Policy Research Working Paper 9143

The Nitrogen Legacy

The Long-Term Effects of Water Pollution on Human Capital

Esha Zaveri
Jason Russ
Sébastien Desbureaux
Richard Damania
Aude-Sophie Rodella
Giovanna Ribeiro

WORLD BANK GROUP
Water Global Practice
February 2020

Policy Research Working Paper 9143

Abstract

The fallout of nitrogen pollution is considered one of the largest global externalities facing the world, impacting air, water, soil, and human health. This paper combines data from the Demographic and Health Survey data set across India, Vietnam, and 33 African countries to analyze the causal links between pollution exposure experienced during the very earliest stages of life and later-life health. The results show that pollution exposure experienced in the critical years of development—from birth until age three—is associated with decreased height as an adult, a well-known indicator of overall health and productivity, and is robust to several statistical checks. Because adult height is related to education, labor productivity, and income, this also implies a loss of earning potential. The analysis begins within an assessment in India, where the data are more available, and is then extended to geographic settings including Vietnam and 33 countries in Africa. The results are consistent and show that early-life exposure to nitrogen pollution in water can lower height-for-age scores during childhood in Vietnam and during infancy in Africa. These findings add to the evidence on the enduring consequences of water pollution and identify a critical area for policy intervention.

This paper is a product of the Water Global Practice. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://www.worldbank.org/prwp. The authors may be contacted at ezaveri@worldbank.org, jruss@worldbank.org, rdamania@worldbank.org, sebdburo@gmail.com, and arodella@worldbank.org.

The Nitrogen Legacy: The Long-Term Effects of Water Pollution on Human Capital[1]

Esha Zaveri, Jason Russ, Sébastien Desbureaux, Richard Damania, Aude-Sophie Rodella, Giovanna Ribeiro

Keywords: Water Pollution, Health, Nitrogen

JEL Codes: O13, O15, Q15, Q25, Q54

[1] This research was part of the World Bank report titled Quality Unknown: The Invisible Crisis of Water Quality. We are grateful to Kelly Baker and seminar and conference participants at the International Food and Policy Research Institute (Delhi), Center for Policy Research (Delhi), Indian Statistical Institute (Delhi), the annual meetings of the Association of Environmental and Resource Economists (AERE) (2019), and the Workshop on Environmental Economics and Data Science (TWEEDS) (2019) for helpful feedback. Ali Sharman provided excellent research assistance. All errors are our own

1. Introduction

A hundred years since the ingenious experiments in nitrogen fixation by Haber and Bosch, which resulted in the development of the first nitrogen-based synthetic fertilizers, the fallout of nitrogen pollution is considered one of the most important environmental issues of the twenty-first century (Kanter, 2018). Recent studies suggest that nitrogen may be the world's largest global externality, due to its effects on human health and the environment (Keeler et al. 2016). The world has also surpassed the *planetary boundary* for nitrogen-- a level of human interference beyond which environmental damage increases dramatically and possibly permanently (Steffen et al. 2015).

In water, excess reactive nitrogen can promote the growth of algae, which can trigger toxic blooms that can kill fish, and nitrate in drinking water can harm human health. It is one of the few water pollutants that is trending upwards nearly everywhere, including in developed countries like the United States, despite strong regulation (Keiser and Shapiro 2018). The legacy effects of nitrogen pollution on the environment are likely to endure decades after nitrogen inputs have ceased, with long time lags between the adoption of conservation measures and any measurable improvements in water quality (Van Meter et al. 2018).[2] In humans, the health impacts can be acute, causing infant death due to methemoglobinemia, or the *blue baby syndrome,* that reduces the blood's ability to transport oxygen.[3] However, causal evidence for the long-term and legacy health impacts of early-life exposure to nitrogen pollution is still limited.

This paper addresses this gap by examining the impact of nitrate-nitrogen pollution on height and well-being in India, along with supporting evidence from Vietnam and 33 countries in Africa. India provides a compelling setting in which to study the impacts of water pollution caused by nitrate-nitrogen. The Green Revolution, starting in the mid-1960s, was a watershed moment in Indian agriculture. Along with a rapid increase in agricultural productivity, it also led to a dramatic rise in the consumption of synthetic nitrogenous fertilizers such as nitrogen-phosphate-potassium (NPK). The five-fold rise in the use of NPK fertilizers per hectare of cultivated land since the mid-1960s resulted in profound changes to the nitrogen cycle with impacts on India's waters — runoff of excess nitrogen from fields increased concentrations of nitrate in the waters to unsafe levels (Fields, 2004). But agriculture is just one of the sources of nitrogen pollution. According to the first-ever decade-long nitrogen assessment conducted for India by the Indian Nitrogen Group (ING), a voluntary body of more than 100 scientists and

---

[2] For instance, even if runoff of nitrogen was fully stemmed, it will still take 30 years to realize the 60% decrease in load needed to reduce eutrophication in the Gulf of Mexico (Van Meter, 2018).

[3] This health hazard was responsible for triggering the creation of drinking water standards for nitrates at 10 parts per million. Note that 10 mg/L as nitrate-nitrogen (NO_3-N) is approximately equivalent to the World Health Organization (WHO) guideline of 50 mg/L as NO_3.

other stakeholders, sewage and organic solid wastes are some of the fastest growing sources of nitrogen pollution in the country (INA, 2017).

Most previous work that provides causally interpretable estimates has primarily focused on short-run and immediate birth outcomes (Brainerd and Menon, 2014; Jones, 2019).[4] These studies show that early-life exposure to nitrogen-related pollution can lead to infant mortality (Brainerd and Menon, 2014) and low birthweight (Jones, 2019). The well-established fetal origins literature suggests that intrauterine health impacts can lead to lasting health damages, and that low birthweight is associated with shorter height in adulthood (Barker, 1990; Almond and Currie, 2011; Currie and Vogl, 2013; Christian et al., 2013; Almond, Currie and Duque, 2018). And low birthweight is a well-known marker for many health problems later in life, including coronary heart disease (Barker 1995), decreased glucose tolerance (and thus a higher propensity for obesity) (Ravelli et al. 1998), and increased rates of all-cause mortality (Risnes et al. 2011). Still, there has been no attempt to quantify the full extent of health damages, especially the irreversible and lagged human capital impacts, as a result of nitrogen pollution in water.

Evidence for lagged human capital impacts in environmentally vulnerable and poor locations is severely limited due to the paucity of longitudinal data that can trace long-term impacts. This is particularly challenging for water pollution, since monitoring of water quality is sparse in space and time and is site-specific. In this paper, we exploit temporal and geographic variation in nitrogen pollution exposure with a newly constructed database of water quality that combines in situ monitoring station data with a geospatial statistical model for stream networks developed by ver Hoef and Peterson (2010). We carefully integrate early-life exposure to nitrogen pollution between the critical years of development – from the period of birth up until year three – with women's health outcomes, climatic factors, correlated pollutants, household inputs and other socioeconomic demographics for our analysis.

Our research design exploits the direction of river flow and the upstream-downstream geographic relationship used in past literature (Do et al. 2018; Garg et al. 2016) to estimate a pollution-health relationship. Because the costs imposed by water pollution are largely felt in downstream regions, the analysis focuses on the impact of upstream pollution on health outcomes in downstream regions. To isolate the average pollution spillover at downstream locations, the analysis uses a rich set of controls. These are meant to control for time-invariant, location-specific characteristics such as local soil quality and natural resource endowments, as well as factors that vary by year and month, such as weather, and national trends in economic

[4] A number of biomedical and epidemiological studies in the United States and other countries have documented a relationship between agrichemical exposure and birth defects such as Down's syndrome and Spina Bifida, especially for children conceived during the crop-sowing months, and among children of agrichemical applicators who are consistently exposed to toxins.

output and technological development. The analysis also controls for time-varying factors that are specific to states to capture state-level policies. To ensure that later-life health outcomes are measured in the same location of conception and birth where exposure occurred, the sample is restricted to individuals who have never migrated from their place of birth. In this way, our empirical strategy controls for a wide number of potential confounders in an effort to identify causal effects. To test for external validity, similar analyses are conducted in geographies outside India - Vietnam and 33 countries in Africa.

Our results find that nitrogen exposure experienced by infants can have durable, long-term impacts that stretch well into adulthood. In India, women exposed to nitrogen pollution in their earliest years of life are shorter on average in adulthood than women of similar circumstances who were not exposed to such pollution. Early-life exposure to nitrogen pollution also lowers later-life labor productivity and depresses adult wages, decreasing overall welfare. This finding is robust to several sensitivity and falsification tests. Analyses across different geographic settings in Vietnam and Africa that measure the impact of nitrogen pollution during early-life provide further supporting evidence for the results found in India. Taken together, this paper provides new evidence that early-life exposure to nitrogen pollution has enduring and irreversible costs on human capital, with decreases in height observed across different life-stages: in adulthood (India), in childhood (Vietnam), and in infancy (Africa).

The rest of the paper is organized as follows. In Section 2 we describe the health and water quality data, the construction of the main variables used in the analysis as well as the procedure we use to match the health data to upstream pollution. Section 3 outlines our empirical strategy and Section 4 discusses the results. Robustness checks are provided in Section 5. Section 6 investigates the external validity of the results reported for India in other geographic settings such as Vietnam and Africa. Section 7 discusses the plausible mechanisms linking nitrogen pollution to height impacts, and Section 8 concludes.

2. Data

In this section, we describe the data sources that were used in the empirical analysis, and the construction of the main variables in the analysis.

2.1 Health Data

The data on our outcomes of interest come from the fourth round of the National Family and Health Surveys (NFHS) conducted in India. The NFHS is the Demographic and Health Survey (DHS) equivalent in India. The survey was conducted between January 2015 – December 2016 and covered all areas of the country. In a departure from the previous DHS surveys, the sample for this survey was designed to be representative at the district level. Close to 600,000 households were interviewed which included 0.7 million eligible women in the age group 15-49.

The main variables in this analysis come from the woman's questionnaire where a number of anthropometric measures are collected.

We make use of adult height as our main health variable. The micro-econometric literature often uses adult height as a proxy indicator for overall health and long-term adult well-being since it reflects the accumulation of shocks to health through childhood and adolescence. A rough consensus drawn from this literature is that an improvement in health associated with a 1-centimeter increase in adult height raises productivity by 3.4 percent (Kraay, 2018). Respondent's height is reported in centimeters in the DHS data.

The DHS also records how long the individual has resided in the current location. We utilize this information to restrict the sample to only those women whose birth-place coincides with the current location. This allows us to guard against the possibility of mis-measuring exposure to nitrogen pollution if the woman was born in a different location than the location in which she currently resides. We trace the birth-year histories of all adult women ranging from 1966 to 1999, a period when the effect of the Green Revolution was already in force yet nitrogen fertilizers were still increasing in use.

2.2 Water Quality Data

The Central Water Commission (CWC) within the Ministry of Water Resources monitors data on ambient water quality throughout the country. We compile and harmonize a rich data set of water pollution measurements between the years 1963-2017 along a network of 375 river monitoring stations throughout India. Many gaps in the data exist since the pollution measures are not consistently recorded over the entire sample time frame and the panel is unbalanced. Over the 1963-2017 period, 75% of water quality data are missing (60% between 1986-2017). To circumvent these problems, we build on a new class of spatial statistical network model for stream data to interpolate and fill in missing observations across monitor-year pairs (ver Hoef and Peterson, 2010). The model takes advantage of the fact that water quality in a station downstream depends on environmental conditions and human activities upstream of the station, and on the water quality "received" from upstream (directed network). Therefore, spatial covariates in a well-defined upstream area and the spatial dependence between observations based on stream distance allow to model water quality and predict it in unsampled locations. We train such model to fill-in missing observations. We then collapse monitor level observations to the district level. A more complete description of the model is provided in Appendix 1.

We focus on cumulative exposure to nitrate-nitrogen when the concentrations exceed safety thresholds of 10mg/l from the year of birth to age 3. Prior work suggests that the first 1,000 days of a child's life are the most critical for early childhood development and for determining whether a child will grow up stunted. It has also been shown that height at age three strongly predicts adult height (Maccini & Yang, 2009). Lower height-for-age scores can lead to severe consequences for cognitive development, overall health, and even socio-economic conditions that carry into adulthood.

We assign each woman a fractional measure of the share of years exposed to high levels of nitrate-nitrogen between the year of birth and age 3 in the district where she was born. Since districts have split over time, we use parent districts to allow comparability across time. We then compare later-life health outcomes among cohorts with more and less pollution exposure after accounting for a rich set of controls.

Even though direct measures of drinking water quality are unavailable, in-situ monitoring data serve as a reasonable proxy for proximate levels of nitrates in drinking water. This is because nitrates are notoriously expensive and difficult to clean out of water, and cannot be sufficiently treated using conventional methods.[5] Evidence from a slew of countries around the world, including Morocco, Niger, Nigeria, Senegal, India, Pakistan, Japan, Lebanon, Philippines, the Gaza Strip and Turkey, show that nitrates in drinking water often cross conventional safety thresholds (Ward et al., 2018).[6]

2.3 Additional Controls

We control for both average rainfall in millimeters and average temperature in degrees Celsius as these have been shown to impact adult outcomes (Maccini and Yang 2009, Fishman et al. 2019, Hyland and Russ 2019) and are known to also interact with nitrate loadings in waterways (Zheng et al. 2016). These are obtained from the Indian Meteorological Department. In some specifications, we also control for fecal coliforms from the CWC data set. They are an oft-used measure of domestic pollution, and are a major focus of water supply, sanitation, and hygiene (WASH) operations. It is measured as the "most probable number" of coliform organisms per 100 mL of water (MPN/100 ml, reported in thousands).

[5] Indeed, even in the United States, the percentage of public water systems that have violated safety limits for nitrates in drinking water has increased in the 15 year period between 1994 and 2009, due to the difficulty of coping with the rising nitrate pollution and the concomitant rise in the costs of water treatment (Ward et al. 2018).

[6] In Senegal, studies have recorded nitrate-nitrogen levels going beyond 40 mg/l, more than 4 times the safety limit for NO3-N. Extremely high levels of nitrate have also been reported in the Gaza Strip, where nitrate reached concentrations of 500 mg/L NO_3 in some areas (10 times the safety limit for NO3), and more than 50% of public-supply wells had nitrate concentrations above 45 mg/L NO_3. Other site-specific studies in India have found nitrates in drinking water supplies to be particularly high in rural areas, where average levels are reported to be between 46 mg/L NO_3 and 66.6 mg/L NO_3 with maximum levels exceeding 100 mg/L NO_3 in several regions.

2.4 Matching Health Data to Water Quality Data

The primary challenge to evaluate the pollution-health relationship is the endogeneity of pollution exposure. Pollution is not randomly assigned and is often the byproduct of productive activities. In the case of nitrates in the water, it is largely a byproduct of intensive agriculture and untreated urban waste. Thus, a naïve approach which examines impacts of local pollution on local impacts will likely conflate the positive effects of increased production with the negative externalities of water pollution, and underestimate the latter's effect. To circumvent this bias, we construct a measure of upstream pollution using the geography of river flow. Similar techniques to identify upstream-downstream relationships have been applied in recent economics literature (Garg et al., 2018' Do et al., 2018 and Keiser, 2018). This choice is predicated on the fact that the decision to pollute upstream is orthogonal to downstream health, while geography dictates that pollution flows downstream.

We make use of a digital elevation model from the Shuttle Radar Topography Mission (SRTM) mission to identify the direction of stream flow and to track upstream and downstream through surface waters in India. We link the districts in our sample to all other districts that are upstream from it as connected by the stream network as shown in Figure 1. Since water quality decays over time, we bound the distance between upstream and downstream district-pairs such that the upstream district is the closest upstream district, and up to 300 km apart. For any given downstream district, we then calculate the average concentrations of nitrate-nitrogen pollution in the upstream districts.

2.5 Data for Vietnam and Sub-Saharan Africa

To test for external validity of the results from India, we also measure the impact of nitrogen pollution on health in other geographic settings: Vietnam and 33 countries in Africa.

Data for Vietnam

The water quality data come from the Mekong River Commission (MRC), which collects data for four countries (Cambodia, Lao PDR, Thailand, and Vietnam) spanning the years 1985-2010, and cover the main tributaries of the Mekong River. Our health data come from the latest Vietnam Living Standards Survey (VLSS) of 1997–98 where we focus on the health outcomes of children aged 4 to 12 years. The VLSS 97–98 was a nationally representative survey that sampled almost 6,000 households across the country. For each member of the surveyed household, the survey contains information on gender, year of birth, age, and anthropometric outcomes. At the household level, information is available on the ethnicity of the household head and the province of residence. Our sample is restricted to those who have always resided in the same place. Because nitrate-nitrogen levels in Vietnam are relatively lower, exposure to nitrate pollution is examined at levels that are above the 75th percentile in the distribution, or roughly

2 mg/L. Following a similar methodology described in section 2.4, each VLSS commune is matched to its upstream pollution counterpart.

Data for Sub-Saharan Africa

We use data from 90 Demographic and Health Surveys (DHS) spanning a period of 23 years to account for all child and household variables presented in the analysis. Figure 2 shows the 31 countries included in the analysis from Sub-Saharan Africa, as well as Morocco and the Arab Republic of Egypt. The dots represent the approximate locations of the communities where households in the survey live. We focus on anthropometric measures of children up to 5 years of age. We convert children's heights into Z-scores using the WHO growth standards (WHO 2006). Doing so allows us to assess child height relative to well-nourished children of the same age and sex. For our main outcome variable, we use height-for-age Z-score (HAZ) and low HAZ (i.e. HAZ below -2), which reflects stunting. Water quality data come from a machine learning algorithm presented in Damania et al. (2019). Each birth record is then matched to nitrate pollution flowing from urban centers that are farther upstream. These urban centers were identified using data of urban agglomerations from Africapolis (OECD/SWAC (2018)).

3. Empirical Methods

To estimate the long-run health impacts of childhood exposure to nitrogen pollution, the research design exploits quasi-random variation in exposure to nitrogen pollution experienced by different birth cohorts in different districts. Specifically, the analysis compares height outcomes between exposed and non-exposed cohorts, controlling for average differences in these outcomes across birth years and across districts. The estimating equation for individual-level outcome Y of person i during time t and born in district d and state s is presented below.

$$Y_{idt} = \alpha + \beta {N^U}_{idt}^{[birth,3]} + \lambda X_{dst} + \gamma D_{id}^W + \rho_{mt} + \rho_s t + \mu_d + \epsilon_{idt} \qquad (1)$$

${N^U}_{idt}^{[birth,3]}$, where superscript U denotes upstream, is the fraction of years from the time of birth to age 3 that individual *i* was exposed to nitrate-nitrogen levels from upstream areas that exceeded safe limits in their birth district *d*. It serves as a measure of cumulative pollution exposure in early life during generally accepted critical periods for biological growth and development. These values are recorded from upstream districts exploiting the natural flow of rivers and the fact that pollution flows downstream even as the decision to pollute upstream is orthogonal to downstream health. In this way, we exploit quasi-random variation in pollution that originates upstream and yet flows downstream to other districts. The analysis then uses these spillovers to ascertain how much of the health impact persists in the next district downstream of pollution incidence.

The analysis compares later-life height among cohorts with duration of nitrate-nitrogen exposure, controlling for birth year, birth month, district fixed effects and state-trends. In this way, the analysis exploits within-district variation in birth timing relative to pollution exposure to identify β.

The birth-year and birth-month fixed effects (ρ_{mt}) are included to account for age effects in health outcomes as well as unobserved national or seasonal shocks such as macroeconomic conditions or seasonal weather patterns, which might otherwise confound the relationship between pollution exposure and height. Similarly, district fixed effects are included to control for any time-invariant unobservable differences between districts that can affect health. For example, access to local nutrition programs is one such factor that may be constant across individuals born in the same location. The analysis also includes state-trends ($\rho_s t$) to flexibly control for heterogeneous changes in demographic factors, technological progress in agriculture and other policies that differ across states.

A number of other district and household specific variables are included in the analysis. X_{dst} are a vector of district time-varying variables (include temperature and precipitation and concentrations of other water quality indicators like fecal coliform). $D_{ids}^{W}$are controls for household characteristics such as religion and caste that are salient to the Indian context. Lastly, we use cluster-robust standard errors to account for within-district clustering of errors and arbitrary correlation of observations across time. Our baseline specification, therefore, compares two women from the same district who are subjected to different levels of nitrate-nitrogen exposure based on their year of birth, over and above any unobserved shocks to height that vary by the year of birth, and any long-run trends (or annual patterns) in height in the state of birth.

Thus, following established statistical methods in applied economics, the relationship between water pollution and height (β) is identified by removing any confounding differences attributable to location and time. The reduced-form relationship provides a causal estimate of the health damages caused by downstream spillovers of pollution, adding to related work on pollution spillovers by Do et al. (2018), Garg et al., (2018), Keiser and Shapiro (2018), Lipscomb and Mobarak (2017) and Sigman (2002, 2005). Further, since the identification strategy uses multiple exposure events over time and space, it alleviates concerns that the results are being driven by confounding factors to health that may be correlated with single events.

So far, our estimation strategy allows us to quantify the persistence of water quality impacts in downstream districts by focusing on downstream spillovers. We are also interested in the within-district externality: to what extent does nitrogen pollution within a given district affect health outcomes in the same district? To address this question, we instrument local pollution concentrations in a given location and time with upstream concentrations. This effectively uses variations in local water quality that are induced by exogenous upstream concentrations. The validity of this approach rests on the assumption that river flow is unidirectional and pollution from far away distances affects health, but only through its effect on local pollution concentrations. The first-stage (equation (2)) and second stage (equation (3)) of the two-stage least squares strategy are presented below.

$$N_{idt}^{[birth,\,3]} = \alpha' + \beta' N_{idt}^{U,[birth,\,3]} + \lambda' X_{dt} + \gamma' D_{id}^{W} + \rho_{mt}{}' + \rho_s t' + \mu_d{}' + \epsilon_{idt} \qquad (2)$$

$$Y_{idt} = \alpha + \beta \widehat{N}_{idt}^{[birth,3]} + \lambda X_{dt} + \gamma D_{id}^{W} + \rho_{mt} + \rho_s t + \mu_d + \epsilon_{idt} \qquad (3)$$

Coefficient β in equation (3) gives us the estimated impact of pollution exposure on health in the average district. Together with the spillover health impact in downstream districts estimated in equation (1), we are able to measure the full external health costs imposed by pollution.

It is important to highlight that this paper does not include a structural model that describes the mechanism(s) for our baseline results. Therefore, we interpret our main result as a reduced form relationship between nitrogen pollution and adult health. We provide discussion on the possible mechanisms in Section 7.

4. Results

Summary statistics are provided in Table 1. About 3% of the sample experienced high levels of nitrate-nitrogen pollution in water (exceeding 10 mg/l) in the year of birth and on average women were exposed to high levels of nitrate-nitrogen pollution for 2% of their lives up to age three. Table 2 presents the main results from estimating equation (1). Column (1) presents results from the preferred specification. We find that exposure to nitrate pollution that exceeds safety standards over the entire period decreases height by 2.24 centimeters. At the mean fraction of early life exposed to pollution, this decrease in height is 0.5 centimeters. Columns (2), (3) and (4) use an indicator for high-level exposure in-utero, in the birth year and at age one rather than a cumulative measure of exposure. The results show a lowering of height with exposure but these effects are not significant compared to the effect from cumulative exposure in column (1). Column (5) includes an indicator for whether concentrations of fecally derived bacteria related to poor sanitation from upstream locations are above desired limits from the time of birth to age three, to confirm that it is not this correlated water quality indicator that is driving the result. Exposure to nitrogen pollution continues to be statistically significant, and the magnitude is even higher. This suggests that exposure to nitrogen pollution matters for health in addition to exposure from excreta-related bacteria. In Column (6), stricter control of birth month-by-birth year fixed effects are included to control for unobserved factors that are constant across all individuals born in the same year and month. Results are qualitatively similar. The results show that exposure to nitrate pollution that exceeds safety standards over the entire period decreases height by 1.96 centimeters.

So far, the results have focused on the persistence of water quality impacts in downstream areas by measuring the direct spillover externality imposed by upstream pollution. In Table 3, we provide the estimated impact of the within-district externality by measuring the impact of pollution on health in the same district using the 2SLS procedure outlined in equations (2) and (3).

The first-stage is strong across all columns and the upstream concentrations are significant. When local concentrations are instrumented with upstream pollution levels in the second-stage, all specifications yield statistically significant estimates of the pollution impact and the effect of nitrogen pollution in water on height is negative, and large in magnitude. Diagnostic statistics for instrument relevance such as the Kleibergen-Paap F (Kleibergen and Paap, 2006) statistic show that the instrument is very strong. The F-statistic exceeds the Stock-Yogo (Stock and Yogo, 2005) weak identification critical value for 10% maximal instrumental variables size.[7] The point estimates from the 2SLS procedure are relatively much larger than the corresponding downstream spillover impact in Table 2, supporting the logic that as water pollution decays with river flow and time, the downstream impacts are likely to be smaller in magnitude than the within-district health impact. The results show that exceeding the nitrate-nitrogen safety standards over the entire period decreases height by 2.81 centimeters.

Because adult height is associated with income, this implies a productivity loss of around 7% using decrease in height estimates under full exposure derived from column 1 and using estimates of the economic returns to height assumed in the World Bank Human Capital Project (Kraay, 2018). When using estimates of a decrease in height under mean exposure derived from column 1, this translates into a 1.7% fall in productivity or earning potential.[8]

5. Robustness Checks

We carry out several robustness exercises to further corroborate our baseline results.

In order to examine the possibility that these results are driven by spurious spatial or temporal patterns, the analysis is subjected to falsification tests. The first test involves re-estimating equation (1) while replacing each individual's exposure condition with exposures that occur for 6 different four-year periods before or after birth up to age 9. The resulting coefficient estimates are plotted in Figure 3 against the different window periods of exposure. All the

[7] Baum et al. (2007) and Bazzi and Clemens (2013) provide explanations of these tests.

[8] As a robustness check, we also make use of the Indian Human Development Survey (IHDS) to measure the impact of early-life exposure to nitrogen pollution on later-life wages using a similar methodology described in Section 2. The IHDS is a nationally representative survey. The survey provided a more complete recording of men's earnings that we use as our main variable of interest. In unreported results, we find that full exposure in early-life decreases wages in adult life by 9%, on average, providing direct proof of the impact on labor productivity.

“shifted” coefficients are smaller than the “true” coefficient, plotted at 0-3, and are all statistically insignificant.

The second test involves replacing the upstream pollution variable with a falsified value using pollution data from the nearest off-river region farther downstream—a location that is disconnected from river flow dynamics and from where the pollution cannot flow (upstream) to areas where the health outcomes are measured. In the case that the ‘falsified’ upstream pollution variable shows a significant impact on health, then it would be likely that our baseline results are capturing spurious spatial correlations. Table 4, however, reveals otherwise. There is no significant impact of the falsified value on health, suggesting that the upstream variable utilized in the analysis is indeed isolating quasi-random variation in pollution.

In Table 5, in addition to the district fixed effect, we also include district time trends to address the concern that broad secular trends at the district level might be influencing our results. The results are of the same sign and magnitude as our baseline estimates, and remain significant at the 5 percent level.

In Table 6, we cluster the standard errors by state in DHS, as well as survey cluster in DHS, instead of district. Standard errors are more or less similar using either of these alternatives, and the results remain significant and unchanged.

6. Evidence from Other Regions

In Vietnam, home to one of the fastest-growing and urbanizing societies in the world, agricultural growth and intensification have played significant roles in spurring development. But in parts of the country, the environmental footprint of the agricultural sector is deepening. In intensively farmed areas, agriculture has become a significant contributor to water pollution. This is particularly so in the intensively farmed Mekong delta region (Cassou, Jaffee, and Ru 2018; Chea, Grenouillet, and Lek 2016). To measure the consequences of nitrogen pollution, the analysis focuses on children aged four to twelve years surveyed in the latest Living Standards Survey of 1997–98. Table 7 shows that repeated exposure to nitrate pollution for the first three years of life substantially lowers height-for-age scores in childhood, with full exposure decreasing height-for-age scores by 0.7 standard deviation. These effects occur despite nitrate-nitrogen concentrations being below the recommended safety thresholds of 10 mg/L and emerge even after accounting for exposures from other contaminants.

In Africa, although present-day fertilizer usage is lower than in Asia, it is growing. Other sources of nitrate exposure include expanding urban centers that lack wastewater treatment facilities and increased livestock farming. The analysis is based on the entire universe of child records up to age 5 years across 33 countries in Africa from DHS records. The results in Table 8 show that in utero exposure to nitrate pollution emanating from upstream urban agglomerations lowers the height-for-age scores and increases the likelihood of stunting for children younger than five years, even at low levels of nitrate exposure. The negative effects are most pronounced downstream from urban centers where nitrate levels are relatively higher. Stunting already remains a widespread problem in Sub-Saharan Africa, where more than 35 percent of children younger than five years are considered stunted (World Development Indicators). This suggests an urgent need for potable water treatment in urban agglomerations.

7. Mechanisms

These results are perhaps the first demonstration of such widespread links between exposure to elevated nitrate levels during early-life and long-run health outcomes. Nevertheless, they are consistent with several well-established streams of biomedical literature that are indicative of such a link. First, increased dietary-nitrate intake has been associated with hypothyroidism and thyroid cancer (Aschebrook-Kilfoy et al. 2012; Ward et al. 2010, 2018). The thyroid is an important gland for regulating hormone production and metabolism regulation. Hypothyroidism in children is therefore linked to stunting of growth and a delay in the process of maturation (Wilkins 1953). Thus, the path from increased nitrate consumption from water, to diseases of the thyroid, to stunted growth and development is seemingly clear and direct. Another potential causal link between nitrates in water and reduced health and growth is through the buildup of algae and bacteria in water. Nitrogen in waterways often causes cyanobacteria fueled algal blooms. These bacteria can emit cyanotoxins that are toxic to humans and, if consumed, can lead to diarrhea-related illnesses. Repeated bouts of diarrhea increase the probability of nutritional deficiencies in children and thus stunted child development. Exposure to such toxins can also adversely affect birth outcomes by lowering infant birthweight (Jones 2019), an important predictor of stunting later in childhood (Christian et al. 2013).

Finally, and related to the prior point, exposure to higher levels of pathogens can disrupt the gut microbiome. The first months after birth are particularly critical for establishing the composition of the gut microbiome that persists for the rest of a person's life (Robertson et al. 2019). There is evidence in the medical literature that this microbiome is difficult to permanently change later in life, although this matter is under debate. If true, then the resulting change in gut microbiome from exposure to nitrate-induced toxins like those from cyanobacteria could permanently handicap the digestive system of individuals and reduce their capacity to absorb nutrients throughout their lives. However, more research is required on how and when exposure of fetuses and young children to high nitrate levels influence microbiome function, growth, and development, especially in settings in which pathogenic infections and food insecurity are problematic.

8. Conclusion

Recent studies have focused attention on the loss of human lives and immediate birth outcomes as a result of water pollution. In a departure from previous studies, this paper underscores the long-lasting health damages, and decreased economic capability that survivors of water pollution endure. We find a statistically significant negative effect of early-life exposure to nitrogen pollution on women's height in India, with supporting evidence of a decrease in child height in Vietnam, and infant HAZ scores and increased incidence of stunting in Africa. These results are robust to several checks for confounding factors. By demonstrating the long-term effects of nitrogen pollution, our results draw attention to the critical role that local environmental spillovers play for population health outcomes and highlight the need for closer policy attention to nitrogen pollution.

The policy relevance of our results is underscored by the fact that health effects also emerge at levels well below prescribed limits, raising questions about what constitutes safe standards for nitrates in water. Emerging evidence from epidemiological studies has also found relationships between nitrate ingestion and cancer, thyroid disease, and adverse pregnancy outcomes, such as neural tube defects, at concentrations below regulatory limits (Temkin et al., forthcoming; Ward et al. 2018). Even as far back as 1977, a report by the U.S. National Academy of Sciences warned that "there is little margin of safety" in the 10 mg/L safety limit for nitrates (National Research Council 1977). More research and assessments across even more geographies and populations are needed to make definitive claims. It is possible that future research will uncover even more health effects as more data become available to link exposures that began decades ago to diseases that develop today. However, the body of evidence so far suggests that there still remains a great deal of uncertainty surrounding drinking water standards for nitrates set by environmental agencies. The magnitude of people impacted by nitrate contaminated water is, therefore, likely to be much larger than presently thought.

This work also speaks to the consequentiality of fertilizer subsidies in developing countries that are tipped in the favor of nitrogen fertilizer use. In India, a system of domestic price controls by way of large subsidies has significantly distorted market prices for nitrogen fertilizer compared to other nutrients, resulting in an inefficient balance of fertilizer application (Gulati and Banerjee 2015). By 2015, subsidy costs amounted to $11.6 billion per year in India, roughly five times more than what was recorded 15 years earlier (Gulati and Banerjee 2015). This is exemplified by the wide gap between global and Indian domestic nitrogen prices ---world prices were almost four times higher than regulated Indian prices in 2014 (Huang, Gulati, and Gregory 2017). In recent years the government has made efforts to improve nitrogen use efficiency in agriculture and has mandated urea manufacturers to produce neem-coated urea. Since neem acts as a nitrification inhibitor, it allows a more gradual release of nitrogen into the soil, thereby improving nitrogen use efficiency. More research is needed to quantify the environmental and economic consequences of such measures, and their impacts on water pollution.

Finally, unlike much of the literature on water quality and health that focuses on developed countries, this work adds to the growing evidence on water pollution impacts in the developing world, which is subject to different exposure profiles, institutions and levels of economic development. While our analysis controlled for correlated pollutants where possible, it was primarily focused on a single pollutant. It is possible that the combined health impacts of co-occurring pollutants are different or even more harmful (Stoiber et al. 2019). More work is needed to investigate these issues in the developing world.

References

Almond, D., and J. Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25 (3): 153–72.

Almond, D., J. Currie, and V. Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." Journal of Economic Literature 56 (4): 1360–446.

Aschebrook-Kilfoy, B., S. L. Heltshe, J. R. Nuckols, M. M. Sabra, A. R. Shuldiner, B. D. Mitchell, M. Airola, T. R. Holford, Y. Zhang, and M. H. Ward. 2012. "Modeled Nitrate Levels in Well Water Supplies and Prevalence of Abnormal Thyroid Conditions among the Old Order Amish in Pennsylvania." Environmental Health 11 (1): 6.

Barker DJ (1990) The fetal and infant origins of adult disease. BMJ 301:1111.

Barker, D. J. (1995). Fetal origins of coronary heart disease. Bmj, 311(6998), 171-174.

Brainerd, Elizabeth and Menon, Nidhiya, 2014. Seasonal effects of Water quality: the hidden costs of the Green revolution to infant and child health in India. Journal of Development Economics. 107, 49–64.

Cassou, E., S. M. Jaffee, and J. Ru. 2018. The Challenge of Agricultural Pollution: Evidence from China, Vietnam, and the Philippines. Washington, DC: World Bank.

Chea, R., G. Grenouillet, and S. Lek. 2016. "Evidence of Water Quality Degradation in Lower Mekong Basin Revealed by Self-Organizing Map." PloS ONE 11 (1): e0145527.

Christian, P., S. E, Lee, M. D. Angel, L. S. Adair, S. E. Arifeen, P. Ashorn, F. C. Barros, C. H. Fall, W.W. Fawzi, W. Hao, and G. Hu. 2013. "Risk of Childhood Undernutrition Related to Small-for-Gestational Age and Preterm Birth in Low- and Middle-Income Countries." International Journal of Epidemiology 42 (5): 1340–55.

Currie, J. and Vogl, T. 2013. Early-life health and adult circumstance in developing countries, Annual Review of Economics, Annual Reviews, vol. 5(1), pp. 1–36, 05

Fields, S. 2004. Global nitrogen: cycling out of control. Environ Health Perspect 112(10):A556-63.

Damania, R., Desbureaux, S., Rodella, A. S., Russ, J., & Zaveri, E. (2019). Quality Unknown: The Invisible Water Crisis. World Bank: Washington, DC.

Do, Quy-Toan, Shareen Joshi, and Samuel Stolper. 2018. Can environmental policy reduce infant mortality? Evidence from the Ganga Pollution Cases. Journal of Development Economics 133 : 306-325.

Fishman, R., Carrillo, P., & Russ, J. (2019). Long-term impacts of exposure to high temperatures on human capital and economic productivity. Journal of Environmental Economics and Management, 93, 221-238.

Garg, Teevrat, Hamilton, Stuart, Hochard, Jacob, Plous, Evan, Talbot, John, 2016. Not So Gently Down the Stream: River Pollution and Health in Indonesia. Grantham Research Institute on Climate Change and the Environment. Working Paper #234.

Gulati, A., and P. Banerjee. 2015. "Rationalizing Fertilizer Subsidy in India: Key Issues and Policy Options." Working Paper 307, Indian Council for Research on International Economic Relations, New Delhi. https://www.academia.edu/14970315/Rationalizing_Fertilizer_Subsidy_in_India-Key_Issues_and_Policy_Options.

Huang, J., A. Gulati, and I. Gregory, editors. 2017. *Fertilizer Subsidies: Which Way Forward?* IFDC/FAI report. https://ifdcorg.files.wordpress.com/2017/02/fertilizer-subsidieswhich-way-forward-2-21-2017.pdf.

Hyland, M., & Russ, J. (2019). Water as destiny–The long-term impacts of drought in sub-Saharan Africa. World Development, 115, 30-45.

The Indian Nitrogen Assessment (INA). 2017. *Eds* Abrol, Y. P. *et al*.

Jones, B. A. 2019. "Infant Health Impacts of Freshwater Algal Blooms: Evidence from an Invasive Species Natural Experiment." Journal of Environmental Economics and Management 96: 36–59.

Kanter, D. R. 2018. "Nitrogen Pollution: A Key Building Block for Addressing Climate Change." Climatic Change 147 (1–2): 11–21.

Keeler, B. L., J. D. Gourevitch, S. Polasky, F. Isbell, C. W. Tessum, J. D. Hill, and J. D. Marshall. 2016. "The Social Costs of Nitrogen." Science Advances 2 (10): e1600219.

Keiser D, The missing benefits of clean water and the role of mismeasured pollution data. 2019. J Assoc Environ Resour Econ 6 (4): 669-707.

Keiser, D.A., and J.S. Shapiro. 2018. "Consequences of the Clean Water Act and the Demand for Water Quality." *Quarterly Journal of Economics* 134 (1): 349–96.

Kleibergen, Frank and Richard Paap. 2006. Generalized reduced rank tests using the singular value decomposition. Journal of Econometrics 133 (1): 97–126.

Kraay, Aart C.. 2018. Methodology for a World Bank Human Capital Index (English). Policy Research working paper; no. WPS 8593. Washington, D.C. : World Bank Group.

Peterson, Erin E., and Jay M. Ver Hoef. 2010. A mixed-model moving-average approach to geostatistical modeling in stream networks. Ecology 91, no. 3: 644-651.

Lipscomb, Molly, Moshfiq Mobarak, A., 2017. Decentralization and Water pollution spillovers: evidence from the redrawing of county boundaries in Brazil. Rev. Econ.Stud. 84 (1), 464–502.

Maccini, S. L., & Yang, D. (2009). Under the weather: Health, schooling, and economic consequences of early-life rainfall. American Economic Review, 99(3), 1006–1026.

OECD/SWAC (2018), Africapolis (database), www.africapolis.org (accessed 05 February 2019)

Ravelli, A. C., van der Meulen, J. H., Michels, R. P. J., Osmond, C., Barker, D. J., Hales, C. N., & Bleker, O. P. (1998). Glucose tolerance in adults after prenatal exposure to famine. The Lancet, 351(9097), 173-177.

Risnes, K. R., Vatten, L. J., Baker, J. L., Jameson, K., Sovio, U., Kajantie, E., ... & Sundh, V. (2011). Birthweight and mortality in adulthood: a systematic review and meta-analysis. International journal of epidemiology, 40(3), 647-661.

Sigman, Hilary, 2002. International spillovers and Water quality in rivers: do countries free ride? Am. Econ. Rev. Pap. Proceed. 92 (4), 1152–1159.

Sigman, Hilary, 2005. Transboundary spillovers and decentralization of environmental policies. J. Environ. Econ. Manag. 50, 82–101.

Steffen, W., K. Richardson, J. Rockstrom, S.E. Cornell, I. Fetzer, E.M. Bennett, R. Biggs, et al. 2015. “Planetary Boundaries: Guiding Human Development on a Changing Planet.” *Science* 347: 736.

Stock, James H. and Motohiro Yogo. 2005. Testing for Weak Instruments in Linear IV Regression. In Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg, ed. Donald W. K. Andrews and James H. Stock. Cambridge University Press.

Stoiber, T., A. Temkin, D. Andrews, C. Campbell, and O. V. Naidenko. 2019. Applying a Cumulative Risk Framework to Drinking Water Assessment: A Commentary.” *Environmental Health* 18 (1): 37

Temkin, A., S. Evans, T. Manidis, C. Campbell, and O. V. Naidenko. Forthcoming. “ Exposure-Based Assessment and Economic Valuation of Adverse Birth Outcomes and Cancer Risk Due to Nitrate in United States Drinking Water.” *Environmental Research.*

Van Meter, K. J., Van Cappellen, P., & Basu, N. B. (2018). Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico. *Science*, *360*(6387), 427-430.

Wilkins, L. 1953. “Disturbances in Growth.” Bulletin of the New York Academy of Medicine 29 (4): 280.

Ward, M., R. Jones, J. Brender, T. de Kok, P. Weyer, B. Nolan, C. M. Villanueva, and S. van Breda. 2018. “Drinking Water Nitrate and Human Health: An Updated Review.” International Journal of Environmental Research and Public Health 15 (7): 1557.

Ward, M. H., B. A. Kilfoy, P. J. Weyer, K. E. Anderson, A. R. Folsom, and J. R. Cerhan. 2010. “Nitrate Intake and the Risk of Thyroid Cancer and Thyroid Disease.” Epidemiology (Cambridge, Mass.) 21 (3): 389.

Zheng, L., Cardenas, M. B., & Wang, L. (2016). Temperature effects on nitrogen cycling and nitrate removal-production efficiency in bed form-induced hyporheic zones. Journal of Geophysical Research: Biogeosciences, 121(4), 1086-1103.

Tables

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Height (cm)	151.584	6.315048	80	209.2
Mean upstream nitrate-N concentrations	1.776662	2.887341	0	20.18191
1[Exceedance of nitrate-N in year of birth]	0.0257691	0.1584502	0	1
Fraction of early childhood nitrate-N exposure	0.0190851	0.0732628	0	0.5
Annual Precipitation (mm)	882.8155	500.6614	99.29888	4463.666
Average temperature in the Wet Season	28.57736	1.785982	22.61299	32.91105
Average temperature in the Dry Season	20.75181	2.642523	13.20256	27.45295

Notes: Table shows descriptive statistics from the DHS surveys as well as the water quality data from CWC. Sample based on 19,138 respondents who have not migrated from their birth place.

Table 2: The long-term impacts of upstream pollution on health

	Dependent variable: Height (cm)					
	(1)	(2)	(3)	(4)	(5)	(6)
Fraction early childhood N exposure	-2.246***				-3.044***	-1.963***
	(0.497)				(0.996)	(0.506)
Exposure in-utero		0.541				
		(0.463)				
Exposure at birth			-0.385			
			(0.458)			
Exposure at age 1				-0.411		
				(0.392)		
Observations	19138	17399	17618	17417	13862	19138
mean Dependent Variable	151.6	151.6	151.6	151.7	151.4	151.6
R-sq	0.0793	0.0812	0.0795	0.0769	0.0656	0.0908
RMSE	6.082	6.076	6.093	6.114	6.089	6.046
Birth-year Fixed Effects	Y	Y	Y	Y	Y	
Birth-Month Fixed Effects	Y	Y	Y	Y	Y	
District Fixed Effects	Y	Y	Y	Y	Y	Y
State Trends	Y	Y	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y	Y	Y
Fraction early childhood FColi exposure					Y	
Birth-Year by Month Fixed Effects						Y

Notes: Table shows results from estimating Eq. (1) via ordinary least squares (OLS). Each column displays estimates from a separate regression. Fraction of early childhood exposed to N pollution is the fraction of years from year of birth to age 3 that nitrate pollution exceeds safety guidelines. Standard errors are clustered at the district level, and are presented in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.

Table 3: The long-term impacts of local pollution on health using instrumental variables

	(1)	(2)	(3)	(4)
	First-stage	Second-stage	First-stage	Second-stage
Upstream: Fraction early childhood N exposure	0.748***		0.745***	
	(0.166)		(0.163)	
Local: Fraction early childhood N exposure		-2.819***		-2.604***
		(0.645)		(0.634)
Observations	17755	17755	17755	17755
mean Dependent Variable		151.6		151.6
R-sq		0.0186		0.0187
RMSE	0.0407	5.956	0.0407	5.921
Kleibergen-Papp F-stat		20.34		20.94
		(F=16.38)		(F=16.38)
Birth-year Fixed Effects	Y	Y		
Birth-Month Fixed Effects	Y	Y		
District Fixed Effects	Y	Y	Y	Y
State Trends	Y	Y	Y	Y
Weather controls	Y	Y	Y	Y
Birth-Year by Birth-Month Fixed Effects			Y	Y

Notes: Table shows results from estimating Eq. (2) and Eq. (3) using Two-Stage Least Squares (2SLS). Each column displays estimates from a separate regression. Fraction of early childhood exposed to N pollution is the fraction of years from year of birth to age 3 that nitrate pollution exceeds safety guidelines. Columns 2 and 4 show 2nd-stage results and columns 1 and 3 show 1st-stage results. The endogenous variable(Local: Fraction early childhood N exposure) is instrumented using its upstream analog. For Kleibergen-Paap rkWald F Stat, Stock-Yogo weak identification critical value for 10% maximal instrumental variable size in parentheses. Critical value for 15% maximal instrumental variable size equals 8.96. Standard errors are clustered at the district level, and are presented in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.

Table 4: Falsification test

	Placebo Districts	
	(1)	(2)
Fraction childhood N exposure	0.106	-0.010
	(0.430)	(0.413)
Observations	23338	23338
R-sq	0.0773	0.0683
RMSE	5.796	5.821
Birth-year Fixed Effects	Y	Y
Birth-Month Fixed Effects	Y	Y
District Fixed Effects	Y	Y
State Trends	N	Y
Weather controls	Y	Y

Notes: Columns 1 and 2 show results from a placebo test, in which the upstream district for each observation is replaced by a different, neighboring district that is not upstream. Each column displays estimates from a separate regression. Fraction of early childhood exposed to N pollution is the fraction of years from year of birth to age 3 that nitrate pollution exceeds safety guidelines. Standard errors are clustered at the district level, and are presented in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.

Table 5: Main results with trends for districts

	Dependent variable: Height (cm)	
	(1)	(2)
Fraction childhood N exposure	-2.273**	-2.392***
	(0.887)	(0.891)
Observations	19450	19138
R-sq	0.0846	0.0835
RMSE	6.047	6.064
Birth-year Fixed Effects	Y	Y
District Fixed Effects	Y	Y
District Trends	Y	Y
Weather controls	N	Y

Notes: Table shows results from estimating Eq. (1) via ordinary least squares (OLS). Each column displays estimates from a separate regression. Fraction of early childhood exposed to N pollution is the fraction of years from year of birth to age 3 that nitrate pollution exceeds safety guidelines. Standard errors are clustered at the district level, and are presented in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.

Table 6: Alternative clustering

	Dependent variable: Height (cm)	
	(1)	(2)
Fraction childhood N exposure	-2.246	-1.963
s.e. clustered by district	(0.497)***	(0.506)***
s.e. clustered by state	(0.552)**	(0.488)**
s.e. clustered by survey cluster	(0.928)***	(0.944)***
Observations	19138	19138
R-sq	0.0793	0.0908
Birth-year Fixed Effects	Y	
Birth-Month Fixed Effects	Y	
District Fixed Effects	Y	Y
State Trends	Y	Y
Weather controls	Y	Y
Birth-Year by Month Fixed Effects		Y

Notes: Table shows results from estimating Eq. (1) via ordinary least squares (OLS). Each column displays estimates from a separate regression. Fraction of early childhood exposed to N pollution is the fraction of years from year of birth to age 3 that nitrate pollution exceeds safety guidelines. Standard errors are clustered at the district level, state level and survey cluster level and are presented in parentheses. ***, **, * denote statistical significance at the 1%, 5% and 10% levels respectively.

Table 7: Impacts in Vietnam

Height-for-age scores	(1)	(2)
Fractional Exposure to Nitrate-Nitrite	-0.776**	-0.779**
	(0.338)	(0.337)
Birth-year Fixed Effects	Y	Y
Birth-Month Fixed Effects	Y	Y
Commune Fixed Effects	Y	Y
Province Trends	Y	Y
Other controls	N	Y
N	691	691
R-sq	0.132	0.156

Notes: Statistical significance is given by * p<0.10 ** p <0.05 ***p < 0.01. Standard errors in parentheses are clustered at the commune level. Other controls include precipitation, temperature, ethnicity (tribe), sex, conductivity, phosphorus, water-treatment at home, household asset value, years of education of head, farm/non-farm household

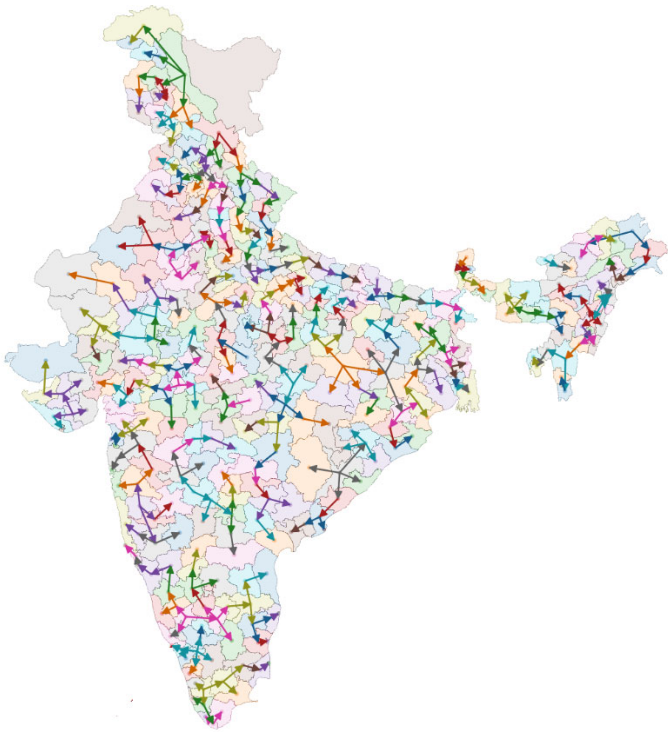
Table 8: Impacts in Africa

In-utero exposure	Stunting	HAZ	Stunting	HAZ
Downstream of N pollution	0.0172***	-0.0729***		
	(0.00636)	(0.0228)		
Downstream of N pollution x Rural			0.0209***	-0.0848***
			(0.00597)	(0.0222)
Fixed effects	Year-Month of Birth, Grid Cell			
Other controls	Y	Y	Y	Y
N	204,886	204,886	204,886	204,886
R-Sq	0.106	0.143	0.106	0.144

Notes: Statistical significance is given by * p<0.10 ** p <0.05 ***p < 0.01. Standard errors in parentheses are clustered at gridcell level. Other controls include household variables – if it is in a rural location, indicator for improved sanitary facilities, improved water source and no sanitation facility (open defecation), child age in months, age of mother at birth giving, if child is a girl, a household wealth index, body mass index (BMI) of mother, an index of mother empowerment (health decisions), mother's years of education and mother's partner's years of education – and community variables – percentage of improved water source, improved sanitation and open defecation, and total population of urban area; temperature and precipitation; and year specific country trend.

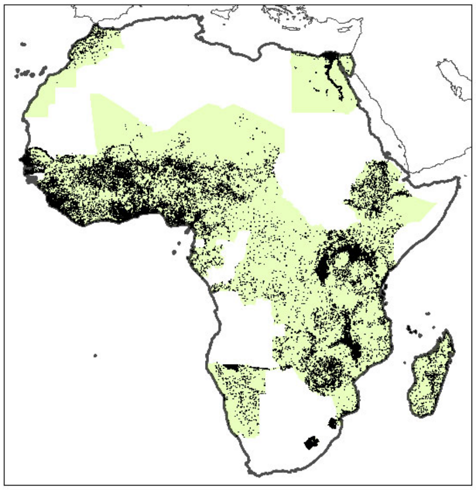
Figures

Figure 1: Upstream-Downstream hydrologic breakdown



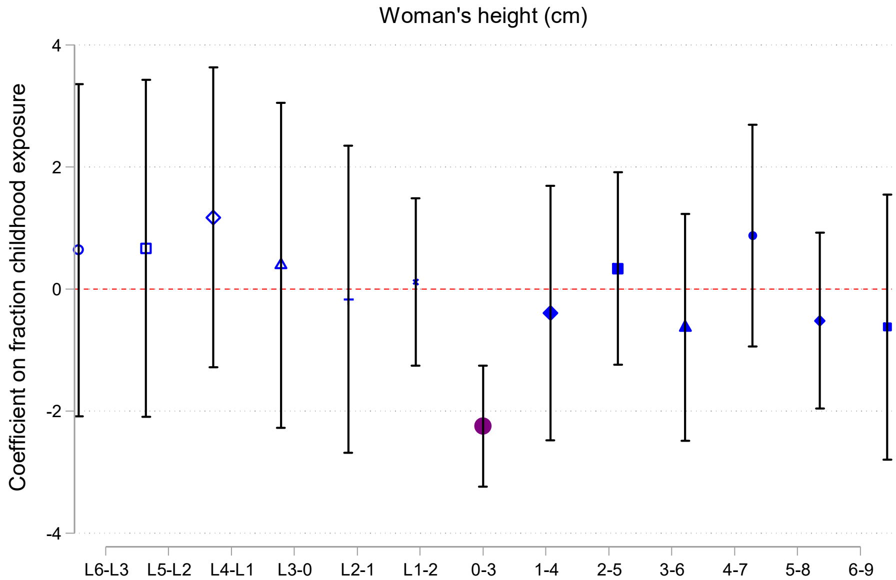
Note: The map shows direction of streamflow from upstream to downstream districts

Figure 2: Countries Studied, Africa



Note: The map shows locations of enumeration areas that were surveyed as part of the DHS Program.

Figure 3: Different window periods of exposure



Note: Estimated coefficients from variants of the main regression equation, in which the period of pollution exposure is shifted by 6 four-year periods (horizontal axis) from the main 0-3 period. Each marker's vertical position therefore measures the estimated impact of exposure at the appropriate period of exposure. For example, the purple marker represents the impact of exposure discussed in the report. Other markers represent the impact of "placebo" exposures. Error bars represent 95% confidence intervals.

Appendix 1: Spatial Stream Network Model

Water quality downstream directly depends on upstream water quality as well as weather related variables and anthropogenic activities happening in between. Models developed by ver Hoef, Peterson and Theobald (2006) and ver Hoef and Peterson (2010) allow to statistically represent these stream dependencies within a network and predict water quality in a spatially valid framework. They allow to flexibly control for spatial auto-correlation between observations belonging to the same river network based on stream distances, to take into account accumulation of pollutants as well as their dilution. They present important improvement to classic geospatial models based on Euclidean distances which were proved to be biased in such settings.

Here, we used the model developed by ver Hoef and Peterson (2010) to fill missing observations in the CWC nitrogen data between 1986 and 2017 where over 60% of the nitrogen observations are missing in this data set, limiting our understanding of the evolution of water quality over the period. More specifically, we used the openSTARS package in R (Kattwinkel and Szcos 2018) to derive a topographically correct stream network for all India. First, a Digital Elevation Model from the SRTM mission was used to derive all streams across India. For computational limits, the original DEM 30 meters model was resampled at a 100m resolution. Second, the upstream area of each CWC station was determined. Third, the stream distance between each station belonging to a given network was calculated. Fourth, the annual level of rainfall, average temperature, average elevation and average slope were computed to better account for dilution of pollution.

Then, we used the SSN package (ver Hoef et al. 2014) to model the determinants of water quality in CWC stations. The original model developed by the authors is:

$$n_i = X_i\beta + S_u + S_d + S_e + W\gamma + \epsilon i$$

Where $\mathrm{n_i}$ is the nitrogen level in station i, $\mathrm{X_i} = \left(\mathrm{X_{i1}} \dots \mathrm{X_{iq}}\right)$ are environmental covariates defined over the upstream area of each station. $\mathrm{S_u}$, $\mathrm{S_d}$, $\mathrm{S_e}$ are a set of spatially auto-correlated random variables that models spatial dependence inside a network. The main dependency we want to capture is the upstream to downstream relation between stations ($\mathrm{S_u}$). The authors also provide the possibility to incorporate downstream to upstream dependencies ($\mathrm{S_d}$), as well as standard Euclidean relationships ($\mathrm{S_e}$). Following common practices in spatial statistics, we assumed exponential spatial dependencies between observations. Finally, $W\gamma$ represents a possible set of fixed effects, such as watershed fixed effects. The model was estimated for each year between 1986 and 2008 – the year of birth of the last woman in the DHS data used in the analysis. Years before 1986 were excluded for an insufficient number of observations (<100).

Our objective year is to find the model that predict best nitrogen levels. To do so, we create loops to estimate for each year 93 models that represent all the possible combinations of covariates and spatial dependencies. Models were validated through a Leave One Out Cross Validation (LOOCV) strategy. The final model was chosen based on a Mean Square Prediction Error (MPSE) criteria. The maximization of the predictive power of the model was achieved by introducing one trick in the original approach: we included as a predictor the average value of nitrate in a station between 1986-2008. For each year, the model was trained on available observations and prediction of nitrate was done for missing observations. The final data set was then used to study the long term impact of nitrogen level on health outcomes.

Additional references

Kattwinkel M, Szocs E. 2018, openSTARS: open source implementation of the STARS ArcGIS toolbox. Seehttps://github.com/MiKatt/openSTARS

Ver Hoef, Jay M., Erin Peterson, and David Theobald. "Spatial statistical models that use flow and stream distance." *Environmental and Ecological statistics* 13.4 (2006): 449-464.

Ver Hoef, Jay, et al. "SSN: An R package for spatial statistical modeling on stream networks." *Journal of Statistical Software* 56.3 (2014): 1-45.