approximately 5.8 billion possible village relationships across Indonesia.[12]

---

[11]Another way to characterize our key independent variable described in equation (4) is as a scalar multiple of the population weighted average of the number of villages where a plurality of households engage in polluting behavior, with zero weight being assigned to villages where the dominant source of bathing (or trash disposal) is not in the river.

[12]We manage canopy interference and computational processing constraints by developing a clustered implantation of the

A high-resolution 30 meter void-filled DEM was used alongside village administrative boundaries from *Podes*. A pour point was then constructed for each individual village that self-reported being located on a river.[13] The use of self-reported locations in this way provides a means to "ground truth" our data to avoid misclassifying on-river (type 1 error) and off-river (type 2 error) villages. The mapping of all upstream and downstream village relationships was conducted independently for each sample year (2000, 2003, 2006 and 2008) to accommodate the redistricting of administrative units.[14] The product of the hydrological analysis was a list of every Indonesian village and its ordered upstream counterparts across the four sample years (approximately 13.7 million upstream observations). The comprehensive nature of the GIS output enables us to (1) calculate the relevant upstream and downstream pollutant measures, (2) define all distinct drainage basins that represented the level of exposure to river water pollutants, and (3) simulate geographically-targeted policy interventions.

4   Estimation and Identification Strategy

The challenge in identifying the effects of water polluting behavior on human health is finding exogenous variation in water pollution that is large enough to capture an economically measurable effect. There are many plausible reasons why exposure to, and consumption of, impure water may be determined endogenously. For instance, poorer individuals who have a lower stock of health may be financially or behaviorally constrained from consuming clean

---

r.watershed and r.water.outlet algorithms in GRASS GIS v7.

[13]Each pour point identified the village's maximum upstream catchment, which is bound by its drainage basin.

[14]Three sets of verifications were conducted. First, the GIS open source algorithms were compared against the ESRI algorithms. Second, the flow accumulations were consistent with the official Indonesian River Network. Third, the construction of drainage basins was verified by ensuring the official rivers network flowed properly through each basin ensuring that the constructed basins were logically sound.

water. Instead of correlating local water pollution with local health outcomes, we focus on the diarrheal incidence in a given village due to individuals who are geographically separated but whose (unregulated or unenforced) polluting behavior may affect downstream villages through river networks. Relying on the identifying assumption that year-to-year changes in upstream polluting behavior are exogenous to downstream diarrheal incidence, we estimate a linear probability model:

$$I(DiarrhealOutbreak = 1)_{vpt} = \sum_j [\beta_{1j} Upstream_{jvpt} + \beta_{2j} Downstream_{jvpt}] + \gamma_v + \eta_{pt} + \epsilon_{vpt}$$

(2)

$I(\cdot)$ is an indicator variable equal to 1 if the village v in province p at time t had an outbreak of diarrheal disease, and equal to 0 otherwise.[15] $Upstream_{jvpt}$ is the number of people upstream of village v that are engaging in type j polluting behavior (i.e. trash, bathing) in province p at time t. Similarly, $Downstream_{jvpt}$ is the number of people downstream of village v that are engaging in type j polluting behavior. γ_v denotes the village fixed effects that control for time-invariant heterogeneity across villages and η_{pt} are the province-year fixed effects that control for provincial level shocks across all years. ϵ_{vrt} is the error term, with standard errors clustered at the drainage basin to allow for arbitrary serial correlation across villages along the same drainage basin over time.

While focusing on geographically-separated polluting behavior can avert some endogeneity concerns, to the extent that individuals could geographically sort over time with wealthier individuals ending up in villages with cleaner water, the coefficients on upstream polluting, β_{1j} would remain biased. Since we use province-year fixed effects and thereby control for

---

[15]We validate the use of the binary outbreak variable as opposed to data on death rates in the appendix (table A.1).

all province-specific changes over time, we will focus on such geographic sorting within a province over time. To overcome these concerns, we test the validity of our identifying assumption with a battery of placebo tests. First, upstream pollution could have an effect on downstream individuals' polluting behavior, but downstream polluting should not have a direct effect on upstream individuals' health. Second, we estimate the effect of polluting behavior on diseases that are not transmitted through ingestion of contaminated water, such as measles, malaria, respiratory infections, and dengue. If we are estimating a spurious geography-health correlation instead of the causal effect of upstream bathing on downstream diarrheal incidence, then we should also see association with these other diseases that are not waterborne. The absence of such effects would support the identifying assumption. Third, we add a range of time-varying control variables for changing demographic and poverty characteristics. As detailed in the results section, each of these tests supports our identifying assumption and thereby increases our confidence that we are estimating the causal effect of polluting behavior on diarrheal incidence.

Following equation (2), we also estimate avoidance behavior. In particular we test whether individuals reduce consumption of drinking water from the river in response to upstream polluting.

$$H(DrinkFromRiver)_{vpt} = \sum_{j} [\beta_{1j}Upstream_{jvpt} + \beta_{2j}Downstream_{jvpt}] + \gamma_v + \eta_{pt} + \epsilon_{vpt} \quad (3)$$

where $H(\cdot)$ is an indicator function equal to 1 if most people in that village drink water from the river.

Three additional econometric issues bear noting. First, there could be potential concerns

over the choice of our estimator. We find that our estimates are robust to different choices of estimator. In particular, we use the fixed effects logit estimator and find that our results are qualitatively similar, as reported in appendix table A.2. We also find that a mere 0.3% of our observations (433 out of 108,991) had predicted values outside the [0,1] range, suggesting that fit is not a concern with the use of a linear probability model. Second, we cluster our standard errors at the drainage basin to allow for errors to be correlated across villages along the same river segment and over time. Given that the pollutants accumulate along a river segment, clustering at the drainage basin allows for conservative inference on the effects of upstream polluting behavior. Third, given different populations across villages, we show in appendix table A.1 that our results are robust to population-weighted generalized least squares (GLS) estimation.

5   Results

We investigate the health effects of two kinds of non-industrial riparian polluting behaviors - bathing and trash disposal. We estimate the effects of these different polluting behaviors simultaneously using equation (2) and find strong evidence that upstream bathing causes increased diarrheal incidence. Specifically, we find that a one sample standard deviation increase in the number of people bathing upstream from a village (182,940 individuals) increases the probability of diarrheal outbreak in that village by 4.59 percentage points (table 3, column 2). Using the within-sample average of 18.29% diarrheal incidence, this corresponds to a 25.10 % effect.[16] The result is stable to the choice of specification (appendix

---

[16]We report the results per 100,000 individuals, which corresponds to a 2.52 percentage point effect or 13.7% effect.

table A.2) and limited to villages where the primary source of water is the river (table 3, column 2-3). The magnitude of the effect of upstream bathing on downstream diarrheal incidence is considerably larger when limiting our sample to only those villages where the primary source of drinking water is the river. Cumulatively yet conservatively, we estimate that upstream bathing can explain 865 deaths, which is 7.5% of all diarrheal deaths in our sample, suggesting that there is a large human cost to seemingly benign polluting behavior.

Notably, we find no evidence on the impact of upstream trash disposal on downstream diarrheal incidence. It is important to note that these are equilibrium effects, net of avoidance behavior specific to the form of upstream polluting behavior. As a result, the absence of effects on in-river trash-disposal could be evidence of avoidance behavior. Later in this section we will separately estimate the avoidance behavior in response to upstream polluting behavior.

Avoidance Behavior

The contrast between the effects of upstream bathing and upstream trash disposal is of interest. Since the effects in table 3 are net of avoidance behavior, these results suggest that downstream populations exhibit avoidance behavior in response to trash and not to pollutants entering the river through bathing in it. While we cannot explicitly or directly test the motivations behind this gap in avoidance behavior between the different sources of pollution, we hypothesize two plausible explanations. First, trash disposal in the river may pose a higher health risk than in-river bathing, and as such individuals are more active in avoiding the risk arising from the former. Yet, given the large health cost of upstream

bathing, this seems less likely. Instead we favor the second explanation, that trash disposal results in pollutants that are visible to the naked eye, in contrast to impurities generated from bathing that are less or not visible to individuals. This explanation is consistent with our results in table 4 that individuals stop drinking water from the river in response to upstream trash polluting but not in response to upstream bathing.

We can quantify the extent to which individuals engage in avoidance behavior in response to upstream polluting behavior (table 4). We find some evidence of avoidance behavior in response to upstream in-river trash disposal, but no evidence for avoidance behavior with respect to river bathing. Specifically, one standard deviation (37,853 individuals) increase in upstream bathers reduces river water consumption by 0.65 percentage points (p-value = 0.06), which corresponds to a 5% effect.[17] While we cannot provide specific behavioral or structural explanations for this gap in avoidance behavior, combined with the evidence on the net effects from table 3, we cautiously conclude that individuals exhibit greater avoidance behavior with respect to visible pollutants such as trash than less visible pollutants arising from bathing, making them potentially more lethal. This is consistent with a negligible net effect of upstream trash disposal on downstream diarrheal incidence.

Topographic and Geographic Factors

Next we isolate effects based on different geographies to deduce whether (1) the results are consistent with the topography of villages (flat versus hilly), which would affect the rate of water flow, and (2) there are systematic differences in the effects of upstream bathing

---

[17]In table 4, we report that that a 100,000 individuals engaging in upstream trash disposal reduces river water consumption by 1.71 percentage points, which corresponds to a 5% effect.

across urban and rural villages in our sample. In table 5 we provide results breaking down the effect of upstream bathing by topography (columns 2-3) and by urbanization (columns 4-5). Consistent with our intuition, we find that the bulk of the human health effect is concentrated in flat rather than hilly villages due to the propensity for eutrophication in stagnant water (Jiménez, 2006). By contrast, we find no evidence of differential impacts in rural versus urban locations, suggesting that avoidance behavior may not be region-specific.

We also address the concern that pollution dissipates over long distances along the river network. In the main specification, the $Upstream_{vt}$ indicator provides equal weight to each individual in a village that partakes in river bathing. However, it is possible that bathing activity for more distant villages may have little to no impact on disease outcomes if the contaminants dissipate over the course of the river. Given the method of riparian network construction discussed in Section 3, we can address this mechanism by weighting individuals in closer villages more than those who are farther away. To do this, we calculate an alternative Upstream (and Downstream) measure for bathing and trash behavior as:

$$Upstream_{vt} = \sum_{n_{vt}} [population_{nt} * bathing_{nt} * f(distance_{nt})] \tag{4}$$

where $distance_{nt}$ is the distance (in meters) of the centroid of village n from village v. The function $f(\cdot)$ weights the relative distance of the villages with the property that $f' \leq 0$.

Table 6 presents main regression results using this distance-weighting indicator.[18] The results are resemblant to those in table 3 in both sign and statistical significance. However, the distance-weighting results are larger by an order of magnitude, which is consistent with

---

[18]In this specification, we use the weighting function $f(distance) = 1/log(distance)$.

the fact that bathing populations farther upstream have a dissipated impact on downstream human health.

Placebo Tests

In this section we provide the results from a battery of placebo and falsification tests to demonstrate the strength of our identification strategy, and consequently the validity of our identifying assumption. Table 7 shows that the effect we find is (1) specific to diarrheal incidence and (2) specific to *upstream* polluting behavior. As shown in the table, we find no evidence of an effect on diarrheal incidence in a given village in response to downstream bathing. Not only is the effect of downstream bathing statistically indistinguishable from zero, it is also an order of magnitude smaller than the effect of upstream polluting. Therefore, concerns over drainage basin specific, time-varying factors that are correlated with both polluting behavior and diarrheal outcomes may be overstated.

Additionally, we follow Garg (2015) and provide placebo tests on other diseases (table 7, columns 2-5). Spurious correlation between downstream health (driven by demographic or political economy factors) and upstream polluting behavior should be shared when predicting the impact of upstream pollutants on other disease outbreaks. For instance, if individuals were geographically sorting over time in response to poor health conditions (but not to upstream bathing), then we should expect to see a correlation with incidence of at least one of the other diseases. The absence of any noticeable or meaningful effect on any of the other diseases adds support to the validity of our identification strategy.

We test for non-linearities in appendix table A.3 and find no evidence to support a non-

linear relationship between exposure to upstream river bathing and downstream diarrheal incidence.

Proxy for River Access

Another possible threat to identification is that the river bathing metric is acting as a proxy for river *access*. Villages in which the river is more accessible (e.g. not flowing through a deep gorge) may promote increased river use for bathing as well as for other purposes. While the placebo tests presented in table 7 cast doubt upon the presence of a non-waterbourne transmission mechanism, it does not rule out other pollution sources besides bathing, such as agriculture run-off, that may be driving the main result.

We address this by using a question from *Podes* regarding the village's river use. The question provides a binary indicator for whether the river is used for a host of economic activities: agriculture, industry, transportation, and "other" miscellaneous activities. We replicate the calculation from equation (2) using these respective indicators in place of the river bathing indicator. In this way, we focus on the total upstream population with unfettered substantial river access.

Table 8 presents the results of using these accessibility-to-river population indicators as a predictor of diarrheal outbreak. None of these activities alone (columns 1-3) have significant predictive power on the probability of a downstream diarrheal outbreak. This absence of a significant relationship between upstream river accessibility and downstream human health outcomes remains when considering all river access indicators jointly (columns 4-5). Even with the inclusion of trash dumping in column 6 we find no significant effects, and the

coefficient estimates are an order of magnitude smaller than the estimated effect of river bathing. As column 6 should display any of the predictive power related to river access but not river bathing, we can be cautiously optimistic that the effect that we find in table 3 is being driven by upstream river pollution related to bathing activities and not generically river access.

6   Policy Simulations

The impact of upstream bathing on downstream health suggests that policy responses to river pollution should be targeted with consideration to geography. Therefore, we conduct a set of policy simulations by imposing increasingly stringent moratoriums on river bathing (table 9).[19] Tailored policy options are particularly relevant in Indonesia and other developing countries where limited resources for enforcement require precision targeting of point-source pollution.

We show that the geography of targeting is essential to cost-effective policy. We categorize all sample villages into deciles based on total downriver population. Villages located near a river's headwaters with a large downstream population are grouped into the first decile while most downstream villages are grouped into the tenth decile (table 9). Targeting upstream villages generates the largest benefit - targeting the most upstream villages is an order of magnitude more effective than targeting downstream villages. Specifically, avoiding a single

---

[19]Simulations are cumulative across all four sample years (2000, 2003, 2006 and 2008) and the exposure variable (upstream bathing population) is updated by recalculating the sample mean with a moratorium imposed on villages that fall under the targeting rule. Epidemics are predicted using the rate of 0.0252 per 100,000 upstream individuals and deaths are predicted using the sample average of 0.58 diarrhea-related fatalities occurring within a village per diarrhea outbreak. Elasticity measures are calculated as the percentage change in diarrhea-related deaths divided by the percentage change in bathing individuals for each policy increment. Note that the marginal elasticities are not forced to be decreasing as they are calculated for incremental adjustments to the policy rather than as percentage changes from the no-regulation baseline (which would be necessarily decreasing in magnitude).

diarrheal death requires preventing 971,000 individuals in the most downstream decile from bathing but only 82,000 individuals in the most upstream decile. Our findings are therefore consistent with recent work on the political economy of water pollution (Lipscomb and Mobarak, 2016).

The baseline case, which most closely resembles the current state of affairs in Indonesia, has no regulation on river bathing activity. Population deciles with the largest downstream populations are then targeted incrementally until a complete moratorium on river bathing is achieved. Column 2 in table 9 shows the number of individuals bathing in the river in each decile.[20] In our two extreme cases, the absence of regulation on river bathing allows the 865 deaths attributable to river bathing to persist while a strict moratorium on river bathing prevents all of these deaths (table 9, column 3).

However, avoided deaths on decile-level moratoriums is not a comparable measure across the different deciles that have varying number of bathers. We generate two measures that allow us to compare moratoriums on different deciles - average and marginal number of individuals who must stop bathing to avoid a single instance of diarrheal mortality (table 9, columns 3 and 4 respectively). These cost calculations are akin to average and marginal costs of the policies per unit of benefit (figure 3). Columns (3) and (4) show that a policy-maker interested in reducing diarrheal deaths would have to inconvenience (or compensate) the fewest number of individuals per avoided death in the most upstream decile - 82,000 individuals who bathe in a river, versus 971,000 individuals in the most downstream decile. Conversely, in columns 5 and 6 (table 9), we develop an elasticity measure that shows reductions in mortality from a 1% reduction in decile-specific river bathers (figure 4). Reducing

---

[20]Deciles are constructed based on total population, not bathing-in-river populations.

top-decile bathers by 1% reduces marginal downstream diarrhea-related mortalities by 2.54% but reducing the lowest-decile bathers by 1% reduces marginal mortalities by only 1.61%.

7 Conclusion

In this paper, we construct and employ a novel data set on Indonesia's drainage basins to provide the first causal evidence that household-level polluting behavior and in particular upstream in-river bathing generates large downstream health externalities. Our results have particular relevance for policymakers for several reasons. First, we uncover a previously ignored source of household level pollution: in-river bathing. We find that upstream river bathing can explain as many as 865 deaths over four years representing 7.5% of all diarrheal deaths in our sample. This represents a large human cost from a source of river pollution that the literature has almost entirely ignored. Second, we find that targeting based on geographic location of the source of pollution can result in substantial health savings. In particular, a 1% decrease in in-river bathers in the most upstream decile reduces fatalities by 2.54%. By contrast, a 1% decrease in in-river bathers in the most downstream decile reduces fatalities by only 1.62%. Third, we find suggestive evidence that individuals exhibit avoidance behavior to visible but not invisible pollutants. If salience drives a wedge between the marginal benefit and marginal cost of avoidance behavior, investment in prevention of these "silent" killers may yield considerable health savings. Instead of large scale government programs aimed at river basin cleanup which may be financially infeasible, policymakers could attempt to enact policies aimed at preventing polluting behavior. Future work could explore the health impacts of avoidance behavior to previously understudied point and non-point sources of

water pollution.

References

AECEN. 2008. "Environmental Compliance and Enforcement in Indonesia: Rapid Assessment."

Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane. 2010. "Providing safe water: Evidence from randomized evaluations." *Annual Review of Resource Economics*, 2(1): 237–256.

Ashraf, Nava, James Berry, and Jesse M Shapiro. 2010. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." *American Economic Review*, 100: 2383–2413.

Barrett, Christopher B, Teevrat Garg, and Linden McBride. 2016. "Well-Being Dynamics and Poverty Traps." *Annual Review of Resource Economics*, 8(1).

Bartik, Timothy J. 1988. "Evaluating the benefits of non-marginal reductions in pollution using information on defensive expenditures." *Journal of Environmental Economics and Management*, 15(1): 111–127.

Bedner, Adriaan. 2010. "Consequences of decentralization: Environmental Impact Assessment and water pollution control in Indonesia." *Law & Policy*, 32(1): 38–60.

Bernhardt, A, and N Gysi. 2013. *Blacksmith Institute and Green Cross Switzerland.*

Bhatnagar, Anita, and P Sangwan. 2009. "Impact of mass bathing on water quality." *International Journal of Environmental Research*, 3(2): 247–252.

Braadbaart, Okke. 1995. "Regulatory strategies and rational polluters: industrial wastewater control in Indonesia, 1982-1992." *Third World Planning Review*, 17(4): 439.

Brollo, Fernanda, Tommaso Nannicini, Roberto Perotti, and Guido Tabellini. 2013. "The Political Resource Curse." *American Economic Review*, 103(5): 1759–1796.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh." *Econometrica*, 82(5): 1671–1748.

Burgess, Robin, Matthew Hansen, Benjamin A Olken, Peter Potapov, Stefanie Sieber, et al. 2012. "The Political Economy of Deforestation in the Tropics." *Quarterly Journal of Economics*, 127(4): 1707–1754.

Carson, Richard T, Phoebe Koundouri, and Céline Nauges. 2011. "Arsenic mitigation in Bangladesh: A household labor market approach." *American Journal of Agricultural Economics*, 93(2): 407–414.

Currie, Janet, Joshua Graff Zivin, Jamie Mullins, and Matthew Neidell. 2014. "What Do We Know About Short-and Long-Term Effects of Early-Life Exposure to Pollution?" *Annual Review of Resource Economics*, 6: 1–1.

Currie, Janet, Joshua Graff Zivin, Katherine Meckel, Matthew Neidell, and Wolfram Schlenker. 2013. "Something in the water: Contaminated drinking water and infant health." *Canadian Journal of Economics*, 46(3): 791–810.

Cutler, David, and Grant Miller. 2005. "The role of public health improvements in health advances: the twentieth-century United States." *Demography*, 42(1): 1–22.

Dasgupta, Purnamita. 2004. "Valuing health damages from water pollution in urban Delhi, India: a health production function approach." *Environment and Development Economics*, 9(1): 83–106.

Deschenes, Olivier, Michael Greenstone, and Joseph S Shapiro. 2012. "Defensive investments and the demand for air quality: Evidence from the nox budget program and ozone reductions." National Bureau of Economic Research.

Ebenstein, Avraham. 2012. "The consequences of industrialization: evidence from water pollution and digestive cancers in China." *Review of Economics and Statistics*, 94(1): 186–201.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2005. "Water for life: The impact of the privatization of water services on child mortality." *Journal of Political Economy*, 113(1): 83–120.

Garg, Teevrat. 2015. "Public health effects of ecosystem degradation: Evidence from deforestation in Indonesia." *Cornell University Working Paper*.

Graff Zivin, Joshua, and Matthew Neidell. 2013. "Environment, Health, and Human Capital." *Journal of Economic Literature*, 51(3): 689–730.

Graff Zivin, Joshua, Matthew Neidell, and Wolfram Schlenker. 2011. "Water Quality Violations and Avoidance Behavior: Evidence from Bottled Water Consumption." *American Economic Review*, 101(3): 448–453.

Greenstone, Michael, and B Kelsey Jack. 2015. "Envirodevonomics: A Research Agenda for an Emerging Field." *Journal of Economic Literature*, 53(1): 5–42.

Greenstone, Michael, and Rema Hanna. 2014. "Environmental Regulations, Air and Water Pollution, and Infant Mortality in India." *American Economic Review*, 104(10): 3038–3072.

GWP, Global Water Partnership. 2013. "National stakeholder consultations on water: supporting the post-2015 development agenda."

Jalan, Jyotsna, and Martin Ravallion. 2003. "Does piped water reduce diarrhea for children in rural India?" *Journal of Econometrics*, 112(1): 153–173.

Jiménez, Blanca. 2006. "Irrigation in developing countries using wastewater." *International Review for Environmental Strategies*, 6(2): 229–250.

Joshi, Namita, and Vandana Sati. 2011. "Assessment of Water Quality of River Ganges at Haridwar during Kumbh Mela-2010." *Report and Opinion*, 3(7): 30–36.

Kerstens, SM, G Hutton, A Van Nes, and I Firmansyah. 2013. "Downstream Impacts of Water Pollution in the Upper Citarum River, West Java, Indonesia Economic Assessment of Interventions to Improve Water Quality (p. 32)." *Jakarta, Indonesia.*

Kitazawa, Yoshitsugu. 2012. "Hyperbolic transformation and average elasticity in the framework of the fixed effects logit model." *Theoretical Economics Letters*, 2(2): 192.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix-Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation and Property Rights Institutions." *Quarterly Journal of Economics*, 126: 145–205.

Leggett, Christopher G, and Nancy E Bockstael. 2000. "Evidence of the effects of water quality on residential land prices." *Journal of Environmental Economics and Management*, 39(2): 121–144.

Lipscomb, Molly, and Ahmed Mushfiq Mobarak. 2016. "Decentralization and the political economy of water pollution: Evidence from the re-drawing of county borders in Brazil." *Forthcoming in Review of Economic Studies.*

Lucas, Anton, and Arief Djati. 2000. *The Dog is Dead So Throw it in the River: Environmental Politics and Water Pollution in Indonesia: an East Java Case Study.* Monash Asia Inst.

Nachmany, Michal, Samuel Fankhauser, Terry Townshend, Murray Collins, Tucker Landesman, Adam Matthews, Carolina Pavese, Katharina Rietig, Philip Schleifer, and Joana Setzer. 2014. "The GLOBE climate legislation study: a review of climate change legislation in 66 countries."

Resosudarmo, Budy P. 2003. "River water pollution in Indonesia: an input-output analysis." *International Journal of Environment and Sustainable Development*, 2(1): 62–77.

Scott, WE, RA Van Steenderen, DI Welch, Water Research Commission, et al. 1985. "Health aspects of eutrophication." *Available from the National Technical Information Service, Springfield VA. 22161, as PB 87-225645.*

Sharma, Vijay, Sushil Bhadula, and BD Joshi. 2012. "Impact of mass bathing on water quality of Ganga river during Maha Kumbh-2010." *Nature and Science*, 10(6): 1–5.

Tyagi, Vinay Kumar, Akanksha Bhatia, Rubia Zahid Gaur, Abid Ali Khan, Muntajir Ali, Anwar Khursheed, Absar Ahmad Kazmi, and Shang-Lien Lo. 2013. "Impairment in water quality of Ganges River and consequential health risks on account of mass ritualistic bathing." *Desalination and Water Treatment*, 51(10-12): 2121–2129.

WHO. 2002. "Eutrophication and health." *Luxembourg: Office for Official Publications of the European Communities. Geneva: WHO.*

WHO. 2014. "Preventing Diarrhea Through Better Water, Sanitation and Hygiene: Exposures and Impacts in Low- and Middle-Income Countries."

Wu, RSS. 1999. "Eutrophication, water borne pathogens and xenobiotic compounds: environmental risks and challenges." *Marine Pollution Bulletin*, 39(1): 11–22.
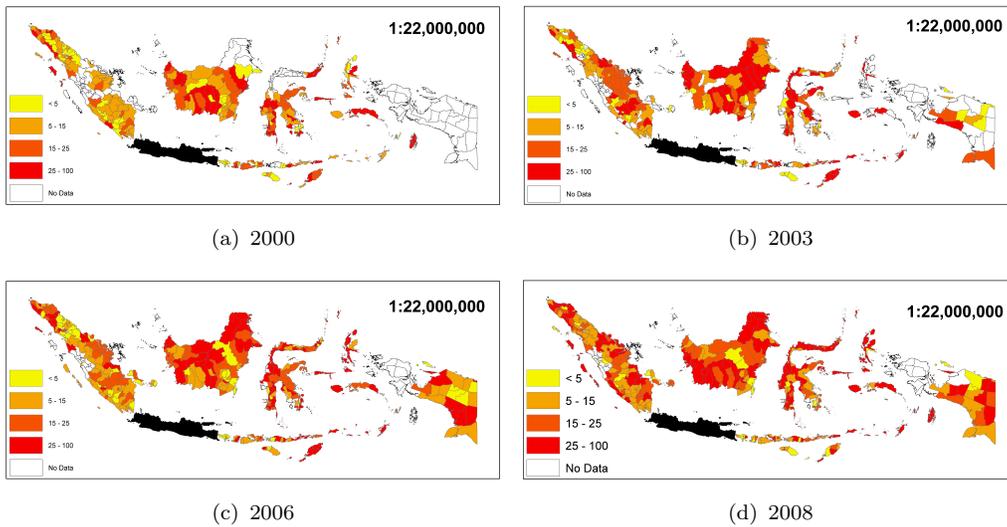
Figures

(a) 2000

(b) 2003

(c) 2006

(d) 2008

Figure 1: Percentage of Villages Declaring a Diarrheal Outbreak in a Year
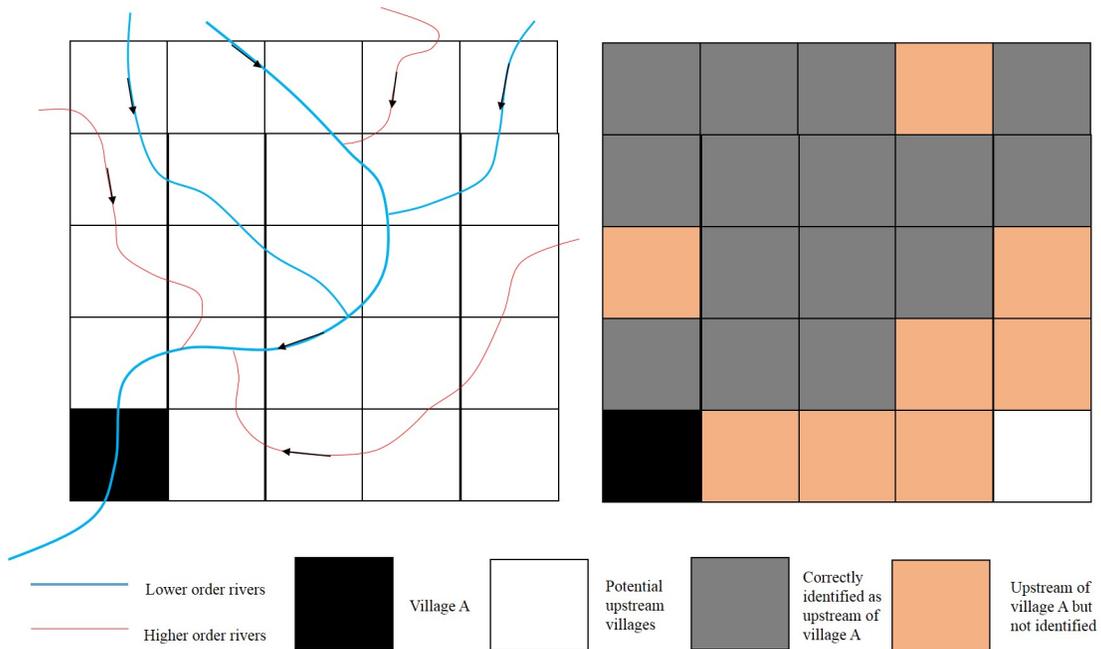
Figure 2: *Construction of River Network Data. (Left) Low and High Order Rivers (Right) Improved Assignment*
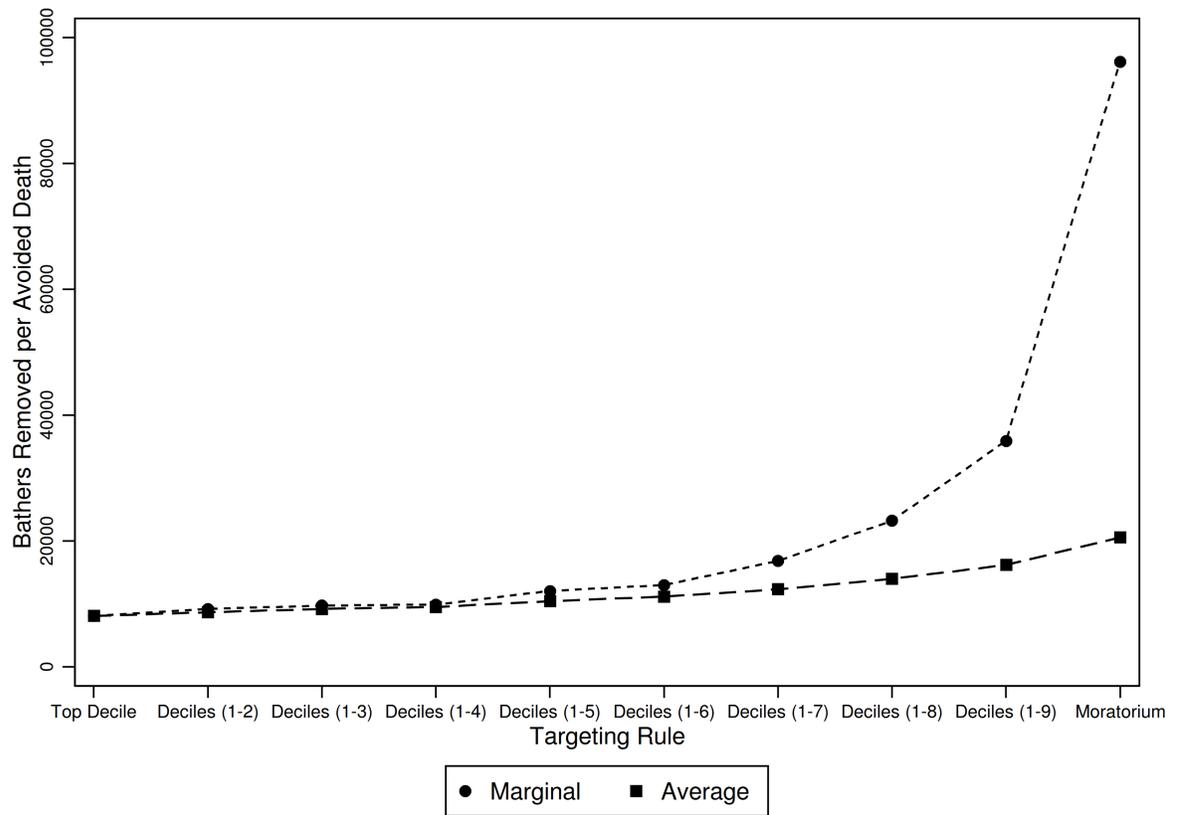
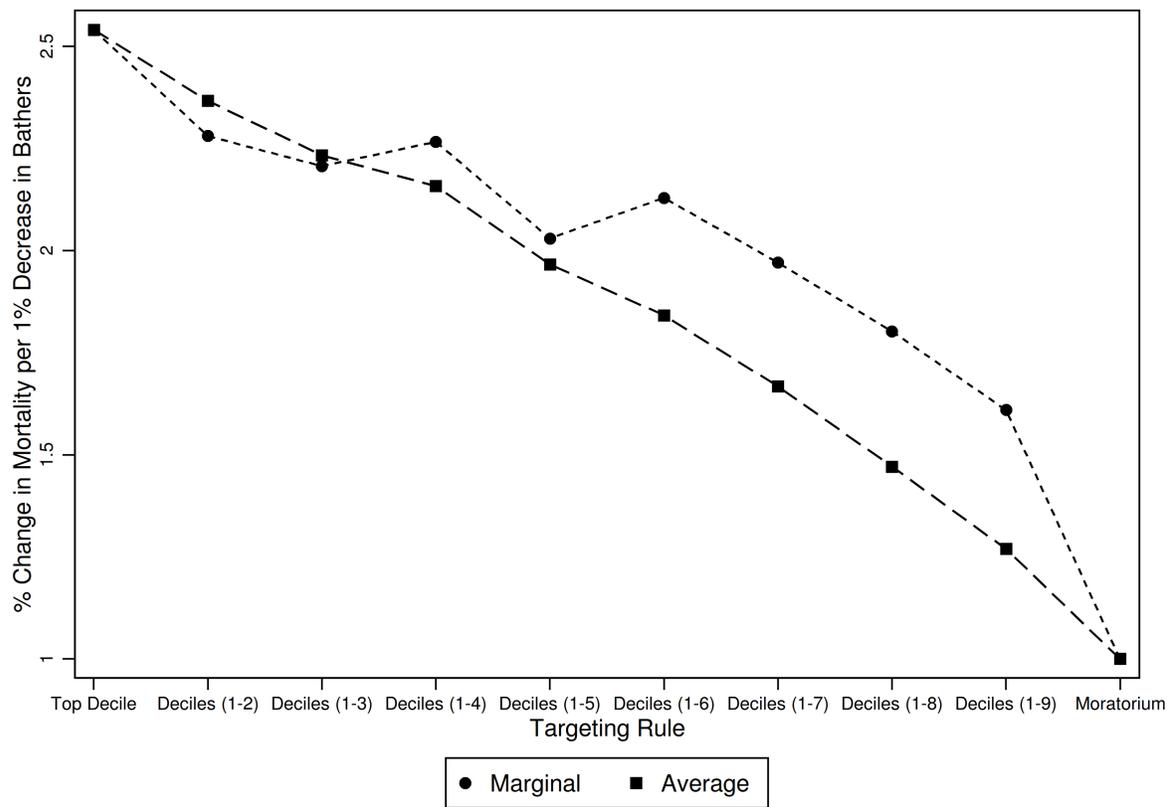Figure 3: *Marginal and Average Cost by Geographic Deciles*

Figure 4: *Percentage Reduction in Mortality Per 1% Decrease in Bathers by Targeting Rule*

Tables

Table 1: Probability of Village-Level Disease Outbreaks

	(1) Full	(2) Flat	(3) Hilly	(4) Urban	(5) Rural
Diarrhea	0.183	0.158	0.188	0.158	0.185
	(0.387)	(0.365)	(0.391)	(0.365)	(0.389)
Respiratory Infections	0.111	0.0978	0.118	0.101	0.112
	(0.315)	(0.297)	(0.323)	(0.301)	(0.316)
Measles	0.0807	0.0748	0.0866	0.0682	0.0820
	(0.272)	(0.263)	(0.281)	(0.252)	(0.274)
Malaria	0.207	0.165	0.231	0.127	0.216
	(0.405)	(0.371)	(0.421)	(0.333)	(0.411)
Dengue	0.0613	0.0698	0.0420	0.213	0.0454
	(0.240)	(0.255)	(0.201)	(0.409)	(0.208)
Observations	108,994	54,122	23,127	10,298	98,696

The mean incidence of the different diseases are statistically different from each other using the Fisher's LSD as well as the more conservative Tukey's test.

Table 2: Dumping Variables Summary Statistics

	Mean(sd)
Do most individuals bathe in the river?	0.791
	(0.407)
Do most individuals drink from the river?	0.343
	(0.475)
Do most individuals dispose trash in the river?	0.0915
	(0.288)
Observations	108,994

Table 3: Impact of Upstream Bathing on Downstream Diarrheal Incidence

Was there an outbreak of diarrhea?	(1) Full Sample	(2) Full Sample	(3) Drink from River	(4) Not Drink from River
Upstream Bathing	0.0225**	0.0252**	0.0564***	-0.000656
(100,000 individuals)	(0.0105)	(0.0107)	(0.0181)	(0.0132)
Upstream Trash	-0.00504	-0.00182	-0.0174	-0.00412
(100,000 individuals)	(0.0111)	(0.0120)	(0.0165)	(0.0133)
Observations	108,991	106,797	36,819	69,978
R-squared	0.012	0.013	0.023	0.012
Village FE	Yes	Yes	Yes	Yes
Province-Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. All specifications include additional controls for total village population, total upstream population and total downstream population. Columns (2) - (4) include additional controls: dominant source of income in village, geography of village, quality of governance (education of village head), access to medical facilities in the village and political status of the village.

Table 4: Avoidance Behavior

Did Villages Drink from River?	(1) Full	(2) Urban	(3) Rural
Upstream Bathing	0.00798	-0.00123	0.0123
(100,000 individuals)	(0.0111)	(0.0181)	(0.0127)
Upstream Trash	-0.0146*	-0.0114	-0.0161*
(100,000 individuals)	(0.00780)	(0.0114)	(0.00936)
Observations	108,991	10,298	98,693
R-squared	0.016	0.043	0.019

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. The specifications include additional controls for total village population, total upstream population and total downstream population. The sample is limited to villages that self-report proximity to a river.

Table 5: Results Disaggregated By Geography

Was there an outbreak of diarrhea?	(1) Full Sample	(2) Flat	(3) Hilly	(4) Urban	(5) Rural
Upstream Bathing	0.0225**	0.0249*	-0.0589	0.0193	0.0211*
(100,000 individuals)	(0.0105)	(0.0142)	(0.0529)	(0.0212)	(0.0117)
Downstream Bathing	0.00315	0.00280	-0.00000	0.00256	0.00579
(100,000 individuals)	(0.00250)	(0.00295)	(0.00923)	(0.00426)	(0.00360)
Observations	108,991	54,122	23,124	10,298	98,693
R-squared	0.012	0.012	0.030	0.047	0.013

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. The specifications include additional controls for total village population, total upstream population and total downstream population. The sample is limited to villages that self-report proximity to a river.

Table 6: Impact of Upstream Bathing Weighted by Distance

Was there an outbreak of diarrhea?	(1) Full Sample	(2) Full Sample	(3) Drink from River	(4) Not Drink from River
Upstream Bathing	0.205*	0.229**	0.580***	-0.0551
(100,000 individuals)	(0.111)	(0.114)	(0.198)	(0.141)
Upstream Trash	-0.0203	0.0280	-0.117	-0.0939
(100,000 individuals)	(0.115)	(0.127)	(0.190)	(0.137)
Observations	108,991	106,797	36,819	69,978
R-squared	0.012	0.013	0.023	0.012
Number of desa_id	41,502	41,319	19,057	33,318
Village FE	Yes	Yes	Yes	Yes
Province-Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. All specifications include additional controls for total village population, total upstream population and total downstream population. Columns (2) - (4) include additional controls: dominant source of income in village, geography of village, quality of governance (education of village head), access to medical facilities in the village and political status of the village.

Table 7: Placebo Tests

Was there an outbreak of	(1) Diarrhea	(2) Respiratory	(3) Measles	(4) Malaria	(5) Dengue
Upstream Bathing	0.0225**	0.00548	0.000182	0.0105	-0.00711
(100,000 individuals)	(0.0105)	(0.00772)	(0.00846)	(0.0116)	(0.00737)
Downstream Bathing	0.00315	0.000529	0.00322	0.00639***	-0.000797
(100,000 individuals)	(0.00250)	(0.00215)	(0.00229)	(0.00243)	(0.00264)
Observations	108,991	108,991	108,991	108,991	108,991
R-squared	0.012	0.011	0.010	0.015	0.026

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. The specifications include additional controls for total village population, total upstream population and total downstream population. The sample is limited to villages that self-report proximity to a river.

Table 8: Measures of River Access

Was there an outbreak of diarrhea?	(1) Full Sample	(2) Full Sample	(3) Full Sample	(4) Full Sample	(5) Full Sample	(6) Full Sample
Upstream Irrigation	-0.0218			-0.0194	-0.0198	-0.0193
(100,000 individuals)	(0.0140)			(0.0144)	(0.0146)	(0.0146)
Upstream Industry		-0.000799		0.0123	0.0107	0.0108
(100,000 individuals)		(0.00102)		(0.00797)	(0.0142)	(0.0141)
Upstream Transportation			-0.000522	-0.00734*	-0.00798*	-0.00784*
(100,000 individuals)			(0.000564)	(0.00444)	(0.00448)	(0.00454)
Upstream Other Activities					0.000960	0.000836
(100,000 individuals)					(0.00421)	(0.00423)
Upstream Trash						0.000514
(100,000 individuals)						(0.0122)
Observations	106,797	106,797	106,797	106,797	106,797	106,797
R-squared	0.013	0.013	0.013	0.013	0.013	0.013
Number of desa.id	41,319	41,319	41,319	41,319	41,319	41,319
Village FE	Yes	Yes	Yes	Yes	Yes	Yes
Province-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. All specifications include additional controls for total village population, total upstream population, total downstream population, dominant source of income in village, geography of village, quality of governance (education of village head), access to medical facilities in the village and political status of the village.

Table 9: Policy Simulations

Targeting Rule	(1) Bathing individuals impacted (millions)	(2) Predicted mortality reduction	(3) Average: Bathers removed per avoided death (thousands)	(4) Marginal: Bathers removed per avoided death (thousands)	(5) Average: % Change in mortality per 1% decrease in bathers	(6) Marginal: % Change in mortality per 1% decrease in bathers
No regulation of river bathing	0	0	-	-	-	-
Top decile only	2.4	29	82	82	2.54	2.54
Deciles (1-2)	5.5	63	88	93	2.36	2.28
Deciles (1-3)	11.5	124	93	98	2.23	2.21
Deciles (1-4)	21.8	227	96	100	2.15	2.27
Deciles (1-5)	37.8	358	106	122	1.96	2.03
Deciles (1-6)	55.8	495	113	132	1.84	2.13
Deciles (1-7)	77.6	623	125	170	1.66	1.97
Deciles (1-8)	103.7	734	141	235	1.47	1.80
Deciles (1-9)	133.6	817	164	363	1.26	1.61
Complete Moratorium on river bathing	179.5	864	208	971	1.00	1.00

Simulations include 108,990 village-year observations (2000, 2003, 2006 and 2008). Each 100,000 upstream bathing individuals increase epidemic rates by 0.0251 while each epidemic yields 0.58 deaths, which is the sample average for all on-river villages across all years (excluding Java). The targeting rule deciles are based on total downstream population. For example, the top decile includes only those top 10% of villages with the largest downstream populations. Column one shows that increasingly stringent moratoriums on river bathing increase the number of bathing individuals affected, which corresponds directly with a reduction in predicted mortality displayed in column 2. Columns 3 and 4 capture the number of bathers that must be removed to prevent a death where column 3 is the average effect across all deciles under the moratorium and column 4 is the marginal effect of the most downstream moratorium decile. Columns 5 and 6 present estimates of the elasticity of placing a moratorium on river bathing. Column 5 captures the change in mortality per 1% decrease in bathers, averaged over the deciles under the moratorium. Column 6 presents the effect of the change in mortality per 1% decrease in bathers for the marginal decile placed under moratorium. Here, elasticities are generally decreasing resembling the impact of targeting the most upstream and detrimental individuals. However, because the policy rule is based on cumulative downstream population along the river, not the downstream bathing population, it is possible for the marginal elasticity measure to increase with the addition of a new decile.

A Appendix

A.1 Validation of Outbreak Variable

In this section, we validate the outbreak variable with data on deaths. We do not use deaths in our primary specification since we have mortality data on only 4,681 village-years with more than one time period of data. In practice we have over 19,000 village-year observations, but over 15,000 villages with just a single year of data on deaths. In table A.1 we estimate two sets of regressions. In Column 1 we use, as the dependent variable, the log of the interaction of the outbreak variable with the ratio of average diarrheal deaths per unit population. This is equivalent to computing the effect on diarrheal outbreaks and then transforming the outbreak variable into the log of mortality rate. In column 2 we use the log of the actual death rate. The coefficient of interest and the corresponding placebo test have qualitatively the same coefficients. To the extent that column 1 and 2 are comparable, by using outbreak data instead of deaths data, we are likely underestimating the impact of upstream bathing on downstream health outcomes.

Table A.1: Validation of Outbreak Variable with Data on Deaths

	(1) Log Mean Death Rate per Epidemic	(2) Log Death Rate
Upstream Bathing (100,000 HH)	0.682***	0.672***
	(0.215)	(0.207)
Downstream Bathing (100,000 HH)	0.0699	0.0710
	(0.116)	(0.120)
Observations	19,933	19,933
R-squared	0.207	0.231
Number of Villages	15,252	15,252

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. The sample is limited to villages that self-report proximity to a river. Column (1) is defined as the log of the average number of diarrheal deaths divided by the average population interacted with the indicator variable denoting an outbreak in a village in a given year. Column (2) is the log of the death rate. Both regressions are weighted by population to account for differential populations across village and compute the average effect per person rather than per village.

A.2 Robustness to Choice of Estimator

In this section, we validate the choice of estimator. Column 1 presents the coefficients of a linear probability model (LPM) in predicting the diarrheal epidemic in a village in a given year. Using an LPM generates only 0.3% of predicted values that are outside the [0,1] range. Column 2 performs a similar estimation using a panel logit model instead of an LPM. The results of the estimation are qualitatively similar to the LPM regression, and maintains both the sign and level of significance for the bathing estimator.

Table A.2: Robustness to Choice of Estimator

	(1) LPM	(2) Logit
Upstream Bathing (100,000 individuals)	0.0225**	0.176**
	(0.0105)	(0.0713)
Downstream Bathing (100,000 individuals)	0.00315	0.0354
	(0.00250)	(0.0304)
Observations	108,991	41,332
R-squared	0.012	0.026

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. Both specifications include additional controls for total village population, total upstream population and total downstream population. Column (1) presents estimation using an OLS linear probability model, with approximately 0.3% of predicted values lying outside the [0,1] range. Column (2) presents the results of the fixed effects logit model. R-squared presented in the table is the psuedo-R2 calculated by the logit estimation. The number of observations using the FE Logit model is smaller than Column (1) due to all-positive or all-negative outcomes for village across all 4 years of the panel being dropped. The fixed effect logit model does not allow for the computation of the marginal effect at the mean value for the variables of interest - for a more detailed discussion, see Kitazawa (2012).

A.3 Testing for Non-Linearities

In this section, we test for whether there exists a non-linear relationship between upstream bathing behavior and diarrheal incidence in downstream villages. Column 1 presents the main result of the paper using the OLS estimator, and is identical to the corresponding column in table 3. Column 2 applies a quadratic fit to the bathing estimator, and finds no significant relationship between diarrheal outbreak and the square of upstream bathing populations. The third column runs an OLS regression using the log of upstream bathing values. This monotonic transform of the explanatory variable is qualitatively similar to the main specification, and although the estimator is an order of magnitude smaller, it still estimates a positive and significant relationship. Thus we find no evidence to support a non-linear relationship between exposure to upstream river bathing and downstream diarrheal incidence.

Table A.3: Testing for Non-Linearities

	(1) Linear	(2) Quadratic	(3) Logarithmic
Upstream Bathing (100,000 individuals)	0.0225**	0.0176	
	(0.0105)	(0.0150)	
Square [Upstream Bathing]		0.000428	
		(0.000536)	
Log[Upstream Bathing]			0.00490***
			(0.00114)
Observations	108,991	108,991	108,991
R-squared	0.012	0.012	0.013
Village FE	Yes	Yes	Yes
Province-Year FE	Yes	Yes	Yes
Controls	No	No	No

Cluster robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ All specifications include village and province-year fixed effects. Standard errors are clustered at the river basin levels. All specifications include additional controls for total village population, total upstream population and total downstream population. The sample is limited to villages that self-report proximity to a river. The mean number of upstream bathing households is 54,840.