Water Pollution and Infant Mortality: Evidence from India*

Gowthami Venkateswaran Andrew Hultgren
Hemant Pullabhotla Kathy Baylis

October 2026

Abstract

Measuring physical determinants of water pollution at unprecedented spatial resolution for a developing country, we study infant mortality effects of fertilizer runoff in India. We find a remarkable elasticity of infant mortality to upstream river nitrate pollution of 2.5 in rural villages, where access to clean drinking water is limited. We find no effects of downstream pollution. Our estimates suggest that a 2010 change in fertilizer subsidies explains ∼28% of national baseline infant mortality. These findings highlight a substantial, unrecognized driver of infant mortality in contexts with high fertilizer subsidies and weak regulatory institutions—common conditions throughout the developing world.

JEL: Q18, H23, Q53, Q58, I18
Keywords: Fertilizer subsidies, Externalities, Water Quality, Health

*Venkateswaran: Keystone (email: gowthamiui79@gmail.com), Hultgren: University of Illinois Urbana-Champaign (email: ahultgr@illinois.edu), Baylis: University of California Santa Barbara, Pullabhotla: Deakin University. We thank conference participants at the NBER 2026 Spring EEE conference, 2026 World Congress of Environmental and Resource Economists, 2024 Heartland Environmental and Resource Economics Conference, 2024 World Bank DIME Conference; seminar participants at University of British Columbia, University of California Santa Barbara, University of Chicago, Iowa State University, University of Massachusetts Amherst, Monash University and University of Sydney; and Daniel Brent, Laura Grant, Michael Greenstone, Corey Lang, Hope Michelson and Lucija Muehlenbachs for helpful comments and suggestions. All remaining errors are our own.

1 Introduction

Infant mortality rates in low- and middle-income countries exceed those of high-income countries by an order of magnitude, and corresponding levels of infant deaths differ by two orders of magnitude (World Bank, 2024), indicating a substantial infant mortality gap between developed and developing nations. Pollution exposure is the largest risk factor in premature deaths worldwide (Fuller et al., 2022), yet constraints on data availability and quality have limited causal characterization of pollution health effects in developing world contexts. Thus, this important driver of the developing world infant mortality gap remains poorly understood. While evidence of developing country air pollution effects has recently started to grow, corresponding evidence for water pollution remains particularly limited. Understanding these health effects is critical to the design of optimal regulatory policy in developing world contexts, where marginal effects of pollution may differ due to information constraints, defensive investments, or baseline pollution levels (Hsiang et al., 2019). Here, we estimate the causal effect of nitrates, a globally ubiquitous water pollutant, on infant mortality in India, the world's most populous country.

India presents a compelling developing country setting in which to study water pollution and health for several reasons. First, over half of India's 1.2 billion people live in rural areas where reliance on river water for bathing and drinking is common and water quality regulations are inconsistently enforced. Second, nitrate pollution of drinking water, which is common throughout the developing world, is widespread in India due to a long history of heavy nitrogen fertilizer subsidies and corresponding overuse and runoff. Third, India is one of the few developing countries that has detailed data on water pollution, fertilizer use and infant mortality, which we combine with a novel measure of the topographical drivers of runoff, allowing us to observe determinants of pollution treatment and outcomes at unprecedented resolution for a developing country.

As our context, we study a large relative price decrease for nitrogen fertilizer under India's 2010 Nutrient-Based Subsidy (NBS) reform, which rolled back price supports on all fertilizers except urea (nitrogen) due to its political sensitivity (Ravinutala, 2016; Kishore et al., 2021).[1] We employ plausibly exogenous variation in predetermined soil

[1]Between 2005 and 2010, the government's fertilizer subsidy burden increased by 500%, stressing budgets. Under the 2010 NBS program, the government reduced price supports for all fertilizers except urea — the dominant source of nitrogen. Prices of other fertilizers increased substantially while urea prices remained under sustained price controls (Section 2). This policy change has been associated with a significant increase in nitrogen use among Indian farmers (Gulati & Banerjee, 2015).

characteristics, the geography of river catchment basins, and the direction of river flow to instrument for post-reform changes in river nitrogen pollution. In a novel application, we use a recently-developed hydrological model (Heberger, 2022) to identify highly-resolved upstream runoff catchment basins for each water quality monitor in our data, enabling a highly localized (versus geographically aggregated) unit of analysis.

We face two main challenges to identifying the causal relationship between fertilizer use and infant mortality in India. First, changes in nitrogen use and runoff after the NBS-policy may be endogenous to unobserved determinants of health outcomes. We draw on findings from the scientific literature to exploit variation in predetermined soil and topographical characteristics – in particular, the soil clay content of runoff catchment basins – as exogenous shifters of post-policy nitrogen use and pollution exposure. As detailed in Section 2, soil clay content shifts both nitrogen fertilizer application and runoff. Fertilizer application shifts because soils with higher clay content increase plant nutrient uptake and thus marginal returns to nitrogen fertilizer (Burke et al., 2019). Surface water runoff shifts because soils with higher clay content exhibit greater retention of applied nutrients near the soil surface, where they run off more easily to local water bodies.[2] Variation in local topography determines the direction of runoff from fields to rivers and thus which soil types are contained within any given runoff catchment basin.

A second key identification challenge is that local fertilizer use can affect local infant health through gains to productivity and incomes (Bharadwaj et al., 2020), which may lead to improved food security and increases in family health investments. This effect makes estimation of the impact of nitrogen pollution difficult as exposed communities may also experience offsetting local fertilizer benefits. We exploit the flow direction of rivers to address this issue. Rather than focusing on local soil characteristics near infant populations, we trace the watershed upstream and use the soil characteristics of these upstream areas to isolate the effects of nitrate water contamination on downstream populations. We link each water station to its nearby villages, and then map upstream watersheds for each water monitoring station, enabling precise identification of watersheds contributing to nitrate runoff and allowing for detailed mapping of local pollution exposure to local health outcomes. To our knowledge, this represents the first developing-country analysis of water pollution mortality effects at this level of

[2]Note that greater retention of fertilizer near plant roots and higher rates of fertilizer runoff in high-clay soils are not mutually inconsistent: high clay soils exhibit less infiltration of fertilizer into deep soils and groundwater (away from crop root zones). Thus more of the fertilizer is "available" both to the plant root zone as well as for runoff.

geographic precision.

Our analysis begins by measuring the effect of the NBS policy on nitrogen fertilizer use. We combine district-level fertilizer consumption data between 2000 and 2015 with a rich soil dataset from the Harmonized World Soil Database, a high resolution soil inventory providing detailed information on dozens of soil properties in a $\sim$1 km grid. Using both event study and difference-in-differences specifications over plausibly exogenous variation in soil clay content, we confirm that districts with high-clay soils – where returns to nitrogen are higher – exhibit significantly increased post-policy nitrogen fertilizer consumption relative to lower-clay districts within the same state-year. The effect we estimate is economically meaningful, representing an 18% increase over the mean nitrogen fertilizer consumption.

Next, we use data from 1,100 river water quality monitoring stations to study associated changes in nitrate water pollution levels. We pair these data with novel, detailed hydrological data from the Global Watershed API (Heberger, 2022) to map the upstream watershed of each river water quality monitoring station. This hydrological dataset provides all the river tributaries and origins upstream of a given water monitoring station, as well as their corresponding watersheds over land areas. These watersheds usefully define the land area within which the flow of water runoff will eventually reach a given water station, enabling us to precisely measure for the first time in a developing country context the soil clay content "upstream" (within the watershed) of each individual water quality station. Using upstream watershed soil clay content as a source of exogenous variation in post-policy fertilizer application and runoff, we find that stations with high-clay upstream soils exhibit a large and statistically significant 5.6 mg/L increase in river nitrate pollution levels relative to lower upstream-clay stations within the same state-year. This increase is substantial: it represents over half of the 10 mg/L safe limit of nitrates in drinking water.[3] Again, results hold for both difference-in-differences and event study estimators.

Finally, to study the effect of river nitrate pollution on health outcomes, we create a pseudo-panel of infant mortality using household data from approximately 10,000 clusters from the 2015/16 Indian wave of the Demographic and Health Surveys (DHS). We compare infant mortality in cohorts born before and after the policy change, using upstream soil clay content as an instrument for post-period changes in local river nitrate pollution. We find that DHS clusters located close to water monitoring stations with high fractions of clay soil in their upstream watersheds face increased risks of infant

[3] U.S. EPA threshold defined in 56 FR 3526.

mortality after the NBS. This effect is driven by rural DHS clusters, where populations are more likely to rely on river water for drinking and bathing. Our IV regression results are conditioned on DHS cluster fixed effects that control for cross-sectional time-invariant local characteristics and state-by-year fixed effects that absorb any differential trends across states over time. We focus on infant deaths within one year of birth to ensure a precise exposure period, avoiding the complications of long-term differential effects. We find a local average treatment effect for rural DHS clusters of 0.08 additional deaths for each 1 mg/L increase in nitrates in nearby rivers, which implies a large rural infant mortality elasticity of 2.5 to river nitrate pollution. This effect is substantial: it explains about 28% of country-wide baseline infant mortality rates in India and has not to our knowledge been previously documented in any developing country context.

We conduct a series of falsification and robustness checks to probe the validity of our results. First, we estimate the first-stage specification for other water quality indicators available in the data that should be unaffected by the NBS. These include measures such as pH, temperature, dissolved oxygen, and fecal coliform, and we find no effects. Next, while we use the entire watershed retrieved for each station in our main specification, we also run a robustness test that contains only upstream areas within 100 km of water stations to account for potential pollution decay over space and find qualitatively similar effects on nitrate pollution, with a slightly smaller point estimate. Crucially, we also conduct placebo tests on the full IV estimates using the downstream watershed catchment area for all monitoring stations, and find no effects on nitrate pollution or infant mortality. Finally, when linking DHS clusters to nearby water stations, we include stations within 20 km of a given cluster to account for the degree to which the cluster coordinates have been locally randomized for data privacy purposes. We test robustness to other distances and find that results remain consistent over larger distances between DHS clusters and water stations, but precision declines.

This paper makes three main contributions to the literature. First, it contributes to the literature on the human health effects of pollution. Much of this literature focuses on air pollution (e.g., Chay & Greenstone, 2003; Currie & Neidell, 2005; Currie et al., 2009; Graff Zivin & Neidell, 2013; Knittel et al., 2016; Schlenker & Walker, 2016; Deryugina et al., 2019; Anderson, 2020), largely in developed nations like the United States. A smaller literature studies water pollution, again largely focused in the United States (Currie et al., 2013; Keiser & Shapiro, 2018; Keiser et al., 2019; Chen, 2026). However, most of the world's population lives in developing economies, where marginal effects of pollution may differ due to information constraints, behavioral

norms, or higher baseline pollution levels (Jalan & Somanathan, 2008; Arceo et al., 2016; Hsiang et al., 2019).[4] While regulatory policies have made some progress in lowering air pollutants in both India and China, regulations of water pollution remain limited and weakly enforced. For example, Greenstone and Hanna (2014) find null effects on water pollution from India's National River Conservation Plan, noting that implementation lacked clear jurisdiction and dedicated funds, and that enforcement was weak. Our findings highlight the magnitude of the welfare loss associated with weak regulation in this important developing-world context.

Second, the paper contributes to the broader understanding of the unintended consequences of a partial deregulation of subsidies that further distorted relative input prices (Lipsey & Lancaster, 1956). While a growing body of literature explores the effect of subsidy-driven distortions on input use, productivity, and welfare (Restuccia et al., 2008; Hsieh & Klenow, 2009; Adamopoulos & Restuccia, 2014; Kurdi et al., 2020; Donovan, 2021; S. Garg & Saxena, 2023; Chakraborty et al., 2024), less is known about the effects of distortionary input subsidies on health outcomes. Further, though a large body of literature has focused on the impacts of fertilizer subsidies when they are first introduced (see Holden (2019) for a review), less attention has been paid to the consequences of selective fertilizer subsidy removals, which can create new sets of distortions (Gulati & Banerjee, 2015). Our results document an important unintended consequence of one such selective subsidy removal. Overall, in the Indian context, the environmental and human capital damages of fertilizer subsidies have yet to be taken into policy consideration, even though populations there are particularly vulnerable to water source pollution (Fuller et al., 2022).

Third, the paper contributes to improvements in measuring downstream pollution and linking it to social outcomes in the developing world context. Most studies outside the U.S. conduct an upstream-downstream framework at some aggregated level (e.g. Greenstone & Hanna, 2014; Dias et al., 2023), for example describing one political region as upstream of another based on the flow direction of major rivers. Unlike the U.S. context (Keiser & Shapiro, 2018), mapping precise upstream watersheds is still a challenge in developing country settings, with some recent progress (T. Garg et al., 2018; Hagerty & Tiwari, 2022). Using an open-source global watershed mapping

[4]Outside the U.S., studies of water pollution and human health have examined sanitation and human waste (Gamper-Rabindran et al., 2010; T. Garg et al., 2018), industrial waste (Ebenstein, 2012; Do et al., 2018), and pesticides (Lai, 2017; M. E. Skidmore et al., 2023; Dias et al., 2023). None of these examines runoff of fertilizers, which are heavily subsidized in many parts of the developing world (Huang et al., 2017). Brainerd and Menon (2014) studies agro-chemicals broadly, and we discuss it below.

platform developed by Heberger (2022), we enhance the spatial analysis of watersheds in India by mapping the entire upstream watershed of each water monitoring station. This mapping employs detailed gridded terrain data in combination with computational efficiencies to produce highly localized definitions of the catchment basin of each water monitoring station in our data. This detailed mapping allows us to conduct our analysis at the water station level instead of at an aggregate political unit, allowing for the most precise and localized measure of determinants of water pollution treatment to-date in a developing country context.

The closest work that complements this study is by Brainerd and Menon (2014). They study the effects of an index of river chemical pollutants on infant health in India and find significant negative impacts, exploiting seasonal prenatal exposure to a range of chemicals to identify impacts on various measures of child health. Our study builds on Brainerd and Menon (2014) in several important ways. First, Brainerd and Menon (2014) are constrained by the fact that their health data are only geocoded at the state level, whereas we observe data for ∼10,000 DHS villages. Further, their measure of agricultural pollution is an index constructed from statewide averages of high concentration indicators across multiple pollutants, complicating the interpretation of marginal effects.[5] In contrast, we are able to link river monitor-specific measures of nitrate and nitrite pollution directly with the DHS villages near them, measuring a well-defined treatment at each village with an interpretable marginal effect. Finally, due to their data constraints, Brainerd and Menon (2014) estimate health effects in the state where chemicals are applied. Thus their estimates reflect both potential health gains from enhanced productivities and household incomes as well as potential losses from chemical exposure (Bharadwaj et al., 2020). In our analysis, the upstream-downstream specification allows us to isolate the health effects of nitrate pollution by focusing on communities downstream of those where fertilizers have been applied.

Additionally, we note that despite the widespread criticisms of the NBS program in India (e.g., Patel, 2016; Praveen, 2021; Mohapatra, 2022), we are not aware of any causal estimates of its impacts on environmental and health outcomes. We provide important and timely evidence on a policy that is still in place and frequently discussed

[5]Brainerd and Menon (2014) measure treatment as an average of seven dummy indicators for statewide levels of different water pollutants, where indicator i takes a value of 1 if the average level of pollutant i across monitors within the state-month exceeds the U.S. EPA safe threshold for that pollutant. The dummy values of the seven pollutant averages are then themselves averaged, which defines the measure of treatment in a state-month. These indicators include measures for fluoride and chromium, which can indicate fertilizer runoff but are also associated with mining, brick kilns, tanneries, and pesticides (Sankhla & Kumar, 2018; Karunanidhi et al., 2021).

by the Indian government, policymakers, and media (Gulati & Banerjee, 2015; Shagun, 2023; Ghosal, 2023; Damodaran, 2024).

Finally, while our estimates for India alone represent a meaningful fraction of global infant mortality (due to India's large population and relatively high baseline infant mortality rate), our findings also have broad relevance to developing country contexts beyond only India. Excessive fertilizer use and the associated pollution of water bodies is a global issue (Mateo-Sagasta et al., 2017). The introduction of fertilizer subsidies during the Green Revolution was instrumental in boosting agricultural yields, nutrition, and economic output in many regions of the world (Bharadwaj et al., 2020; von Der Goltz et al., 2020). However, prolonged subsidies have also encouraged excessive use of fertilizers (Huang et al., 2017; Cassou et al., 2017; Kurdi et al., 2020), and recent evidence suggests that the benefits of increased fertilizer use have begun to plateau as higher application rates translate into reduced productivity gains (Wuepper et al., 2020; Mueller et al., 2017; Itin-Shwartz, 2024). Globally, only about 40% of nitrogen applied as fertilizer is absorbed by crops; the rest runs off to surface water or infiltrates to groundwater (Mateo-Sagasta et al., 2017), making nitrates a ubiquitous water pollutant in many developing world countries even though measurement may be limited. The large infant mortality effects we find here may thus have important implications for pollution and health that extend far beyond the borders of India.

The remainder of the paper proceeds as follows. Section 2 provides background details on the NBS policy, nitrate pollution, and the key geographical characteristics that compose our instrument. Section 3 describes the data, and Section 4 details our identification strategy. We report results in Section 5, and Section 6 provides a brief interpretation and concludes.

2 Background

Our empirical design leverages variation in soil types, river catchment basins, and river flow direction to estimate impacts of nitrogen runoff under the NBS policy on infant mortality. This section provides particular contextual and scientific background information focused on: 1) the NBS policy in India; 2) nitrogen fertilizer uptake by plants and runoff, as well as health effects of exposure; and 3) the role of soil clay content in driving both agricultural returns to nitrogen fertilizers as well as runoff to waterways. River catchment basins and flow direction are conceptually straightforward, and we describe those data in Section 3.

2.1 Nutrient-Based Subsidy

The Indian government began subsidizing fertilizers in 1977, following the Green Revolution. However, between 2005 and 2010, the fertilizer subsidy burden increased by 500%, driven by a substantial increase in international prices (Ravinutala, 2016). At the height of the global food price spike in 2008-09, fertilizer subsidies accounted for 65% of the total government agricultural subsidy expenditure (Gulati & Banerjee, 2015), prompting the government to introduce changes to the subsidy structure to alleviate the fiscal burden.[6]

In April 2010, the Indian government significantly changed how fertilizer subsidies were handled. It transitioned from a product-based subsidy, under which the government set fixed prices for specific fertilizer products, to the nutrient-based subsidy (NBS) system still in place today, where the government provides specific (per unit) subsidies to manufacturers but does not generally regulate final prices.[7] One key exception under this new policy was urea, the most widely used fertilizer in India and the dominant source of nitrogen for Indian farmers (Pawar et al., 2025), which continued to be subsidized at the same product-specific subsidy rate. Urea prices were left untouched because policymakers considered the decentralization of the entire urea sector to be highly politically sensitive (Kishore et al., 2021).

[6]Earlier changes to the fertilizer subsidy policy tended to be short-lived. For instance, a similar increase in the subsidy fiscal burden in 1991 led to the deregulation of phosphorus and potassium fertilizers, but these subsidies were re-introduced within a year. Notably, nitrogen remained subsidized throughout. The time period used in our analysis does not include any other fertilizer subsidy policy changes beyond the 2010 policy change we study.

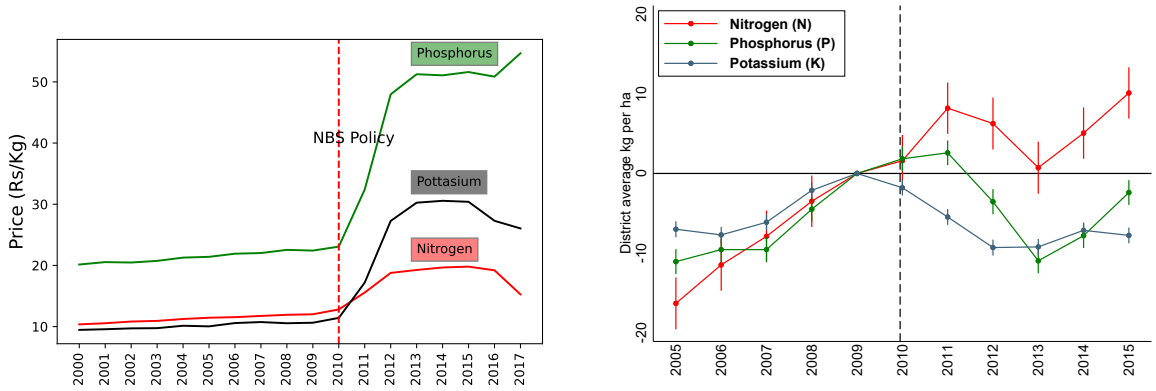
[7]The NBS subsidizes fertilizers based on their content of key nutrients nitrogen (N), phosphorus (P), potassium (K) and sulphur (S). The major nutrients used in India are N, P, and K. Diammonium Phosphate (DAP) is the most widely used phosphate fertilizer and Muriate of Potash (MOP) is the most widely used potassium based fertilizer. NBS values are available for DAP, MOP, and 14 other grades of complex fertilizers.

In practice, the NBS policy is implemented as follows. Each year, the government announces a fixed subsidy rate in rupees per kilogram for each nutrient, based on international prices, inventory levels, and expected demand. These per-kg rates are then converted into per-ton subsidies for each product based on that product's nutrient composition. For example, before 2010, the government set a fixed price for DAP; but since the NBS reform, DAP has been subsidized in India based on its nutrient composition of 18-46-0-0, representing the ratio of N:P:K:S. The 2023-2024 rabi (winter) season subsidy rate for N was 47.02 Rs/kg and for P was 20.82 Rs/kg (since DAP does not contain any K or S, we ignore them here). The subsidy rate per metric ton of DAP is then:

Nitrogen subsidy: 18% of 1,000 kg = 180 kg,	180×47.02 = Rs.	8,463
Phosphorus subsidy: 46% of 1,000 kg = 460 kg,	460×20.82 = Rs.	9,577
Total DAP subsidy per metric ton:	Rs.	18,040

Following the announcement of the nutrient subsidies each season, manufacturers set their own maximum retail prices (MRP). The government transfers the subsidies directly to fertilizer companies based on actual sales made by retailers, and farmers pay the final subsidized price for each product.

Figure 1. Trends in fertilizer prices and change in levels of usage after NBS policy



(a) Farmer retail price trends

(b) Changes in fertilizer application rates

Notes: Panel (a) shows average annual farmer-reported retail prices (INR/kg) for N, P, and K fertilizers using data from the Cost of Cultivation Surveys (Section 3). Panel (b) plots the coefficients of N, P and K application rates (kg/ha) on year indicator variables from a district-level panel regression with district fixed effects, illustrating the change in fertilizer application relative to 2009.

The introduction of the NBS policy led to substantially different price changes across inputs, with persistently low prices (and high subsidies) for N, via urea, and increasing prices (and reduced subsidies) for P and K. Panel (a) of Figure 1 plots the prices of each fertilizer component from the Cost of Cultivation Surveys, demonstrating that P and K prices more than doubled following introduction of the NBS policy, while N prices exhibited a much smaller increase.[8] Thus, under the NBS, Indian farmers faced a relative decrease in the price of N as compared to P and K inputs. Panel (b) of Figure 1 shows changes in district-level fertilizer application rates compared to 2009. After the policy change, N use remained elevated while use of both P and K fell.

The descriptive patterns we see in Figure 1 echo what previous researchers and government policy documents have noted. For instance, Gulati and Banerjee (2015) argue that excluding urea from the NBS scheme was a significant oversight in India's fertilizer policy reform. They cite the Planning Commission's twelfth plan document, which acknowledges that "NBS roll-out was seriously flawed since urea was kept out of its ambit. Urea prices remain controlled with only a 10 percent rise at the time of adoption of the NBS in 2010. Meanwhile, prices of decontrolled products doubled"

[8]Even though urea prices remained fixed, average N prices increased due to the presence of N in other fertilizer products that were affected by the NBS, such as DAP.

(Planning Commission, 2013, p. 14).

2.2 Nitrate runoff, water pollution and health

The application of nitrogen-based fertilizers like urea introduces inorganic nitrogen into the soil where it undergoes a series of chemical transformations.[9] Initially, nitrogen is decomposed to ammonia, which is absorbed by plants. However, soil bacteria also oxidize ammonia into nitrites and eventually to nitrates. While nitrates are absorbed by plants during their growth, they are also highly water-soluble and leach into groundwater or run off into local water bodies, becoming a source of pollution. Additionally, the degree to which a plant absorbs nitrogen is a function of other nutrient availability (von Liebig, 1843). For example, if a plant does not have sufficient phosphorus, it will not be able to absorb nitrogen as efficiently, leading to reduced nitrogen uptake and potentially greater runoff. Under-application of phosphorus and other N-complementary inputs has been a long-standing problem in the Indian context (Kanwar et al., 1982).

Nitrate contamination of water sources poses severe risks to human health and the environment (C.-W. Liu et al., 2014; Li et al., 2021). Infants are particularly vulnerable as high nitrate levels in drinking water can lead to methemoglobinemia ("blue baby syndrome"), a potentially fatal condition that reduces the blood's capacity to carry oxygen. The regulatory standard for nitrates in drinking water was originally set to protect against infant methemoglobinemia rather than broader reproductive or developmental harms, and epidemiological evidence suggests that adverse outcomes, such as preterm births, low birth weight and fetal growth restrictions may arise even at concentrations below current regulatory thresholds (Knobeloch et al., 2000; Ward et al., 2018). Methemoglobinemia has been a widespread concern in India for decades and through the present-day (Majumdar, 2003; Verma et al., 2023); for example, over one quarter of children in Northern India are estimated to face health risks from nitrate exposure in drinking water (Verma et al., 2023).

The challenge of nitrogen pollution is further complicated by the high cost of removing nitrates from water at home. Methods such as carbon filtration or boiling do not reduce nitrate levels in water. The process of reverse osmosis can remove nitrates; however, it is an expensive technology, making it infeasible for most Indian households. Lack of low-cost treatment technologies may therefore amplify the health impacts of increases in environmental nitrate exposure.

[9] See, e.g., NDSU 2022 for a summary of this basic soil science.

2.3 Clay soil characteristics

Clay-rich soils have distinct properties that enhance both nutrient retention and water-holding capacity, making clay soils particularly responsive to fertilizer applications (Churchman, 2018; Cambouris et al., 2016).[10] Because of their finer particle sizes, clay soils typically retain water and nutrients more effectively than coarse sandy soils, where large pore spaces allow water to drain rapidly and reduce the time over which nutrients remain available in the root zone. Clay soils also carry a more negatively charged surface, which increases the soil's cation exchange capacity (CEC). A higher CEC increases the ability of soil particulate surfaces to temporarily hold positively charged nutrient ions such as ammonium, a key form of nitrogen nutrient, and release them gradually to plants, rather than allowing them to be washed away. In contrast, nitrate, formed by ammonium decomposition, is a negatively charged ion and thus is repelled by the negatively charged surfaces of clay soil particles. This causes soil nitrates to even more readily enter runoff to surface waters.

These properties help explain why fertilizer can be more productive on clay-rich soils. Because clay soils generally hold moisture for longer periods, crops are less likely to face short-run water stress after fertilizer application, and key nutrients like ammonium remain in the biologically active root zone for longer, both of which allow for greater nutrient uptake by crops. In contrast, sandy soils drain quickly, so both water and soluble nutrients are more likely to move below the root zone before plants can fully use them. As a result, clay-rich soil can increase the marginal return to fertilizer application,[11] as suggested by plot-level regression evidence from the Indian Cost of Cultivation Surveys.[12]

Notably, clay-rich soils also increase the potential for fertilizer runoff to water bodies. The key hydrologic mechanism is that clay soils have smaller pores than sandy soils and thus lower infiltration rates. When rainfall intensity exceeds soil infiltration capacity, excess water flows away as runoff. Further, as previously mentioned, the negatively charged surfaces of clay particles repels negatively nitrate ions, enhancing nitrate runoff. These characteristics make clay-rich areas more prone to transporting

[10] We seek to provide an accessible summary of the soil science literature; see Churchman (2018), Wang and Li (2019), and Musei et al. (2024) for detailed discussions.

[11] Soils with a clay content greater than 70 percent may not be ideal as excessive clay can hinder water drainage and aeration, potentially reducing crop productivity. Our data does not contain areas with such extremely high levels of clay content.

[12] Translog-style production-function estimates with farmer-crop and state-crop-year fixed effects indicate nitrogen has a positive average association with both yield and revenue, and the estimated return to nitrogen is higher on plots with greater clay content (Appendix Tables B.1 and B.2).

fertilizer residues and dissolved nutrients into nearby streams, rivers, and other surface water bodies (Bombino et al., 2019; Wang & Li, 2019).

Taken together, these two linked features of clay-rich soils — greater agronomic responsiveness to fertilizer and greater propensity for surface runoff — motivate our use of upstream clay content as a shifter of downstream nitrate pollution exposure.

3 Data

3.1 Fertilizer Usage

We measure nitrogen use before and after the NBS policy using district-level fertilizer consumption data available from the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) District-Level Database (DLD). The ICRISAT-DLD provides annual district-level totals of nitrogen (N), phosphorus (P), and potassium (K) consumption and per-hectare use for 1990–2017.

In addition to district aggregates, we conduct supplementary analyses of farmer-level input choices using the Cost of Cultivation Surveys (CCS), conducted by the Ministry of Agriculture and Farmers Welfare, Government of India. The CCS collect detailed plot-level information on input use, crop-specific input prices and output prices for a rotating sample of farmers followed across all planting seasons for three consecutive years within a wave. We use annual CCS waves from 2008–2009 through 2018–2019, covering major crops including rice, wheat, cotton, sugarcane, and maize.

3.2 River Water Quality

We measure surface water quality using station-level data from the National Water Quality Monitoring Programme (NWMP).[13] The NWMP reports multiple indicators, including nitrate and nitrite, fecal coliform, pH, dissolved oxygen (DO), and temperature, for more than 1,100 river monitoring stations. Our data are constrained in that nitrate-nitrite data became available for a larger number of monitoring stations only after 2006, three years before the implementation of the NBS program in 2010. Thus, we focus our analysis on 2007–2014, the period with sufficient spatial coverage of nitrate/nitrite water pollution measures.[14]

[13] NWMP is collected and managed by India's Central Pollution Control Board. The list of stations and their geographic coordinates is sourced from https://cpcb.nic.in/wqm/WQMN_list.pdf.

[14] Our analysis period ends in 2014, which is the last complete year of infant mortality reported in the data.

Our first outcome of interest is nitrogen pollution, measured as the maximum concentration of nitrate-nitrite (mg/L) at the station-year level. Higher nitrate-nitrite levels correspond to more nitrogen contamination. In addition to nitrate-nitrite levels, we examine other indicators of water pollution such as dissolved oxygen, fecal coliform, temperature and pH. Panel A of Figure 2 maps the NWMP monitoring network for stations reporting nitrate-nitrite levels over our study period.

3.3 Upstream-Downstream Watersheds

A key component of our identification strategy is that pollution observed at a monitoring station reflects upstream activities. We therefore construct hydrologically consistent upstream exposure measures for each water quality monitoring station using the Global Watershed API developed by Heberger (2022). For any point on a river network, this hydrological model returns flow direction, river tributary structure, upstream origins, and the associated watershed polygons. Using this API, we construct exposure to upstream watersheds for each water monitoring station in our data (Figure 2, Panel C).

The use of the Global Watershed API represents a meaningful data enhancement relative to earlier research on fertilizer pollution. Prior work in this context often aggregates exposure within administrative boundaries such as districts or counties (T. Garg et al., 2018; M. E. Skidmore et al., 2023) or uses fixed-radius buffers around stations (Cai et al., 2016). Such strategies can result in measurement error through the mis-attribution of downstream or cross-basin exposure to a station, masking variation in the area that actually drains to a station. In contrast, our watershed-based approach addresses such limitations by (a) restricting exposure to upstream areas only, aligning the exposure set to the hydrologically appropriate drainage polygons, and (b) allowing watershed size and shape to vary by station according to regional topography. These choices reduce measurement error and sharpen the link between upstream regions containing agricultural activity and observed nutrients in downstream surface water.

Our watershed approach requires precise locations of the water stations. However, some of the water stations do not have reliable latitude and longitude information. We quality-check CPCB station coordinates, snap stations to the nearest river flowline when minor offsets occur, and discard stations with unreliable geolocations. To avoid spurious catchments (e.g., canal segments or tiny headwater polygons), we exclude stations whose delineated upstream watershed area is smaller than 100 km^2.

3.4 Soil Characteristics

We use data from the Harmonized World Soil Database (HWSD) v2.0, from the Food and Agriculture Organization (FAO), to measure soil characteristics. The HWSD v2.0 is a high resolution global soil inventory, providing detailed information on various soil properties in a 30 arc-second (approximately 1km) raster format worldwide. We focus on the topsoil layer (0–20 cm), which is most relevant for nitrogen application and runoff. For each grid cell, we compute shares of clay, sand, and silt.

We extract soil information at two scales. First, using 2011 district boundaries, we compute district-level averages of soil fractions for our district-based analysis. Second, for the station-level water quality analysis, we compute the same fractions within each delineated upstream watershed polygon for each NWMP station.

3.5 Health Outcomes

We measure infant mortality using the 2015–2016 National Family Health Survey (NFHS-4), India's Demographic and Health Survey (DHS) round. NFHS-4 collected data from January 2015 to December 2016, surveying approximately 600,000 households nationwide using four instruments: household, women's, and men's questionnaires and biomarker data for a sub-sample of individuals.

Here we focus on the women's questionnaire from the NFHS-4 which records detailed reproductive histories, including the year and month of birth of every child, for all women in the sample between the ages of 15 and 49. The reproductive histories also include data on the child's gender, twin status, birth order and survival status. The data indicate whether each child born is currently alive and, if not, the age of death. Our primary health outcome is infant mortality (death within 12 months of birth), given the established vulnerability of infants to nitrate exposure through methemoglobinemia. Following Do et al. (2018), we use women's birth histories to construct a cluster-by-year pseudo-panel of infant deaths. Due to limited earlier coverage of nitrates in the NWMP, we exclude pre-2007 births from the analysis. Appendix Figure D.2 shows the annual distribution of births and deaths for the NFHS-4 sample. Both births and deaths as reported by mothers in 2015/2016 are stable well before our sample pre-period begins in 2007, suggesting that recall was not on average an issue during our time-frame.

The NFHS-4 also provides latitude and longitude coordinates for around 28,000 sample clusters (Figure 2, Panel B). The locations of these clusters are displaced slightly

for confidentiality reasons. Urban clusters are displaced up to 2 kilometers and rural clusters are displaced by 5–10 kilometers. To address this displacement, and because a river may pass closer to a village than its associated water monitoring station, we compute the average maximum nitrate–nitrite concentration (by year) and upstream clay content for the three closest stations within a 20 km buffer around each NFHS-4 cluster. We test different buffer sizes in Appendix D.

3.6 Descriptive patterns in the data

Table 1 reports descriptive statistics for key observables in the data. The column "All" reports unconditional means and standard deviations, subsetted to the pre-treatment period (before 2010) where possible. We then report conditional means for low and high clay regions ("Means by soil clay") and their differences. We additionally control for state-year fixed effects and report the residualized conditional means and their differences ("Within state-yr means by soil clay"). Panel A displays fertilizer application rates, which are reported by district. Panel B displays observables from the DHS data, reported by DHS cluster. Panels C and D display water quality and upstream soil characteristics linked to DHS clusters as described previously in this section.

Mean differences between low and high clay regions are common, but largely disappear when examining the within state-year variation employed in estimation. Differences in the cross section are unsurprising, as states in India's decentralized governance structure have substantial autonomy to set their own policy decisions, and clay content differs across states. Some within state-year differences in educational attainment, religion, and access to pit latrines (but not other human waste handling options) remain across low and high clay regions (Panel B and Appendix A), and we address such level differences in our empirical design. Because soil composition affects runoff to rivers, within state-year differences in water quality between low and high clay regions are to be expected (Panel C and Appendix A). Additional descriptive statistics are reported in Appendix A, and patterns follow those described here.

3.7 Defining "high upstream clay" water stations and DHS clusters

Our identification strategy relies on pre-determined variation in soil clay content, the geography of river watersheds, and river flow direction. Figure 2 depicts how we combine these data to define "high upstream clay" water stations and DHS clusters.

Panel A depicts the distribution of soil clay content and of water stations reporting nitrate pollution during our study period, and panel B depicts the coverage of DHS

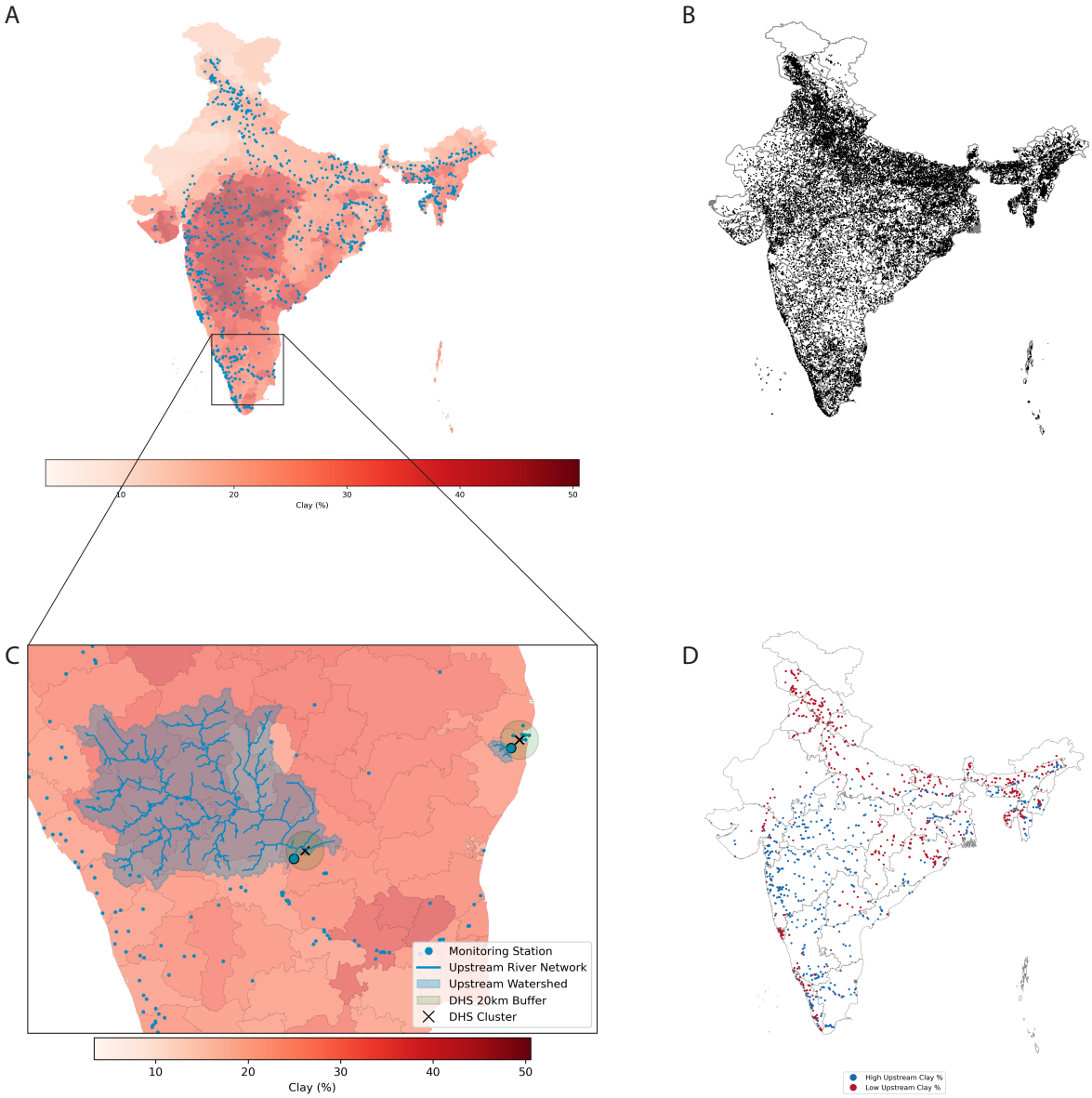
Table 1. Descriptive statistics

		Means by soil clay			Within state-yr means by soil clay		
	All	Low Clay	High Clay	Diff	Low Clay	High Clay	Diff
		Panel A: Fertilizer application, by district					
N (kg/ha)	80.95	92.98	68.88	24.10***	0.40	-0.40	0.81
	(68.35)	(75.71)	(57.65)		(57.24)	(50.60)	
P (kg/ha)	32.88	31.21	34.55	-3.34**	-0.70	0.70	-1.39
	(28.16)	(25.94)	(30.14)		(21.60)	(26.80)	
K (kg/ha)	17.83	12.10	23.59	-11.49***	-0.37	0.37	-0.75
	(23.51)	(17.13)	(27.35)		(14.30)	(21.30)	
		Panel B: Observables from DHS data, by DHS cluster					
Infant deaths	0.08	0.09	0.06	0.03***	0.00	-0.00	0.00
	(0.29)	(0.33)	(0.26)		(0.32)	(0.26)	
Births	1.66	1.80	1.56	0.24***	-0.00	0.00	-0.00
	(1.42)	(1.46)	(1.38)		(1.42)	(1.33)	
Male births (%)	0.53	0.53	0.52	0.01***	0.00	-0.00	0.00
	(0.08)	(0.08)	(0.08)		(0.08)	(0.08)	
No educ. (%)	0.34	0.37	0.33	0.04***	0.01	-0.00	0.01**
	(0.26)	(0.26)	(0.26)		(0.22)	(0.22)	
No toilet (%)	0.29	0.29	0.29	-0.00	0.00	-0.00	0.01
	(0.33)	(0.33)	(0.33)		(0.25)	(0.28)	
Has electricity (%)	0.88	0.86	0.90	-0.05***	-0.00	0.00	-0.01
	(0.19)	(0.22)	(0.17)		(0.18)	(0.15)	
Rural (%)	0.61	0.68	0.56	0.11***	-0.00	0.00	-0.01
	(0.49)	(0.47)	(0.50)		(0.44)	(0.47)	
		Panel C: Water quality observables, by DHS cluster					
Max. B.O.D.	7.30	7.63	7.05	0.58	0.51	-0.38	0.89**
(mg/L)	(14.47)	(18.63)	(10.28)		(17.72)	(6.58)	
Max. ammonia	2.36	0.90	3.66	-2.76***	-0.06	0.05	-0.11*
(mg/L)	(3.67)	(0.59)	(4.65)		(0.47)	(1.41)	
Max. fec. coli.	205	343	91	251***	8.0	-6.5	14.5
(mg/L /100)	(707)	(999)	(251)		(937)	(185)	
Max. NO_3/NO_2	3.88	3.33	4.29	-0.96***	0.21	-0.16	0.37***
(mg/L)	(6.27)	(2.86)	(7.90)		(2.07)	(4.86)	
		Panel D: Upstream catchment basin soil characteristics, by DHS cluster					
Clay (%)	0.28	0.20	0.33	-0.12***	-0.01	0.01	-0.02***
	(0.078)	(0.031)	(0.060)		(0.024)	(0.036)	
Sand (%)	0.39	0.40	0.39	0.01***	0.001	-0.001	0.002*
	(0.051)	(0.038)	(0.058)		(0.024)	(0.038)	

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports unconditional descriptive statistics ("All"), subsetted to the pre-treatment period before 2010 for time-varying data (Panels A and C, and infant deaths and births in Panel B). Additionally, we report means ("Means by soil clay") for "Low Clay" and "High Clay" regions and their differences ("Diff"), and corresponding residualized means ("Within state-yr means by soil clay") after controlling for state-by-year fixed effects. Mean differences between low and high clay regions are common, but largely disappear when examining the within state-year variation employed in estimation. Because soil composition affects runoff to rivers, within state-year water quality differences between low and high clay regions are to be expected. Additional descriptive statistics reported in Appendix A.

Figure 2. Illustration of upstream clay definition.



Notes: Panel A shows the distribution of soil clay content (background) and water monitoring stations (blue circles) throughout India, and panel B shows the coverage of DHS clusters. Panel C illustrates the calculation of clay content for upstream watersheds. Two example DHS clusters (black x's) are shown, each with one example water monitoring station within 20km. One station has a large upstream watershed (light blue shape, center) while the other has a small upstream watershed (upper right). Average upstream clay content for each water station is calculated over each watershed area. Panel D shows the distribution of high- versus low-clay water monitoring stations in blue and red respectively, with "high" clay stations as those with above median clay content.

clusters. Water station coverage is prevalent in both high- and low-clay regions (except for the arid regions of Rajasthan and Gujarat in the west), and DHS clusters are dense throughout India.

Panel C demonstrates how the geographic variation in watersheds combines with variation in soil clay content to define the upstream clay content of a location. Two example DHS clusters are depicted (black x's), each with an example water quality station (blue circles with black outlines) within a 20km buffer of the respective cluster (light green, translucent circles). Upstream river segments and associated watersheds are depicted for each station. The geography around the station in the center of the plot is such that the associated watershed is large, encompassing a wide range of soil clay contents. Conversely, the geography around the station near the east coast is such that the associated watershed is small and composed of more geographically proximate soil clay content. For each water quality station, the "upstream clay" content is calculated as the average clay content over all areas encompassed by its respective upstream watershed.[15] For each DHS cluster, we then calculate the average upstream clay content over the three (or fewer) closest water stations within its 20km buffer. This yields a distribution of upstream clay contents over all DHS clusters with nearby water quality stations in the data. Finally, we define a water quality station or a DHS cluster as "high upstream clay" if its upstream clay content is above the median of the respective station or cluster distribution.[16] Soil science finds that the influence of soil clay content does not emerge in very high sand soils (Section 2), so we calculate an analogous distribution of upstream soil sand content and exclude a station or cluster from the "high upstream clay" group if its upstream sand content is above the 75^{th} percentile of the respective soil sand distribution. Panel D depicts the distribution of "high" (blue) and "low" (red) clay water quality stations, with state boundaries outlined in gray. Note that we employ state-year fixed effects in estimation, thus absorbing across-state variation in the distribution of high and low clay water stations or DHS clusters.[17] Examination of panel D reveals that within many states, high and low clay water quality stations exist in close proximity. This is due to variation in local topography that drives variation in associated watersheds, along with variation in the soil clay content of those watersheds.

[15] The figure depicts soil clay averaged by district for visual clarity, but the actual calculation of watershed clay content is done on the raw raster data.

[16] Appendix D displays robustness to an alternative specification that bins clay content.

[17] Using state-year fixed effects implies that states in Northern India with exclusively high-clay DHS clusters (Figure 2 panel D) are dropped from estimation.

4 Estimation Framework

We face two key challenges in our empirical setting. First, fertilizer application and associated run-off are endogenous. We need plausibly exogenous variation in pollution exposure to estimate causal effects on infant health. Second, increased use of nitrogen fertilizer can also directly affect local child health through increased agricultural productivity, confounding estimation of pollution effects.

To address the first challenge, the endogeneity of nitrogen use, we rely on exogenous variation from predetermined soil characteristics and local topography. Specifically, we leverage two aspects of clay soils, based on findings from soil science, that affect nitrogen runoff. First, areas with a higher percentage of clay in their soil composition tend to exhibit higher returns to fertilizer use. Second, soils with a higher percentage of clay tend to exhibit increased surface runoff of nitrates. See Section 2.3 for a detailed discussion of both of these aspects. Additionally, variation in local topography defines the geographic scope of the watersheds over which we evaluate soil clay characteristics.

To address the second challenge, potentially confounding health gains from local fertilizer use, we employ variation in river flow direction to study effects of increased nitrogen use in upstream watersheds on downstream infant mortality. Because local clay content and thus local fertilizer use may be correlated with geographically proximate upstream clay, we include state-by-year controls for local soil clay content in all of our upstream/downstream regressions. Gains to local health through, for example, regional trade present a threat to identification, and we employ a downstream clay placebo test to rule out such effects. Overall, differences in the size and shape of upstream watersheds, along with differences in their soil composition, give us plausibly exogenous variation in upstream fertilizer use and associated runoff, and thus downstream nitrate pollution exposure.

4.1 Difference-in-differences design

Our ultimate objective is to estimate the effect of the change in nitrate pollution under the NBS policy on infant mortality, for which we will employ an instrumental variables estimator. However, to do so we must first demonstrate that 1) nitrogen fertilizer application increased in high-clay areas post policy (to validate our mechanism), and 2) that river water quality stations with high upstream clay content experienced a post-policy increase in nitrate runoff (demonstrating a first stage). We use a difference-in-differences approach to estimate the impact of the NBS policy change on these

outcomes.

By design, the NBS policy deregulated pricing of P and K fertilizers while holding subsidies for urea constant, resulting in a relative decrease in the price of N inputs (Section 2.1). However, the timing of this policy change may be endogenous. We thus interact a post-policy indicator with plausibly exogenous variation in soil clay content, allowing us to estimate causal effects of the NBS policy in high-clay regions relative to low-clay regions.

We employ event-study and DiD estimators of the following form:

$$\mathrm{y}_{ist} = \sum_{k \neq 2009} \beta_k \; \mathit{High\ Clay}_i \cdot \mathbf{1}\{k = t\} + \phi_i + \gamma_{st} + \zeta Z_{ist} + \epsilon_{it} \tag{1}$$

$$\mathrm{y}_{ist} = \beta \; \mathit{High\ Clay}_i \times \mathit{Post}_t + \phi_i' + \gamma_{st}' + \zeta' Z_{ist} + \epsilon_{it}'. \tag{2}$$

Here, y_{ist} can either refer to fertilizer usage (in kg/ha) in district i or nitrate pollution (in mg/L) at water quality monitoring station i, both occurring in state s and year t. Unit of observation fixed effects ϕ_i absorb time-invariant characteristics. State-by-year fixed effects γ_{st} account for time-varying factors that differ across states, which is important in the Indian context where states have a high degree of independence. In the fertilizer usage estimates over districts, "*High Clay*" refers to the clay content of the respective district (upstream/downstream issues do not matter in this case because we study fertilizer application directly, not runoff). In the water quality estimates over monitoring stations, "*High Clay*" refers to the clay content of the upstream watershed. Our comparison of high- versus low-clay units produces a lower bound of the effects following the policy change because units that are classified as low-clay may still be affected by the policy.

We also use Eq. 1 and its DiD variant to generate reduced-form estimates of infant mortality impacts, where y_{ist} refers to total infant mortality in DHS cluster i, state s, and year t, and "*High Clay*" refers to the average clay content of the upstream watersheds of nearby water quality monitoring stations (see Section 3.7). In our water quality regression, Z_{ist} includes state-by-year controls for local soil clay content. In infant mortality estimates, Z_{ist} includes state-by-year controls for: local soil clay content, average fraction of male children, and average proxies for household wealth (indicators for household electrification, ownership of a refrigerator, and ownership of a car or truck).[18] These controls serve to reduce noise in estimation, and Appendix D

[18] We define "local" soil using a 10km radius for water quality outcomes and a 20km radius for DHS

displays robustness to excluding them.

Because our regressions employ state-year fixed effects, our identifying assumption in this design is that both high- and low-clay units within state-year would have seen their outcomes develop along parallel trends in the absence of the NBS policy, which we obtain if within-state variation in the predetermined clay content of predetermined geographies is as good as random with respect to post-policy farmer actions. In our event-study designs, pre-period estimates provide suggestive evidence supporting the parallel trends assumption. Further, for water quality effects we test a wide range of placebo outcomes. Finally, for both water quality and infant mortality effects, we additionally test a placebo "downstream clay" treatment.

4.2 Instrumental variables design

Estimation of Equation 2 for infant mortality yields the reduced-form impact of the NBS policy on infant mortality in high-clay relative to non-high-clay locations, but the model does not provide causal estimates for the effect of nitrates themselves. To address this, we employ an IV design, exploiting variation in local topography defining catchment basins, soil characteristics within topographically-defined catchment basins, and river flow direction to instrument for nitrate pollution. In particular, we use our upstream high clay indicator interacted with a post-NBS indicator as an instrument for downstream nitrate concentrations, where upstream clay serves as a plausibly exogenous shifter of post-policy increases in upstream nitrogen fertilizer application and runoff to downstream areas. Section 2 motivates the instrument and details its construction.[19]

We estimate the following two-stage least squares regressions, where Equations 3 and 4 are the second and first stage regressions, respectively:

$$Infant\ Mortality_{ist} = \beta\, \widehat{Nitrates}_{ist} + \mu_i + \eta_{st} + \zeta Z_{ist} + \varepsilon_{ist}, \tag{3}$$

$$Nitrates_{ist} = \pi\,(High\ Clay_i \times Post_t) + \mu_i' + \eta_{st}' + \zeta' Z_{ist} + u_{ist}. \tag{4}$$

Here, $Nitrates_{ist}$ denotes average nitrate concentrations measured across stations

survey outcomes. While the positions of water quality monitors are reported precisely, positions of DHS clusters are jittered by up to 10km for data privacy purposes.

[19] In brief, higher levels of soil clay in the upstream areas of nearby river water quality stations increase nitrate runoff through both greater nitrogen fertilizer application in upstream areas and increased surface runoff. We adopt the upstream-downstream specification (Figure 2 panel C) to avoid local effects of soil quality (via fertilizer application) on health, operating through increased local access to nutrition and increased household income and health investments.

within 20km of a DHS cluster i in state s and year t, and $\widehat{Nitrates}_{ist}$ is the fitted nitrate exposure for cluster i. Cluster fixed effects μ_i absorb time-invariant DHS cluster characteristics, while state-year fixed effects η_{st} absorb state-specific shocks over time. As in the reduced form estimator, Z_{ist} includes state-by-year controls for local soil clay content, average fraction of male children, and average proxies for household wealth (indicators for household electrification, ownership of a refrigerator, and ownership of a car or truck). These controls serve to reduce noise in estimation (Appendix D displays robustness to excluding these controls). Standard errors are robust to correlations at the DHS cluster level.

The key exclusion restriction underlying our IV strategy is that *High Clay* $\times$ *Post* affects infant health only through the channel of downstream nitrate pollution, after controlling for DHS cluster fixed effects, state-year fixed effects, and state-year controls for local clay and other observables. Regional trade or other linkages between regional soil characteristics and local outcomes present a threat to our exclusion restriction assumption. To probe this assumption, we include placebo tests for a range of other water quality characteristics and multiple variants estimating a downstream clay placebo treatment, with results discussed in Section 5.

5 Results

We first demonstrate the effect of the NBS policy on fertilizer use in high versus low clay districts. Second, we estimate the effect of the policy change on nitrate pollution measured at water quality stations downstream of high versus low clay watersheds. Third, we estimate the effect of both the policy and (instrumented) nitrates on infant mortality at the DHS cluster level. We discuss the results for each of these steps in turn.

5.1 Effects on fertilizer use

We first ask whether the 2010 NBS policy affected nitrogen and other fertilizer use. Using event-study regressions on fertilizer use that control for district and state-year fixed effects (Eq. 1), we find a significant and sustained increase in nitrogen consumption per hectare of farmland in districts characterized by high clay content following the 2010 policy change (Figure 3). Notably, we do not find significant changes in the use of phosphorus and potassium fertilizers in high versus low clay districts after the NBS, suggesting that not only did the rate of nitrogen application increase, but so

did the ratio of nitrogen to potassium and phosphorus. Recall from Section 2.2 that nitrate runoff can be exacerbated by the relative under-application of potassium and phosphorus, as a lack of these nutrients can limit crop nitrogen uptake. Looking ahead to the next step of the analysis, nitrate runoff may increase not only because of direct increases in nitrogen fertilizer use, but also because crops may uptake a reduced fraction of that additional fertilizer. Pre-period placebo effects are flat in all estimates, consistent with plausibly exogenous variation in treatment (of the 15 pre-period point estimates reported, only one is significant).

Table 2 presents findings from the DiD estimator (Eq. 2, which also controls for district and state-year fixed effects). These findings show a significant impact of the policy on annual nitrogen application per hectare in districts with high clay content relative to low-clay districts, consistent with farmers differentially shifting into nitrogen fertilizers in high clay regions where returns to nitrogen are higher. Like the event-study results, we find no such effects on phosphorus or potassium applications. The estimated effect of the NBS for nitrogen is 15 kg/ha, which represents a substantial 18% increase in nitrogen use relative to mean application rates of 84 kg/ha. Such an increase is consistent with the relative price decrease for nitrogen under the NBS policy, where subsidies for urea (the dominant source of nitrogen) were held fixed but prices of other fertilizer products increased under liberalization of price controls.

Table 2. Impact of the NBS policy on fertilizer use

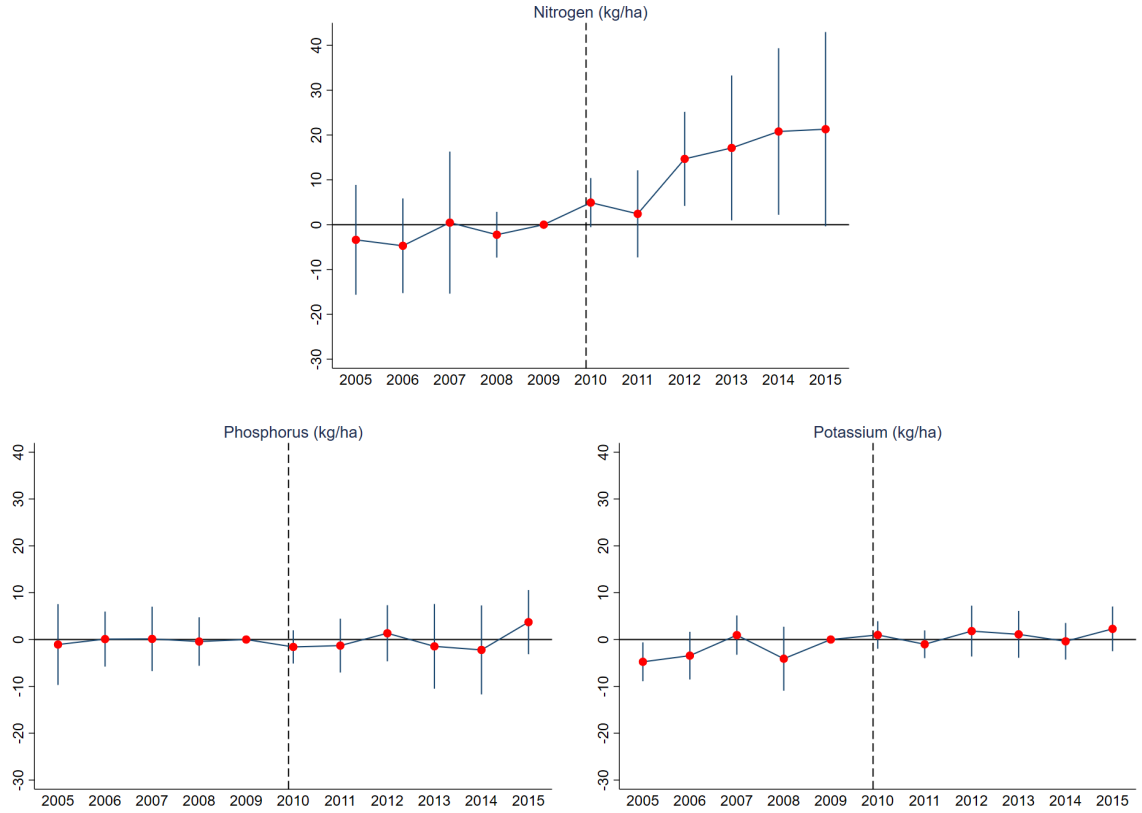
	Nitrogen (kg/ha)	Phosphorus (kg/ha)	Potassium (kg/ha)
	(1)	(2)	(3)
High Clay x Post	15.37**	-0.0241	3.031
	(5.949)	(3.259)	(2.204)
Observations	4,162	4,162	4,162
Dep var mean	84.48	33.99	16.90
District FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports estimates of the Nutrient-Based Subsidy reform on fertilizer use in districts with high versus low levels of soil clay content. "High Clay" refers to an indicator for districts with soil clay levels above the median (see Section 3.7). Standard errors (in parentheses) are robust to correlations at the district level.

By using soil clay as a shifter in fertilizer application, our econometric approach identifies post-policy effects in high clay districts relative to low clay districts, but does not identify how fertilizer applications changed in absolute levels. Appendix Figure D.1 descriptively plots the time series of nitrogen fertilizer application rates in both

Figure 3. Event study plots of N, P and K use in high vs. low clay districts



Notes: The figure shows the regression coefficients of fertilizer use for N, P and K on interaction terms between high clay and year dummies. "High Clay" refers to an indicator for districts with soil clay levels above the median (see Section 3.7). The vertical dashed lines denote the timing of the NBS policy. 95% confidence intervals are displayed as solid lines. Standard errors are robust to correlated shocks at the district level. The model includes district fixed effects and state-year fixed effects.

sets of districts from the beginning of the data in 1990, and we observe that nitrogen use continued to increase in both sets of locations after the NBS policy. We also observe remarkably parallel pre-trends. We note that this descriptive analysis results from regressing nitrogen usage on year indicators separately for high clay and low clay districts, and thus cannot control for important state-year fixed effects. However, it provides suggestive evidence that the estimated relative increase in nitrogen fertilizers in high clay regions (Table 2) may be an absolute increase as well.

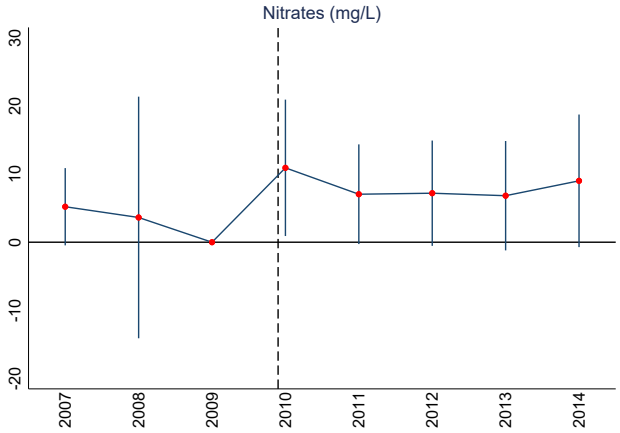
5.2 Effects on nitrate pollution

Having provided evidence that nitrogen applications increased in high clay relative to low clay areas after the NBS policy, we next estimate the effects of the NBS on ni-

trate runoff to downstream water quality monitoring stations. Specifically, we compare nitrate levels measured downstream of high clay versus low clay watersheds (Section 3.7).

Figure 4 displays results from our event study design (Eq. 1). Following the policy change, river water nitrate contamination increased substantially in areas with high clay upstream watersheds relative to those with low clay watersheds. This increase is consistent with estimates from the previous section of higher levels of nitrogen fertilizer application in high clay soils (Table 2), and is also consistent with findings from the soil science literature that nitrate pollution runoff will be higher in areas with high soil clay content (Section 2). Event study results show the estimated increase in nitrate pollution persisted over time — it is significant in the first year ($p = 0.036$) and is consistently positive and marginally significant in all remaining post-policy years (p-values range 0.059 – 0.095). Further, though they exhibit some noise, pre-period estimates appear stable, suggesting that there were no significant differences in nitrate pollution trends between stations with high levels and low levels of soil clay in their upstream watersheds before the NBS program was introduced.

Figure 4. Monitoring station nitrates in high clay vs. low clay upstream watersheds



Notes: The figure shows the regression coefficients of nitrate-nitrite levels in mg/L on interaction terms between upstream clay indicators and year dummies. The vertical dashed line denotes the timing of the NBS policy. 95% confidence intervals are displayed as solid lines. Standard errors are robust to correlations at the water station level. The model includes water station fixed effects, state-year fixed effects, and state-year controls for local soil clay.

Table 3 reports findings from the DiD estimator (Eq. 2). We provide estimates for nitrate pollution (column 1) and also for a range of other water quality measures reported by monitoring stations that serve as placebo outcomes (columns 2–5). In

Table 3. Impact of the NBS policy on nitrate pollution and placebo outcomes

	Nitrates	Fecal Coliform	Temperature	D.O.	pH
	(1)	(2)	(3)	(4)	(5)
High Clay x Post	5.602**	-4,659.6	0.243	0.295	-0.130
	(2.241)	(3,998.3)	(0.414)	(0.244)	(0.113)
Observations	2,734	3,638	4,104	4,121	4,151
Dep var mean	3.092	25,527	28.39	8.139	18.07
Station FE	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: "High Clay" refers to an indicator for water quality monitoring stations with soil clay levels above the median in their upstream catchment basins (see Section 3.7). Standard errors (in parentheses) are clustered at the water station level. Column 1 reports coefficients for nitrate pollution, which is recorded as maximum nitrate-nitrite levels (mg/L). Columns 2 – 5 report placebo water quality indicators such as maximum fecal coliform (values divided by 100), maximum temperature, maximum dissolved oxygen, and maximum pH. All estimates include state-year controls for local soil clay.

column 1, we find that post policy, stations downstream from high clay watershed regions experienced a very large increase in nitrate-nitrite levels of 5.5 mg/L on average, nearly tripling the mean level of 3.1 mg/L. The magnitude of this increase is particularly concerning given that the safe limit for nitrate-nitrite concentrations in drinking water is 10 mg/L. An increase of 5.5 mg/L in nitrate pollution is itself over half of the total safe concentration and could push many stations over the 10 mg/L safety threshold.

Columns 2–5 of Table 3 report results for other outcomes reported by water quality stations which we would not expect the NBS policy to primarily affect: maximum fecal coliform, maximum temperature, dissolved oxygen, and pH. These outcomes are linked to a range of other human activities such as sewage and waste dumping, industrial discharge to rivers, or heavy air pollution (via acid rain).[20] We estimate null effects for each of these placebo outcomes, suggesting that nitrate runoff from fertilizer applications is indeed driving changes to river water quality and not other human activities that happen to also change nitrate levels or are correlated with fertilizer runoff.

Because chemical concentrations can decay in water over time, there is no consensus in the literature on how far upstream is "upstream" for pollution mapping. If the distance employed is large, the mapping might not be accurate due to decay in pollution measures (though decay of nitrate pollution in rivers is slow). If this distance employed is small, the mapping may not capture all important upstream pollution inflows. We

[20]Dissolved oxygen and pH can decrease with nitrates if nitrate levels are high enough to drive large algal blooms in the presence of available organic matter, but such algal blooms tend to occur in stagnant waters rather than flowing rivers.

Table 4. Impact of the NBS policy on infant mortality

	Reduced Form			IV		
	Full Sample	Rural	Urban	Full Sample	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)
Upstream Clay x Post	0.0535*	0.124***	-0.0510			
	(0.0284)	(0.0367)	(0.0460)			
Nitrates				0.0400*	0.0849***	-0.0353
				(0.0215)	(0.0299)	(0.0363)
Observations	31,162	18,260	12,880	31,162	18,260	12,880
Dep var mean	0.0723	0.0907	0.0461	0.0723	0.0907	0.0461
KP F-Stat				43.81	57.54	15.02
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports coefficient estimates from reduced form specifications and instrumental variables. Standard errors (in parentheses) are robust to correlations at the DHS cluster level. The first three columns report estimates for the full, rural and urban samples of DHS clusters under the reduced form specification. Columns 4, 5 and 6 show results from the IV specification for the full, rural and urban samples. All regressions include state-year controls for observables including local soil clay. The sample is limited to DHS clusters that have at least one monitoring station within a 20 km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

show robustness of our water station estimates to capping each upstream watershed at a 100 km radius from a given station (Appendix Table D.11), and results are qualitatively similar.

5.3 Effects on infant mortality

Having established links between the NBS fertilizer policy change, nitrogen fertilizer use in regions with high clay soils, and nitrate-nitrite pollution concentrations at river water quality stations downstream from watersheds with high clay soils, we now estimate the impacts of this policy on infant health. Our primary outcome is the number of infant deaths within the first year of life in a given DHS cluster-year. We use our indicator for high soil clay content in the upstream watersheds of water stations near a given DHS cluster (Section 3.7), interacted with the policy timing, as an intent to treat measure and then as an instrument for nitrate pollution in nearby rivers. Because rural households are more likely than urban households to rely on river water for drinking and bathing, and therefore may be more exposed to river water pollution, we also estimate effects separately for rural and urban DHS clusters.

Table 4 reports the reduced-form and instrumental variable findings based on Eqs. 2 and 3, respectively. All specifications are conditioned on DHS cluster fixed effects,

state-by-year fixed effects, and state-by-year controls for observables including local soil clay, and standard errors are robust to correlations at the DHS cluster level. Reduced-form estimates (columns 1–3) indicate that DHS clusters near river segments with high clay upstream watersheds experienced increases in infant deaths following the NBS policy relative to their low clay counterparts, with marginally significant effects in the full sample (column 1) driven by highly significant effects in rural clusters (column 2). In contrast, the reduced-form estimate for urban clusters cannot be distinguished from zero (column 3). Our finding of stronger rural effects and null urban effects is consistent with greater dependence on untreated or lightly treated surface water for drinking, cooking, and bathing in rural India.

These reduced-form estimates indicate that the NBS policy affected infant deaths, especially in rural areas downstream of high clay soils, but they leave much available information untouched. Our previous results have demonstrated that nitrogen fertilizer rates increased substantially in high clay regions, that nitrate runoff from these high clay regions increased substantially downstream, and that downstream water quality measures associated with other human activities were unchanged. These findings suggest we should have a strong first stage for IV estimation, and that pollution exposure changes are indeed driven by upstream nitrogen fertilizer application and runoff rather than correlated co-pollutants.

We thus instrument for nearby river water nitrate pollution with an indicator for high upstream clay post-policy, producing local average treatment effect (LATE) estimates for the impact of post-period changes in nitrate concentrations (in high clay watersheds relative to low clay watersheds) on nearby infant mortality. We find that a one mg/L increase in nitrates at nearby water stations results in a marginally significant 0.04 additional infant deaths in the full sample (column 4 of Table 4). As with the reduced-form estimates, this full-sample result appears to be driven entirely by rural areas. We estimate that rural infant mortality increases by a remarkable and highly significant 0.08 deaths per mg/L increase in nitrate concentrations (column 5), but we find no effect in urban areas (column 6). Our first stage is strong, with full sample and rural sample Kleibergen–Paap F-statistics of 44 and 58, respectively. Appendix Table D.1 reports the first-stage regressions for nitrate pollution exposure at DHS clusters.[21]

[21]The full-sample point estimate in Table 3 column 1 is larger than the first stage estimate of Appendix Table D.1 column 1, but of similar sign and only marginally statistically different. However, results in Table 3 are estimated at the water-station level, versus estimation at the DHS cluster level in Appendix Table D.1. Thus, estimates in Appendix Table D.1 represent averages in nitrate concentrations over multiple water stations, some of which may be along different rivers in cases where multiple rivers pass near the same cluster. Additionally, "high clay" is defined in Appendix Table D.1

As noted in Section 4.2, regional trade or other linkages between regional soil characteristics and local outcomes present a threat to our exclusion restriction assumption, which requires that upstream clay only affect post period downstream infant mortality via downstream nitrate pollution (after controlling for fixed effects and state-year controls for local soil clay and other observables). We test for such regional linkages by constructing a placebo measure of *downstream* soil clay content. Downstream clay cannot affect pollution runoff to upstream river locations, but should pick up regional correlations in agricultural outcomes that might have spillover effects to an upstream location (thus threatening our exclusion restriction). Directly analogous to our main specification instrument, we construct a placebo instrument for "high downstream clay" in the post period. Reduced form and IV placebo results are displayed in Table 5. All placebo estimates are statistically indistinguishable from zero, and the important rural placebo estimates are economically small as well.

Given the dependent variable mean in rural areas is 0.09 deaths per year, the rural effects we estimate in column 5 of Table 4 represent a striking change in infant mortality and imply an elasticity of infant mortality to river water nitrates of 2.5 in rural areas of India near rivers.[22] That is, a 10 percent increase in river nitrate pollution raises nearby rural infant deaths by roughly 25 percent. Based on this effect in rural areas and the null effects in urban areas, our estimates explain roughly 28 percent of baseline infant mortality in India (Section 6), indicating that fertilizer runoff and associated nitrate pollution of river water represents an important, previously-undocumented driver of rural infant health in the Indian context.

In magnitude, our estimates are larger than those in Brainerd and Menon (2014), who find that a 10 percent increase in agrichemicals in water increases infant mortality by 4.64 percent in India.[23] This difference is not surprising given substantial differences in measurement and design. The measurement of exposure in Brainerd and Menon (2014) is a state-level average of seven dummy indicators for the harmful presence of seven distinct pollutants in water during gestation, with each indicator representing an average over all water stations within a state before all seven indicators are themselves averaged. Our estimates use observed nitrate concentrations in mg/L measured at monitoring stations near each DHS cluster, and therefore are able to capture a more

at the DHS cluster level, and represents the average upstream clay content of nearby water stations. Estimates in Table 3 are robust to winsorization of the data (results available upon request).

[22] Calculated as $\frac{0.0849/0.0907}{1/2.7} = 2.5$, where 2.7 is the mean level of water station nitrates near rural DHS clusters.

[23] Brainerd and Menon (2014) find that a one unit increase in their state-level measure of agrichemical presence in water increases infant mortality by 0.08 deaths.

direct, precise, and spatially localized measure of exposure. Further, identification in Brainerd and Menon (2014) relies on differences in wheat and rice planting seasons or on differential trends in wheat and rice cropped areas across predominantly wheat states (in Northern India) and rice states (in Southern and Eastern India). However, differences in planting seasons are correlated with seasonalities in health outcomes, and may confound seasonal effects of temperature and rainfall on infant health (McMahon & Gray, 2021). Additionally, differential trends in cropped area are correlated with multiple determinants of health outcomes, such as differences in women's labor-force participation (Carranza, 2014). In contrast, we employ plausibly exogenous variation in clay soil distributions, water monitor catchment basin geographies, and river flow direction along with a specific policy shock to measure within-state variation in nitrate pollution and health.

More broadly, other studies of overall water quality improvements — not necessarily focused on fertilizer runoff — provide a useful benchmark for the magnitude of our estimates. Galiani et al. (2005) find that water privatization in Argentina reduced child mortality by about 8 percent overall and by 26 percent in the poorest areas, while Cutler and Miller (2005) estimate that clean water technologies account for roughly 59 percent of the decline in infant mortality in major U.S. cities in the early twentieth century. Viewed against this broader literature, the 28 percent of baseline infant mortality explained by our nitrate-specific estimates is large, but well within the range suggested by existing studies, particularly in rural or historical settings where households face limited access to treatment or avoidance options. Appendix C provides a more detailed comparison with the related literature.

5.4 Robustness Checks

We conduct a range of additional tests and robustness checks across our outcomes, and discuss each briefly here.

5.4.1 Testing nitrogen-clay complementarities

As discussed in Section 2, the soil science literature suggests that farmers will use more nitrogen on higher clay content soils. We test this assumption by using the cost of cultivation survey data to estimate a translog-like production function for six common crops in India that are grown on both high clay and lower clay soils: rice, wheat, maize, soy, sugarcane and cotton. We include farmer fixed effects so that estimates

Table 5. Downstream placebo: Impact on infant mortality (IV)

	Reduced Form			IV		
	Full Sample	Rural	Urban	Full Sample	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)
Downstream Clay x Post	-0.0070	0.0050	-0.0411			
	(0.0374)	(0.0594)	(0.0373)			
Nitrates				-0.0123	0.00380	0.341
				(0.0662)	(0.0449)	(1.032)
Observations	31,162	18,260	12,880	31,162	18,260	12,880
Dep var mean	0.0723	0.0907	0.0461	0.0723	0.0907	0.0461
KP F-Stat				4.07	9.99	0.10
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Table reports placebo estimates using downstream soil properties as an instrument for upstream nitrate pollution, otherwise following our main IV specification. Downstream buffer depicted in Figure D.3. Standard errors (in parentheses) are robust to correlations at the DHS cluster level. The first three columns report estimates for the full sample, rural and urban samples of DHS clusters under the reduced form specification. Columns 4, 5 and 6 show results from the IV specification for the full sample, rural and urban samples. All regressions include state-year controls for observables including local soil clay. The sample is limited to DHS clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

are based on within-farmer variation across plots and years, and we include state-by-year fixed effects to control for state-specific time-varying policies such as minimum support prices. We demean all independent variables and interact reported nitrogen application with soil clay content as well as phosphorous and potassium applications. Results are displayed in Appendix B. Examining associations with crop yields (Table B.1), at the mean level of inputs we find that nitrogen application is associated with a positive marginal product for all crops, and a higher marginal product in higher versus lower clay soils (the significance of these estimates varies across crops). Estimates with farmer revenue as the outcome and input expenditures as the independent variables give similar results with more noise (Appendix Table B.2). These estimates suggest that the benefits of nitrogen are higher in high-clay soils, consistent with our finding that farmers in higher clay regions use more nitrogen fertilizer after the relative decrease in urea (thus, nitrogen) prices under the NBS policy (Table 2).

5.4.2 Water quality placebo tests

As discussed previously, we estimate whether other water quality indicators were affected by being downstream from high-clay watersheds after the NBS. We evaluate

changes in four other water quality indicators unrelated to nitrogen pollution (fecal coliform, temperature, dissolved oxygen, and pH) and find no statistically significant differences coming from high versus low-clay watersheds for any of them. Results are reported in Table 3. This lack of effect on unrelated water quality measures suggests that the observed increase in nitrate-nitrite levels is indeed the result of changes in fertilizer application and does not carry with it substantial changes in other pollutants.

Further, as a second placebo test, we assess whether upstream nitrates or other waterborne pollutants are affected by downstream clay soil levels (testing the opposite river flow direction). If our results are picking up a general within-state watershed effect that changed after the NBS policy instead of one specifically associated with upstream nitrogen use, we might expect that downstream soil quality might be correlated with nitrate levels. Table D.8 reports estimates using clay levels downstream of water monitoring stations. "Downstream Clay" is an indicator variable to denote water stations with high clay levels (above the median) for soils downstream of the focal water monitoring station. Regressions include state-by-year controls for local soil clay, and standard errors are robust to correlations at the water station level. Reassuringly, we find no effect of downstream soil clay on nitrate or other pollutant levels.

5.4.3 Infant mortality - simultaneous estimation of treatment and placebo

As discussed previously, we use downstream clay as a placebo instrument for local nitrate pollution in our infant mortality IV, and find no effects (Table 5). We further test variants of this placebo in joint estimation with the upstream clay treatment. We use Upstream Clay $\times$ Post as our instrument while including Downstream Clay $\times$ Post as an exogenous control (Appendix Table D.6); and we use our Downstream Clay $\times$ Post placebo as an instrument while including Upstream Clay $\times$ Post as an exogenous control (Appendix Table D.7). Our main IV estimates are robust to inclusion of the placebo control, and our placebo estimates remain null when including the upstream clay treatment as a control.

5.4.4 Infant mortality - Flexible clay soil categorization

We explore whether an alternative definition of the soil variable affects our findings. Rather than use above-median upstream clay levels to instrument for high-runoff exposure, we instead employ a flexible binned estimator of upstream clay content, with soil clay under 25% as the omitted category (Appendix Table D.4). Results are robust, with similar and statistically indistinguishable point estimates.

5.4.5 Infant mortality - DHS cluster buffers

To estimate the effect of upstream nitrate pollution on infant mortality, we create 20 km buffers around each infant cluster and calculate the average nitrate levels from water monitoring stations located within these buffers in our main specification. Given that the DHS cluster locations as reported in the data are spatially "jittered" by up to 10 km for confidentiality and that the true cluster locations may be closer to rivers than the associated water monitoring stations would indicate, we consider the 20 km buffer to be a reasonable choice for our analysis. However, we test robustness to this choice by using buffers of 10 and 40 km (half and double our preferred buffer). Results in Appendix Table D.5 show that the estimated effect of nitrate pollution on rural infant mortality is most precise within the 10 and 20 km buffers. Beyond 20 km, as we include surface water stations that are farther from DHS clusters, the effect remains consistent but loses significance, suggesting that proximity to contaminated water sources is a key aspect of measuring exposure.

5.4.6 Alternative outcomes

We finally test robustness to using infant mortality rates (deaths per live birth) as our outcome (Appendix Table D.9). Results are noisier, but the key rural IV result in particular is robust to this specification, with a very similar magnitude as a fraction of the mean. Note that this specification has fewer observations as some DHS clusters report deaths of infants born in the previous year, without any births in the contemporaneous year. Examining births as the outcome, we do find an increase in high-clay rural clusters (Table D.10), but this increase as a fraction of the mean is quite small relative to the mortality and mortality rate magnitudes we estimate, suggesting that any changes in births do not drive our findings. Applying our mortality estimates to any additional live births, such live births would only explain about 13% of our estimated rural mortality effect.[24] We account for this in Section 6.

6 Discussion and Conclusion

Infant mortality rates in low- and middle-income countries exceed those of high-income countries by an order of magnitude. We assemble data at unprecedented spatial res-

[24]Baseline infant mortality rates are 0.047 deaths per live birth (Appendix Table D.9) and our estimated mortality effect in rural areas is 0.0849 (Table 4), so $0.047 \times 0.24/0.0849 = 0.13$

olution on the physical determinants of agricultural water pollution for a developing country, and we find that nitrogen fertilizer runoff from the agricultural sector explains fully one quarter of the developed- versus developing-country infant mortality gap in the Indian context. Causal estimates of developing-country infant mortality effects of nitrogen pollution do not to our knowledge exist, even though nitrogen fertilizer subsidies are widespread throughout the developing world, water quality regulatory regimes are often weak, and populations frequently lack access to clean drinking water.

Our estimation strategy draws on findings from the scientific literature that link soil clay content to increases in both nitrogen fertilizer returns and also nitrogen runoff to nearby waterways. We thus assemble detailed, geo-located data on India's soils as well as the geographies of the catchment basins upstream of each river water quality monitor at unprecedented precision for a developing country, and we pair these data with highly-resolved village-level infant mortality data and individual station-level water quality measures. This allows us to employ for estimation plausibly exogenous variation in soil clay characteristics which shift nitrogen application and runoff, landscape topography which determines the direction of runoff from fields to rivers, and river flow direction — all precisely linked to each river water quality monitor and its nearby villages. To our knowledge, this represents the first developing-country analysis of water pollution mortality effects at this level of geographic precision. Finally, India's 2010 Nutrient-Based Subsidy program induced a large relative price decrease for nitrogen fertilizer, and we further leverage this temporal variation and its differential consequences by soil type for estimation.

We document three key findings. First, we find that the NBS policy induced a significant and substantial 18% increase in nitrogen use in high clay versus low clay districts, highlighting how the policy disproportionately incentivized nitrogen use in these regions. Second, we find that this increased fertilizer use led to large increases in nitrate pollution in downstream rivers. Water quality monitors with high clay catchment basins saw their nitrate concentrations nearly triple after the policy, reaching dangerously high pollution levels. Finally, we find that this increased pollution caused a substantial increase in downstream infant mortality, with estimated effects concentrated among rural communities where reliance on river water for consumption and bathing is common.

We provide a brief interpretation of our key infant mortality finding. Our results are based on differences between high- and low-clay upstream watersheds. To understand the overall effect of the policy change, we need information on whether the policy

increased nitrogen use and nitrate pollution on net, or only in relative terms. To that end, we descriptively note that nitrogen use continued at a similar rate of increase in low clay locations after the NBS policy (Appendix Figure D.1). Thus, if we assume that the policy did not affect nitrogen use and pollution in low clay watersheds, the first stage of our IV regression indicates that nitrates increased by 1.45 mg/L downstream of high-clay soils in rural areas, where our estimated effects occur. This increase resulted in an additional 0.12 annual infant deaths per rural DHS cluster near a river water quality station (1.45 mg/L $\times$ 0.08 deaths per mg/L). With 1.91 mean annual births per rural cluster in our DHS data, the implied death rate is 61 deaths per 1,000 live births, which represents 28% of baseline infant mortality rates or about 380,000 deaths in a baseline year.[25] Applying a value of statistical life for India of 275,000 \$2013 (Viscusi & Masterman, 2017)[26] gives a gross welfare loss of roughly \$105 billion per year. Estimated returns to the agricultural sector from increased nitrogen use are small in comparison.[27]

Two comments on the above interpretation of our findings bear noting. First, in seeking to provide a "back of the envelope" characterization of the welfare implications of our findings, we acknowledge that we make many simplifying assumptions in order to assess rough levels of infant deaths and their welfare consequences. Second, though concerns of methemoglobinemia ("blue baby syndrome") from nitrate pollution of water sources are already well documented in the Indian context, it is highly unlikely that a majority of the infant deaths we document were directly caused by methemoglobinemia. However, nitrate exposure is also associated with preterm births and low birthweight,

[25]During our 2007–2009 pre-period, India's average infant mortality rate, crude birth rate, and population were 50.7 deaths per 1,000 births, 22 births per 1,000 people, and 1.21 billion people, respectively (World Bank, accessed June 23, 2026). Using DHS women's sampling weights, treated rural DHS clusters represent 27.3% of maternal-aged women in India. Thus, 61 $\div$ 50.7 $\times$ 27.3% = 32.8% of baseline infant mortality, which assumes zero effects in urban clusters and in rural clusters far from river water quality stations. Finally, we net out 13% of deaths that may result from additional live births (Section 5.4.6), for 32.8% $\times(1-13\%)$ = 28.4% of baseline infant mortality. Given baseline birth rates and population, this translates to about 380,000 deaths in a baseline year (28.4% $\times$ 50.7 $\div$ 1,000 $\times$ 22 $\div$ 1,000 $\times$ 1.21 billion).

[26]The Viscusi and Masterman (2017) estimate employs a base VSL of \$9.6 million and an income elasticity of unity, which are very similar or identical to other papers employing these estimates (Carleton et al., 2022; Pullabhotla & Souza, 2022). To be conservative, we do not account for a value of life years, which would increase the estimated welfare burden of infant deaths.

[27]We estimate an average crop net revenue nitrogen elasticity of 0.044 (Appendix Table B.2) and an 18% increase in nitrogen application in high clay districts (Table 2). From the Cost of Cultivation data, 64.2% of baseline year agricultural revenue is in high-clay farms, and subsidized nitrogen fertilizer costs represent 4.0% of revenues. We thus estimate a 0.05% increase in net revenues (64.2% $\times$ 18% $\times$ (0.044 $-$ 0.040) = 0.05%), over a baseline average of 275 billion \$2015 in agriculture, forestry, and fishing value added (World Bank, accessed June 23, 2026), or \$0.13 billion.

which are themselves associated with higher rates of infant mortality (Sherris et al., 2021; Coffman et al., 2022; Royal et al., 2024).

Our findings have important implications for water pollution and infant mortality throughout much of the developing world. First, fertilizer subsidies are by no means limited to the Indian context, and are common throughout the developing world. Because large fractions of developing country populations engage in agriculture, these subsidies can be highly politically sensitive, suggesting they are likely to remain in place. Further, more than half of global nitrogen fertilizer runs off to surface water or infiltrates to groundwater (Mateo-Sagasta et al., 2017), indicating that both fertilizer subsidies and associated water pollution are widespread developing world issues. Second, water quality regulations are limited and weakly enforced in many parts of the world beyond India, particularly where populations lack access to clean drinking water. Thus, populations that most rely on untreated surface water for consumption may be those where water quality regulatory enforcement is weakest and agricultural pollution is most likely. We document a scale of human health and welfare implications from nitrogen fertilizer runoff that warrants study in other country contexts. Finally, we note that treatment of nitrates in drinking water supplies is difficult and costly, and is particularly infeasible in most developing country contexts. This highlights the urgent need for policies to prevent runoff to local waterways. Such policies are well documented (Kanter et al., 2020; P. Liu et al., 2023; M. Skidmore et al., 2023) and we provide a benchmark estimate for their economic evaluation should they be implemented and enforced.

References

Adamopoulos, T., & Restuccia, D. (2014). The size distribution of farms and international productivity differences. *American Economic Review*, *104*(6), 1667–1697. https://doi.org/10.1257/aer.104.6.1667

Anderson, M. L. (2020). As the wind blows: The effects of long-term exposure to air pollution on mortality. *Journal of the European Economic Association*, *18*(4), 1886–1927.

Arceo, E., Hanna, R., & Oliva, P. (2016). Does the effect of pollution on infant mortality differ between developing and developed countries? evidence from mexico city. *The Economic Journal*, *126*(591), 257–280.

Bharadwaj, P., Fenske, J., Kala, N., & Mirza, R. A. (2020). The Green revolution and infant mortality in India. *Journal of Health Economics*, *71*, 102314. https://doi.org/10.1016/j.jhealeco.2020.102314

Bombino, G., Denisi, P., Gómez, J. A., & Zema, D. A. (2019). Water infiltration and surface runoff in steep clayey soils of olive groves under different management practices. *Water*, *11*(2), 240. https://doi.org/10.3390/w11020240

Brainerd, E., & Menon, N. (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. https://doi.org/10.1016/j.jdeveco.2013.11.004

Burke, W. J., Frossard, E., Kabwe, S., & Jayne, T. S. (2019). Understanding fertilizer adoption and effectiveness on maize in zambia. *Food Policy*, *86*, 101721. https://doi.org/https://doi.org/10.1016/j.foodpol.2019.05.004

Cai, H., Chen, Y., & Gong, Q. (2016). Polluting thy neighbor: Unintended consequences of china s pollution reduction mandates. *Journal of Environmental Economics and Management*, *76*, 86–104.

Cambouris, A. N., Ziadi, N., Perron, I., Alotaibi, K. D., St. Luce, M., & Tremblay, N. (2016). Corn yield components response to nitrogen fertilizer as a function of soil texture. *Canadian Journal of Soil Science*, *96*(4), 386–399. https://doi.org/10.1139/cjss-2015-0134

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Kopp, R. E., McCusker, K. E., Nath, I., et al. (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics*, *137*(4), 2037–2105.

Carranza, E. (2014). Soil Endowments, Female Labor Force Participation, and the Demographic Deficit of Women in India. *American Economic Journal: Applied Economics*, *6*(4), 197–225. https://doi.org/10.1257/app.6.4.197

Cassou, E., Tran, D. N., Nguyen, T. H., Dinh, T. X., Nguyen, C. V., Cao, B. T., Jaffee, S., & Ru, J. (2017). An overview of agricultural pollution in vietnam: Summary report 2017.

Chakraborty, P., Chopra, A., & Contractor, L. (2024). The equilibrium impact of agricultural support prices and input subsidies.

Chay, K. Y., & Greenstone, M. (2003). The impact of air pollution on infant mortality: Evidence from geographic variation in pollution shocks induced by a recession. *The Quarterly Journal of Economics*, *118*(3), 1121–1167.

Chen, T.-H. J. (2026). Glyphosate use, water contamination, and neonatal health in the united states. *Journal of the Association of Environmental and Resource Economists*, *13*(3), 755–790.

Churchman, G. J. (2018). Game changer in soil science: Functional role of clay minerals in soil. *Journal of Plant Nutrition and Soil Science*, *181*(1), 99–103. https://doi.org/10.1002/jpln.201700605

Coffman, V. R., Jensen, A. S., Trabjerg, B. B., Pedersen, C. B., Hansen, B., Sigsgaard, T., Olsen, J., Schullehner, J., Pedersen, M., & Stayner, L. T. (2022). Prenatal exposure to nitrate from drinking water and the risk of preterm birth: A danish nationwide cohort study. *Environmental Epidemiology*, *6*(5), e223.

Currie, J., Graff Zivin, J., Meckel, K., Neidell, M., & Schlenker, W. (2013). Something in the water: Contaminated drinking water and infant health. *Canadian Journal of Economics/Revue canadienne d'économique*, *46*(3), 791–810.

Currie, J., & Neidell, M. (2005). Air pollution and infant health: What can we learn from california's recent experience? *The Quarterly Journal of Economics*, *120*(3), 1003–1030.

Currie, J., Neidell, M., & Schmieder, J. F. (2009). Air pollution and infant health: Lessons from new jersey. *Journal of health economics*, *28*(3), 688–703.

Cutler, D. M., & Miller, G. (2005). The role of public health improvements in health advances: The twentieth-century united states. *Demography*, *42*(1), 1–22. https://doi.org/10.1353/dem.2005.0002

Damodaran, H. (2024). Explained: Argument for deregulating non-subsidised fertilisers. *The Indian Express*. Retrieved July 12, 2024, from https://indianexpress.com/

article/explained/explained-economics/deregulating-non-subsidised-fertilisers-9447888/

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. (2019). The mortality and medical costs of air pollution: Evidence from changes in wind direction. *American Economic Review*, *109*(12), 4178–4219.

Dias, M., Rocha, R., & Soares, R. R. (2023). Down the River: Glyphosate Use in Agriculture and Birth Outcomes of Surrounding Populations. *Review of Economic Studies*, *90*(6), 2943–2981. https://doi.org/10.1093/restud/rdad011

Do, Q.-T., Joshi, S., & Stolper, S. (2018). Can environmental policy reduce infant mortality? Evidence from the Ganga Pollution Cases. *Journal of Development Economics*, *133*, 306–325. https://doi.org/10.1016/j.jdeveco.2018.03.001

Donovan, K. (2021). The equilibrium impact of agricultural risk on intermediate inputs and aggregate productivity. *The Review of Economic Studies*, *88*(5), 2275–2307.

Ebenstein, A. (2012). The consequences of industrialization: Evidence from water pollution and digestive cancers in china. *Review of Economics and Statistics*, *94*(1), 186–201.

Fuller, R., Landrigan, P. J., Balakrishnan, K., Bathan, G., Bose-O'Reilly, S., Brauer, M., Caravanos, J., Chiles, T., Cohen, A., Corra, L., et al. (2022). Pollution and health: A progress update. *The Lancet Planetary Health*, *6*(6), e535–e547.

Galiani, S., Gertler, P., & Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, *113*(1), 83–120. https://doi.org/10.1086/426041

Gamper-Rabindran, S., Khan, S., & Timmins, C. (2010). The impact of piped water provision on infant mortality in brazil: A quantile panel data approach. *Journal of Development Economics*, *92*(2), 188–200.

Garg, S., & Saxena, S. (2023). Distributional effects of indian agricultural interventions [Working Paper].

Garg, T., Hamilton, S. E., Hochard, J. P., Kresch, E. P., & Talbot, J. (2018). (not so) gently down the stream: River pollution and health in indonesia. *Journal of Environmental Economics and Management*, *92*, 35–53.

Ghosal, S. (2023). Profitability for phosphatic fertiliser players to improve at current nutrient-based subsidy rates: Report. *The Economic Times*. Retrieved January 19, 2023, from https://economictimes.indiatimes.com/industry/indl-goods/svs/chem-/-fertilisers/profitability-for-phosphatic-fertiliser-players-to-

improve-at-current-nutrient-based-subsidy-rates-report/articleshow/97122078.cms

Graff Zivin, J., & Neidell, M. (2013). Environment, health, and human capital. *Journal of Economic Literature*, *51*(3), 689–730. https://doi.org/10.1257/jel.51.3.689

Greenstone, M., & Hanna, R. (2014). Environmental Regulations, Air and Water Pollution, and Infant Mortality in India. *American Economic Review*, *104*(10), 3038–3072. https://doi.org/10.1257/aer.104.10.3038

Gulati, A., & Banerjee, P. (2015). *Rationalising fertiliser subsidy in India: Key issues and policy options* (tech. rep.). Working Paper, No. 307, Indian Council for Research on International Economic Relations (ICRIER), New Delhi.

Hagerty, N., & Tiwari, A. (2022). Industrial water pollution and agricultural production in india. *Working Paper*.

Heberger, M. (2022). Global Watersheds [Web Application].

Holden, S. T. (2019). Economics of farm input subsidies in africa. *Annual Review of Resource Economics*, *11*(1), 501–522.

Hsiang, S., Oliva, P., & Walker, R. (2019). The distribution of environmental damages. *Review of Environmental Economics and Policy*.

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly Journal of Economics*, *124*(4), 1403–1448.

Huang, J., Gulati, A., & Gregory, I. (2017). Fertilizer subsidies-which way forward?

Itin-Shwartz, B. (2024). Survey based assessment of sustainable agricultural practices: Evidence from indian plots. *Agribusiness*, *40*(2), 416–457.

Jalan, J., & Somanathan, E. (2008). The importance of being informed: Experimental evidence on demand for environmental quality. *Journal of development Economics*, *87*(1), 14–28.

Jensen, A. S., Coffman, V. R., Schullehner, J., Trabjerg, B. B., Pedersen, C. B., Hansen, B., Olsen, J., Pedersen, M., Stayner, L. T., & Sigsgaard, T. (2023). Prenatal exposure to tap water containing nitrate and the risk of small-for-gestational-age: A nationwide register-based study of danish births, 1991–2015. *Environment International*, *174*, 107883. https://doi.org/10.1016/j.envint.2023.107883

Kanter, D. R., Chodos, O., Nordland, O., Rutigliano, M., & Winiwarter, W. (2020). Gaps and opportunities in nitrogen pollution policies around the world. *Nature Sustainability*, *3*(11), 956–963.

Kanwar, J., Goswami, N., & Kamath, M. (1982). Phosphorus management of indian soils-problems and prospects. *Fertiliser News*, *27*(2), 43–52.

Karunanidhi, D., Aravinthasamy, P., Subramani, T., Kumar, D., & Venkatesan, G. (2021). Chromium contamination in groundwater and sobol sensitivity model based human health risk evaluation from leather tanning industrial region of south india. *Environmental Research*, *199*, 111238.

Keiser, D. A., Kling, C. L., & Shapiro, J. S. (2019). The low but uncertain measured benefits of us water quality policy. *Proceedings of the national academy of sciences*, *116*(12), 5262–5269.

Keiser, D. A., & Shapiro, J. S. (2018). Consequences of the Clean Water Act and the Demand for Water Quality. *Quarterly Journal of Economics*, *134*(1), 349–396.

Kishore, A., Alvi, M., & Krupnik, T. J. (2021). Development of balanced nutrient management innovations in South Asia: Perspectives from Bangladesh, India, Nepal, and Sri Lanka. *Global Food Security*, *28*, 100464. https://doi.org/10.1016/j.gfs.2020.100464

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, *133*(1), 97–126. https://doi.org/10.1016/j.jeconom.2005.02.011

Knittel, C. R., Miller, D. L., & Sanders, N. J. (2016). Caution, drivers! children present: Traffic, pollution, and infant health. *Review of Economics and Statistics*, *98*(2), 350–366.

Knobeloch, L., Salna, B., Hogan, A., Postle, J., & Anderson, H. (2000). Blue babies and nitrate-contaminated well water. *Environmental health perspectives*, *108*(7), 675–678.

Kurdi, S., Mahmoud, M., Abay, K. A., & Breisinger, C. (2020). *Too much of a good thing? evidence that fertilizer subsidies lead to overapplication in egypt* (Vol. 27). Intl Food Policy Res Inst.

Lai, W. (2017). Pesticide use and health outcomes: Evidence from agricultural water pollution in china. *Journal of environmental economics and management*, *86*, 93–120.

Li, D., Zhai, Y., Lei, Y., Li, J., Teng, Y., Lu, H., Xia, X., Yue, W., & Yang, J. (2021). Spatiotemporal evolution of groundwater nitrate nitrogen levels and potential human health risks in the songnen plain, northeast china. *Ecotoxicology and Environmental Safety*, *208*, 111524.

Lin, L., St Clair, S., Gamble, G. D., Crowther, C. A., Dixon, L., Bloomfield, F. H., & Harding, J. E. (2023). Nitrate contamination in drinking water and adverse

reproductive and birth outcomes: A systematic review and meta-analysis. *Scientific Reports*, *13*(1), 1–18. https://doi.org/10.1038/s41598-022-27345-x

Lipsey, R. G., & Lancaster, K. (1956). The general theory of second best. *The review of economic studies*, *24*(1), 11–32.

Liu, C.-W., Sung, Y., Chen, B.-C., & Lai, H.-Y. (2014). Effects of nitrogen fertilizers on the growth and nitrate content of lettuce (lactuca sativa l.) *International journal of environmental research and public health*, *11*(4), 4427–4440.

Liu, P., Wang, Y., & Zhang, W. (2023). The influence of the environmental quality incentives program on local water quality. *American Journal of Agricultural Economics*, *105*(1), 27–51.

Majumdar, D. (2003). The blue baby syndrome: Nitrate poisoning in humans. *Resonance*, *8*(10), 20–30.

Mateo-Sagasta, J., Zadeh, S. M., Turral, H., & Burke, J. (2017). *Water pollution from agriculture: A global review. executive summary*. Rome, Italy: FAO Colombo, Sri Lanka: International Water Management.

McMahon, K., & Gray, C. (2021). Climate change, social vulnerability and child nutrition in south asia. *Global Environmental Change*, *71*, 102414.

Mohapatra, J. K. (2022). Biting the bullet: Direct benefit transfer of fertiliser subsidy. *IC Centre for Governance*, *171*, 171.

Mueller, N. D., Lassaletta, L., Runck, B. C., Billen, G., Garnier, J., & Gerber, J. S. (2017). Declining spatial efficiency of global cropland nitrogen allocation. *Global Biogeochemical Cycles*, *31*(2), 245–257.

Musei, S. K., Kuyah, S., Nyawira, S., Ng'ang'a, S. K., Karugu, W. N., Smucker, A., & Nkurunziza, L. (2024). Sandy soil reclamation technologies to improve crop productivity and soil health: A review. *Frontiers in Soil Science*, *4*, 1345895. https://doi.org/10.3389/fsoil.2024.1345895

Patel, A. (2016). Subsidies to farm sector in india need to rationalize & its efficient use. *Int. J. Sci. Res. Sci. Eng. Technol*, *2*, 1–11.

Pawar, N. D., Singhal, K., Bhushan, C., Pregger, T., & Jochem, P. (2025). Decarbonization of urea production in india and its impact on water withdrawal and costs: A cost optimization approach. *Journal of Cleaner Production*, *486*, 144433.

Praveen, K. (2021). Indian fertiliser policy: Retrospect and prospect. In *Soil science: Fundamentals to recent advances* (pp. 765–779). Springer.

Pullabhotla, H. K., & Souza, M. (2022). Air pollution from agricultural fires increases hypertension risk. *Journal of Environmental Economics and Management*, *115*, 102723.

Ravinutala, S. (2016). *Redesigning India's Urea Policy* [MPA/ID Thesis]. John F. Kennedy School of Government, Harvard University.

Restuccia, D., Yang, D. T., & Zhu, X. (2008). Agriculture and aggregate productivity: A quantitative cross-country analysis. *Journal of Monetary Economics*, *55*(2), 234–250.

Royal, H., 't Mannetje, A., Hales, S., Douwes, J., Berry, M., & Chambers, T. (2024). Nitrate in drinking water and pregnancy outcomes: A narrative review of epidemiological evidence and proposed biological mechanisms. *PLOS water*, *3*(1), e0000214.

Sankhla, M., & Kumar, R. (2018). Fluoride contamination of water in india and its impact on public health.

Schlenker, W., & Walker, W. R. (2016). Airports, air pollution, and contemporaneous health. *The Review of Economic Studies*, *83*(2), 768–809.

Shagun. (2023). Cacp recommends centre to bring urea under nbs regime to check overuse. *DownToEarth*. Retrieved June 8, 2023, from https://www.downtoearth.org.in/agriculture/cacp-recommends-centre-to-bring-urea-under-nbs-regime-to-check-overuse-89907

Sherris, A. R., Baiocchi, M., Fendorf, S., Luby, S. P., Yang, W., & Shaw, G. M. (2021). Nitrate in drinking water during pregnancy and spontaneous preterm birth: A retrospective within-mother analysis in california. *Environmental health perspectives*, *129*(5), 057001.

Skidmore, M., Andarge, T., & Foltz, J. (2023). Effectiveness of local regulations on nonpoint source pollution: Evidence from wisconsin dairy farms. *American Journal of Agricultural Economics*, *105*(5), 1333–1364.

Skidmore, M. E., Sims, K. M., & Gibbs, H. K. (2023). Agricultural intensification and childhood cancer in brazil. *Proceedings of the National Academy of Sciences*, *120*(45), e2306003120. https://doi.org/10.1073/pnas.2306003120

Stayner, L. T., Almberg, K., Jones, R., Graber, J., Pedersen, M., & Turyk, M. (2017). Atrazine and nitrate in drinking water and the risk of preterm delivery and low birth weight in four midwestern states. *Environmental Research*, *152*, 294–303. https://doi.org/10.1016/j.envres.2016.10.022

Verma, A., Sharma, A., Kumar, R., & Sharma, P. (2023). Nitrate contamination in groundwater and associated health risk assessment for indo-gangetic plain, india. *Groundwater for sustainable development*, *23*, 100978.

Viscusi, W. K., & Masterman, C. J. (2017). Income elasticities and global values of a statistical life. *Journal of Benefit-Cost Analysis*, *8*(2), 226–250.

von Der Goltz, J., Dar, A., Fishman, R., Mueller, N. D., Barnwal, P., & McCord, G. C. (2020). Health impacts of the green revolution: Evidence from 600,000 births across the developing world. *Journal of health economics*, *74*, 102373.

von Liebig, J. F. (1843). *Chemistry in its application to agriculture and physiology*. JM Campbell.

Wang, Z.-H., & Li, S.-X. (2019). Nitrate n loss by leaching and surface runoff in agricultural land: A global issue (a review). *Advances in Agronomy*, *156*, 159–217. https://doi.org/10.1016/bs.agron.2019.01.007

Ward, M. H., Jones, R. R., Brender, J. D., De Kok, T. M., Weyer, P. J., Nolan, B. T., Villanueva, C. M., & Van Breda, S. G. (2018). Drinking water nitrate and human health: An updated review. *International Journal of Environmental Research and Public Health*, *15*(7). https://doi.org/10.3390/ijerph15071557

World Bank. (2024). World development indicators [Accessed on October 20, 2026]. *World Bank Group*.

Wuepper, D., Le Clech, S., Zilberman, D., Mueller, N., & Finger, R. (2020). Countries influence the trade-off between crop yields and nitrogen pollution. *Nature Food*, *1*(11), 713–719.

Online Appendix

A Expanded descriptive statistics

		Means by soil clay			Within state-yr means by soil clay		
	All	Low Clay	High Clay	Diff	Low Clay	High Clay	Diff
			Panel A: Time invariant data				
Fraction male births (%)	0.53	0.53	0.52	0.01***	0.00	-0.00	0.00
	(0.08)	(0.08)	(0.08)		(0.08)	(0.08)	
Mother illiterate (%)	0.39	0.40	0.38	0.02***	0.01	-0.00	0.01
	(0.26)	(0.26)	(0.26)		(0.22)	(0.23)	
Mother educ.: none (%)	0.34	0.37	0.33	0.04***	0.01	-0.00	0.01**
	(0.26)	(0.26)	(0.26)		(0.22)	(0.22)	
Mother educ.: primary (%)	0.16	0.16	0.16	-0.00	-0.00	0.00	-0.00
	(0.13)	(0.13)	(0.13)		(0.12)	(0.12)	
Mother educ.: secondary+ (%)	0.50	0.47	0.51	-0.04***	-0.01	0.00	-0.01*
	(0.27)	(0.27)	(0.27)		(0.23)	(0.24)	
Mother Hindu (%)	0.77	0.76	0.78	-0.01*	-0.01	0.01	-0.02**
	(0.33)	(0.32)	(0.33)		(0.28)	(0.27)	
H.H. toilet: flush (%)	0.58	0.60	0.56	0.04***	0.01	-0.00	0.01
	(0.35)	(0.35)	(0.34)		(0.28)	(0.30)	
H.H. toilet: pit/other (%)	0.13	0.11	0.15	-0.04***	-0.01	0.01	-0.02***
	(0.22)	(0.19)	(0.23)		(0.14)	(0.16)	
H.H. toilet: none (%)	0.29	0.29	0.29	-0.00	0.00	-0.00	0.01
	(0.33)	(0.33)	(0.33)		(0.25)	(0.28)	
H.H. has electricity (%)	0.88	0.86	0.90	-0.05***	-0.00	0.00	-0.01
	(0.19)	(0.22)	(0.17)		(0.18)	(0.15)	
H.H. has fridge (%)	0.34	0.37	0.32	0.05***	0.00	-0.00	0.00
	(0.32)	(0.35)	(0.30)		(0.24)	(0.25)	
H.H. has car/truck (%)	0.08	0.10	0.07	0.03***	0.00	-0.00	0.00
	(0.14)	(0.16)	(0.13)		(0.14)	(0.12)	
Rural (%)	0.61	0.68	0.56	0.11***	-0.00	0.00	-0.01
	(0.49)	(0.47)	(0.50)		(0.44)	(0.47)	
Clay (%)	27.77	20.45	32.62	-12.17***	-1.30	0.86	-2.16***
	(7.80)	(3.05)	(6.00)		(2.38)	(3.64)	
Sand (%)	39.48	40.22	38.99	1.23***	0.09	-0.06	0.15*
	(5.12)	(3.82)	(5.77)		(2.36)	(3.84)	
Silt (%)	30.01	33.78	27.51	6.27***	1.05	-0.70	1.74***
	(4.94)	(4.80)	(3.10)		(3.08)	(2.02)	

		Means by soil clay			Within state-yr means by soil clay		
	All	Low Clay	High Clay	Diff	Low Clay	High Clay	Diff
		Panel B: Time varying data					
Infant deaths	0.08	0.09	0.06	0.03***	0.00	-0.00	0.00
	(0.29)	(0.33)	(0.26)		(0.32)	(0.26)	
Births	1.66	1.80	1.56	0.24***	-0.00	0.00	-0.00
	(1.42)	(1.46)	(1.38)		(1.42)	(1.33)	
Max. temp (°C)	28.72	26.83	30.14	-3.31***	0.07	-0.05	0.12**
	(4.30)	(5.54)	(2.15)		(2.59)	(1.55)	
Max. D.O. (mg/L)	8.13	8.40	7.93	0.47***	0.02	-0.02	0.04
	(1.58)	(1.89)	(1.25)		(1.56)	(0.99)	
Max. pH	8.12	8.06	8.17	-0.11***	0.01	-0.01	0.01*
	(0.44)	(0.38)	(0.47)		(0.28)	(0.33)	
Max. cond.	1,622	536.4	2,465	-1,929***	21.30	-16.54	37.84
	(4,807)	(340.3)	(6,273)		(258.8)	(5,320)	
Max. C.O.D. (mg/L)	25.27	26.44	23.89	2.56	3.00	-3.51	6.52***
	(32.60)	(40.59)	(19.40)		(38.00)	(15.12)	
Max. B.O.D. (mg/L)	7.30	7.63	7.05	0.58	0.51	-0.38	0.89**
	(14.47)	(18.63)	(10.28)		(17.72)	(6.58)	
Max. nitrate (mg/L)	3.06	2.20	3.70	-1.50***	0.14	-0.11	0.25***
	(5.24)	(2.16)	(6.60)		(1.54)	(3.83)	
Max. nitrite (mg/L)	0.91	1.04	0.81	0.22***	0.07	-0.05	0.12***
	(1.70)	(0.91)	(2.10)		(0.69)	(1.59)	
Max. ammonia (mg/L)	2.36	0.90	3.66	-2.76***	-0.06	0.05	-0.11*
	(3.67)	(0.59)	(4.65)		(0.47)	(1.41)	
Max. fecal coliform (100 mg/L)	204.6	342.6	91.1	251.5***	7.96	-6.54	14.5
	(707.5)	(999.1)	(251.4)		(937.1)	(185.4)	
Max. total coliform (100 mg/L)	2,411	5,824	166.1	5,658***	1,122	-737.6	1,860**
	(25,176)	(39,729)	(508.8)		(38,577)	(3,605)	
Max. NO_3/NO_2 (mg/L)	3.88	3.33	4.29	-0.96***	0.21	-0.16	0.37***
	(6.27)	(2.86)	(7.90)		(2.07)	(4.86)	

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports expanded descriptive statistics for time invariant data (Panel A) and time varying data (Panel B). Table structure follows that of Table 1 in the main text. Note that because soil composition affects runoff to rivers, within state-year differences between low and high clay regions are to be expected.

B Production function estimates

We estimate farmers' returns to fertilizer application and the extent to which marginal returns vary with clay soil content using plot-level panel data from the Cost of Cultivation Surveys. Tables B.1 and B.2 report estimates from a translog production function using log yield and log revenue as outcomes, respectively. For both, we run estimates separately for the six major crops in India - rice, wheat, maize, soybean, sugarcane and cotton (columns (2) - (7)) - as well as pooled results for all crops (column (1)). Nitrogen, phosphorus, and potassium enter as main effects, and the specifications include selected pairwise interactions between nitrogen and clay content, phosphorus, and potassium. All input variables are centered on their sample means, so the coefficient on nitrogen can be interpreted as the average marginal return to nitrogen when other inputs are evaluated at their mean levels. To keep the specification parsimonious and avoid instability from overfitting, we do not include the full set of other two-way and higher-order interactions. Our specifications include farmer-crop and state-crop-year fixed effects, accounting for constant plot-level attributes such as soil endowments and farmer ability, as well as state-specific policy changes that may affect local agricultural input and output markets.

Across all crops, we find that returns to nitrogen application remain positive (for both yield and revenue) and increase with clay content. These results complement the findings from soil science literature and provide additional support for our research design, leveraging soil clay content as a shifter for fertilizer use.

Table B.1. Translog production function estimates: Crop yield

	(1) All	(2) Rice	(3) Wheat	(4) Maize	(5) Soybean	(6) Sugarcane	(7) Cotton
log N (kg)	0.090***	0.106***	0.075***	0.093***	0.024	0.043***	0.120***
	(0.004)	(0.006)	(0.007)	(0.019)	(0.013)	(0.009)	(0.016)
log P (kg)	0.016***	0.017***	0.016***	0.035***	-0.002	0.014***	-0.003
	(0.002)	(0.003)	(0.004)	(0.009)	(0.010)	(0.003)	(0.009)
log K (kg)	0.024***	0.027***	0.007*	0.032***	-0.001	0.022***	0.034***
	(0.002)	(0.003)	(0.003)	(0.007)	(0.008)	(0.004)	(0.005)
log N (kg) × log clay (%)	0.009*	0.015*	0.010	0.040	-0.039	0.032*	0.086***
	(0.005)	(0.007)	(0.005)	(0.031)	(0.038)	(0.013)	(0.017)
log N (kg) × log P (kg)	0.004***	0.003*	0.003*	0.017**	0.001	-0.001	0.004
	(0.001)	(0.001)	(0.002)	(0.006)	(0.003)	(0.002)	(0.004)
log N (kg) × log K (kg)	0.001	0.004***	-0.001	-0.018***	0.006	0.004	-0.013***
	(0.001)	(0.001)	(0.002)	(0.005)	(0.004)	(0.002)	(0.004)
log area (ha)	0.136***	0.164***	0.105***	0.144***	0.037**	0.068***	0.092***
	(0.005)	(0.006)	(0.009)	(0.023)	(0.014)	(0.013)	(0.019)
N	251,591	132,989	63,728	13,702	13,695	13,442	14,035
Adjusted R^2	0.92	0.69	0.79	0.74	0.57	0.61	0.47
State-Crop-Year FE	X	X	X	X	X	X	X
Farmer-Crop FE	X	X	X	X	X	X	X

Notes: This table shows the coefficients on production function parameters for a translog-like production function estimated using crop-specific farmer-plot-level input and output data from the Cost of Cultivation Surveys. The data consist of plot-level panel data, with each farmer followed for three years in the survey. All independent variables were centered on their means before estimation. Standard errors in parentheses, adjusted for clustering at the district-year level. Significance level at 10, 5, and 1 percent shown by *, **, and ***, respectively.

Table B.2. Translog production function estimates: Crop revenue

Table B.3. log(Revenue)

	(1) All	(2) Rice	(3) Wheat	(4) Maize	(5) Soybean	(6) Sugarcane	(7) Cotton
log N (Rs)	0.044***	0.053***	0.039**	0.050*	0.025	0.031	0.100***
	(0.006)	(0.008)	(0.013)	(0.020)	(0.020)	(0.019)	(0.027)
log P (Rs)	0.024***	0.025***	0.017***	0.029	-0.002	0.022***	0.006
	(0.003)	(0.003)	(0.004)	(0.020)	(0.017)	(0.006)	(0.011)
log K (Rs)	0.025***	0.033***	0.004	0.041***	-0.001	0.020**	0.037***
	(0.002)	(0.003)	(0.003)	(0.011)	(0.008)	(0.007)	(0.007)
log N (Rs) × log clay (%)	0.010	0.017*	0.002	0.020	0.037	0.052**	0.070***
	(0.007)	(0.008)	(0.005)	(0.031)	(0.058)	(0.019)	(0.020)
log N (Rs) × log P (Rs)	-0.003*	-0.003	-0.001	0.004	0.000	-0.008	-0.014**
	(0.001)	(0.001)	(0.002)	(0.007)	(0.009)	(0.006)	(0.005)
log N (Rs) × log K (Rs)	0.003*	0.002	0.001	-0.003	-0.003	0.008	-0.003
	(0.001)	(0.002)	(0.002)	(0.007)	(0.005)	(0.006)	(0.005)
log area (ha)	0.086***	0.106***	0.063***	0.082***	0.030	0.060**	0.074**
	(0.007)	(0.008)	(0.014)	(0.024)	(0.027)	(0.021)	(0.028)
N	251,703	133,024	63,728	13,747	13,725	13,444	14,035
Adjusted R^2	0.77	0.72	0.83	0.63	0.48	0.58	0.52
State-Crop-Year FE	X	X	X	X	X	X	X
Farmer-Crop FE	X	X	X	X	X	X	X

Notes: This table shows the coefficients on production function parameters for a translog-like production function estimated using crop-specific farmer-plot-level input and output data from the Cost of Cultivation Surveys. The data consist of plot-level panel data, with each farmer followed for three years in the survey. All independent variables were centered on their means before estimation. Standard errors in parentheses, adjusted for clustering at the district-year level. Significance level at 10, 5, and 1 percent shown by *, **, and ***, respectively.

C Comparison with related literature

Appendix Table C.1 places our estimates alongside selected estimates from the related literature. We note that studies differ in outcomes, exposure definitions, and empirical design and estimates are not directly comparable. For that reason, we report each study in its original units.

Although the estimates across these studies are not directly comparable, they suggest that the magnitudes implied by our findings are economically meaningful and broadly plausible relative to the existing literature. Using the IV coefficients, first stage coefficients and sample means from Tables 4 and D.1, our estimates imply that a 10 percent increase in nitrate around the sample mean raises the number of infant deaths per DHS cluster-year by roughly 17.7 percent in the full sample and 25.4 percent in the rural sample.[28] These magnitudes are larger than those in Brainerd and Menon (2014), who find that a one unit increase in their measure of average state-level agrichemical presence increases infant mortality by 0.08 units which implies a 10 percent increase in agrichemicals in water leads to a 4.64 percent increase in infant mortality in India. The larger magnitude of impact that we find is not surprising given substantial differences in measurement and design. Their exposure is a state-level agrichemical index across multiple pollutants rather than specific chemical concentrations measured in mg/L, and their estimates are likely attenuated because exposure is assigned at a much more aggregate level.

At the same time, our estimates are qualitatively consistent with the broader evidence that nitrate-contaminated water can adversely affect early-life health even at concentrations below current regulatory thresholds (Ward et al., 2018). Direct nitrate studies in the epidemiological literature generally focus on birth outcomes rather than mortality, but they likewise find adverse effects of elevated nitrate exposure on preterm birth, low birth weight, fetal growth restriction, and some congenital anomalies. For example, Stayner et al. (2017) find adverse effects of prenatal nitrate exposure on fetal health outcomes in the United States Midwest, while Jensen et al. (2023) report an increased risk of small-for-gestational-age birth in Denmark at nitrate concentrations as low as 2–5 mg/L. More broadly, Lin et al. (2023) summarize evidence linking higher nitrate exposure to increased risk of preterm birth in a meta-analysis.

In terms of overall water quality, studies such as Galiani et al. (2005) in Argentina and Cutler and Miller (2005) in the early twentieth-century United States show that

[28] These calculations use the elasticity at sample means, $\beta\bar{x}/\bar{y}$, where β is the IV coefficient, $\bar{x}$ is mean nitrate, and $\bar{y}$ is mean infant deaths in the IV sample.

improvements in water quality can produce substantial reductions in child and infant mortality. In India, however, Greenstone and Hanna (2014) find no measurable infant-mortality benefits from broad water regulations. Their null result pertains to the average effect of water-quality regulation in a setting where implementation and enforcement appear to have been limited, whereas our design isolates a specific post-NBS pathway from fertilizer use to nitrate contamination and infant mortality. Relatedly, Brainerd and Menon (2014) note that India's water regulations were primarily directed at industrial pollution and sewage treatment rather than agrichemical runoff, and that implementation of water-quality policy appears to have been weak and underfunded. Taken together, these comparisons suggest that our estimates are large relative to the direct nitrate literature, but they are not out of line with the wider evidence that water quality can have first-order effects on infant survival, especially in rural settings with limited treatment or avoidance options.

Table C.1. Comparison of Magnitudes with Related Literature

Study	Setting	Exposure	Outcome	Estimate
Panel A. Nitrate or fertilizer-related water contamination				
This paper (full sample)	India	1 mg/L increase in nitrate	Infant deaths (< 12 months), cluster-year count	+0.0400 (0.0215)
This paper (rural sample)	India	1 mg/L increase in nitrate	Infant deaths (< 12 months), cluster-year count	+0.0849 (0.0299)
Brainerd and Menon (2014)	India	1 unit increase in state-level average agrichemicals in water	Infant mortality; neonatal mortality	+0.08 infant mortality; +0.07 neonatal mortality
Panel B. Broader water-quality interventions				
Galiani et al. (2005)	Argentina	Water privatization	Child mortality	−8% overall; −26% in poorest areas
Greenstone and Hanna (2014)	India	Water regulation	Infant mortality	No measurable effect
Cutler and Miller (2005)	U.S.	Filtration and chlorination	Infant mortality	Clean water explains 59% of the infant mortality decline, 1900–1936

Notes: The table reports selected estimates from the literature.

D Additional Figures and Tables

Table D.1. First-stage Regressions

	Full Sample	Rural	Urban
	(1)	(2)	(3)
Upstream Clay x Post	1.360***	1.451***	1.244***
	(0.255)	(0.191)	(0.321)
Observations	31,162	18,260	12,880
Dep var mean	3.192	2.712	3.872
No. of clusters	6,948	4,289	2,655
KP F-Stat	43.81	57.54	15.02
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports first stage estimates associated with the IV estimates in Table 4. The sample consists of DHS clusters from 2007 to 2014. The dependent variable is nitrate concentration (the average of three nearest stations within 20 km to each DHS cluster). Upstream Clay refers to DHS clusters with clay levels above the median in their upstream watershed regions and is a dummy variable with the value 1 for high clay upstream and 0 otherwise. Upstream Clay interacted with the post-NBS indicator of a reference DHS cluster. Standard errors (in parentheses) are clustered at the DHS cluster level. All regressions include controls for observables. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic on the excluded instrumental variables. Column 1 includes the full sample of DHS clusters located within 20 km of a water monitoring station. Column 2 and 3 only uses the rural and urban clusters respectively.

Table D.2. First-stage Regressions (no observables)

	Full Sample	Rural	Urban
	(1)	(2)	(3)
Upstream Clay x Post	1.352***	1.614***	1.160***
	(0.149)	(0.152)	(0.238)
Observations	31,165	18,264	12,891
Dep var mean	3.192	2.713	3.870
No. of clusters	6,948	4,289	2,658
KP F-Stat	79.41	42.37	23.68
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports first stage estimates, following Table D.1 but without including controls for observables. Sample consists of DHS clusters from 2007 to 2014. The dependent variable is nitrate concentration (the average of three nearest stations within 20 km to each DHS cluster). Upstream Clay refers to DHS clusters with clay levels above the median in their upstream watershed regions and is a dummy variable with the value 1 for high clay upstream and 0 otherwise. Upstream Clay interacted with the post-NBS indicator of a reference DHS cluster. Standard errors (in parentheses) are clustered at the DHS cluster level. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic on the excluded instrumental variables. Column 1 includes the full sample of DHS clusters located within 20kms of a water monitoring station. Column 2 and 3 only uses the rural and urban clusters respectively.

Table D.3. Nitrogen pollution and infant mortality (no observables)

	Reduced Form			IV		
	Full Sample	Rural	Urban	Full Sample	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)
Upstream Clay x Post	0.0424	0.0879**	-0.0394			
	(0.0260)	(0.0367)	(0.0377)			
Nitrates				0.0320*	0.0513**	-0.0238
				(0.0191)	(0.0237)	(0.0315)
Observations	31,165	18,264	12,891	31,165	18,264	12,891
Dep var mean	0.0723	0.0907	0.0463	0.0723	0.0907	0.0463
KP F-Stat				79.41	42.37	23.68
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports coefficient estimates from reduced form specifications and instrumental variables, following Table 4 but without including controls for observables. Standard errors (in parentheses) are clustered at the DHS cluster level. The first three columns report estimates for the full sample, rural and urban samples of DHS clusters under the reduced form specification. Columns 4, 5 and 6 show results from the IV specification for the full sample, rural and urban samples. The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

Table D.4. Binned upstream clay instrument

	Full Sample	Rural	Urban
	(1)	(2)	(3)
Panel A: IV			
Nitrates	0.0276**	0.0603***	0.00009
	(0.0125)	(0.0229)	(0.0107)
Panel B: First Stage			
Upstream Clay (25-30%) × Post	1.349***	1.492***	1.186***
	(0.224)	(0.220)	(0.329)
Upstream Clay (30-35%) × Post	1.391***	1.026***	1.779***
	(0.269)	(0.299)	(0.390)
Upstream Clay (35+%) × Post	3.277***	2.567***	4.535***
	(0.606)	(0.835)	(0.928)
Observations	31,162	18,260	12,880
KP F-Stat	16.18	17.42	9.743
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports coefficient estimates from an alternative instrument which bins upstream soil clay content over 5 percentage point intervals, starting at 25%. Clay content values below 25% respresent the "low-clay" omitted category. All regressions include controls for observables. Standard errors (in parentheses) are clustered at the DHS cluster level. The three columns report estimates for the full sample, and rural and urban samples of DHS clusters under the IV specification (Panel A) along with the associated first stage estimates (Panel B). The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

Table D.5. Rural sample, alternative bandwidths for impact of NBS policy on infant mortality

	10 km	20 km	40 km
	(1)	(2)	(3)
	Panel A: IV		
Nitrates	0.117*	0.0849***	0.102
	(0.0637)	(0.0299)	(0.0781)
	Panel B: Reduced Form		
Upstream Clay x Post	0.154**	0.123***	0.0366
	(0.0775)	(0.0374)	(0.0235)
Observations	6,773	18,290	41,691
Dep var mean	0.0836	0.0905	0.0920
KP F-Stat	21.09	57.55	5.377
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table presents estimated coefficients for the effect of upstream nitrate pollution on infant mortality across different buffer distances around each DHS rural cluster. Columns 1–3 report estimates by averaging water station nitrogen pollution for stations within 10 km, 20 km (main specification), or 40 km of a given DHS cluster. Panel A displays instrumental variable estimates, while Panel B displays the reduced form estimates. Standard errors (in parentheses) are clustered at the DHS cluster level. All regressions include controls for observables.

Table D.6. IV regressions controlling for downstream placebo

	Full Sample	Rural	Urban
	(1)	(2)	(3)
Nitrates	0.0320	0.0673**	-0.0407
	(0.0209)	(0.0264)	(0.0389)
Observations	31,162	18,260	12,880
Dep var mean	0.0723	0.0907	0.0461
No. of clusters	6,948	4,289	2,655
KP F-Stat	42.46	59.07	13.09
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports instrumental variables estimates, following Table 4 but also controlling for a placebo indicator for high-downstream-clay × Post as an exogenous variable in estimation. Standard errors (in parentheses) are clustered at the DHS cluster level. Columns show results from the IV specification for the full sample, rural and urban samples. The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

Table D.7. Downstream placebo regressions controlling for upstream instrument

	Full Sample	Rural	Urban
	(1)	(2)	(3)
Nitrates	-0.0868	-0.0413	0.162
	(0.0665)	(0.0358)	(0.267)
Observations	31,162	18,260	12,880
Dep var mean	0.0723	0.0907	0.0461
No. of clusters	6,948	4,289	2,655
KP F-Stat	4.18	9.80	0.48
Cluster FE	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports instrumental variables estimates of the downstream clay placebo, following Table 5 but also controlling for high-upstream-clay × Post as an exogenous variable in estimation. Standard errors (in parentheses) are clustered at the DHS cluster level. Columns show results from the IV specification for the full sample, rural and urban samples. The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

Table D.8. Downstream placebo for water quality: Impact of NBS policy on water quality measures using downstream soil

	Nitrates	Fecal Coli.	Temperature	DO	pH
	(1)	(2)	(3)	(4)	(5)
Downstream Clay x Post	1.419	317.9	-0.702	0.0446	-0.0250
	(3.967)	(1,264.3)	(0.523)	(0.311)	(0.0948)
Observations	2,215	2,998	3,371	3,383	3,409
Dep var mean	3.081	30,967	28.97	8.063	8.279
Station FE	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Table reports placebo estimates using downstream soil properties. Sample consists of water quality measurements recorded at stations all over India from 2007 to 2014. Downstream Clay refers to water stations with clay levels above the median in the region downstream of the water monitoring station and is a dummy variable with the value 1 for high downstream stations and 0 otherwise. Standard errors (in parentheses) are clustered at the water station level. Column 1 reports coefficients for nitrate pollution, which is recorded as maximum nitrate-nitrite levels in (mg/L). Columns 2 - 5 include other water quality indicators such as maximum fecal coliform (values divided by 100), maximum temperature, dissolved oxygen, and pH.

Table D.9. Impact of NBS policy on infant mortality rates (deaths per live birth)

	Reduced Form			IV		
	Full Sample	Rural	Urban	Full Sample	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)
Upstream Clay x Post	0.0110	0.0115	0.0105			
	(0.0073)	(0.0090)	(0.0135)			
Nitrates				0.0193	0.0403**	-0.0230
				(0.0124)	(0.0171)	(0.0266)
Observations	23,706	14,702	8,981	23,706	14,702	8,981
Dep var mean	0.0411	0.0472	0.0307	0.0411	0.0472	0.0307
KP F-Stat				19.58	35.33	2.72
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports coefficient estimates from reduced form specifications and instrumental variables, following Table 4, but specifying the outcome as infant death rates (deaths per live birth). Standard errors (in parentheses) are clustered at the DHS cluster level. The first three columns report estimates for the full sample, rural and urban samples of DHS clusters under the reduced form specification. Columns 4, 5 and 6 show results from the IV specification for the full sample, rural and urban samples. All regressions include controls for observables. The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

Table D.10. Impact of NBS policy on births

	Reduced Form			IV		
	Full Sample	Rural	Urban	Full Sample	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)
Upstream Clay x Post	0.213**	0.349**	-0.126			
	(0.107)	(0.139)	(0.200)			
Nitrates				0.157*	0.240**	-0.101
				(0.0831)	(0.105)	(0.162)
Observations	31,162	18,260	12,880	31,162	18,260	12,880
Dep var mean	1.70	1.91	1.39	1.70	1.91	1.39
KP F-Stat				43.81	57.54	15.02
Cluster FE	Yes	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Notes: Table reports coefficient estimates from reduced form specifications and instrumental variables, following Table 4, but with births as the outcome. Standard errors (in parentheses) are clustered at the DHS cluster level. The first three columns report estimates for the full sample, rural and urban samples of DHS clusters under the reduced form specification. Columns 4, 5 and 6 show results from the IV specification for the full sample, rural and urban samples. All regressions include controls for observables. The sample is limited to clusters that have at least one monitoring station within a 20km buffer. The KP F-Stat is the Wald version of the Kleibergen and Paap (2006) statistic.

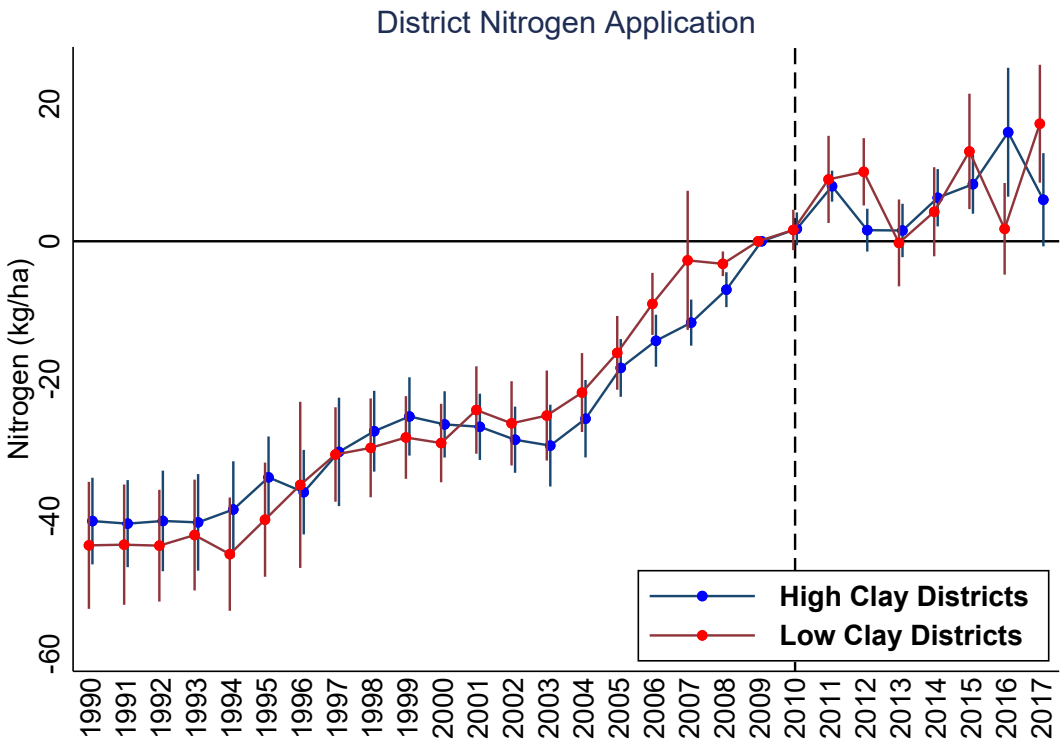
Table D.11. Robustness of water station results: Impact of the NBS policy on nitrate pollution and placebo outcomes

	Nitrates	Fecal Coli.	Temperature	D.O.	pH
	(1)	(2)	(3)	(4)	(5)
	Catchment basin restricted to 100 km				
High Clay x Post	2.778**	-609.3	0.212	0.429*	0.0088
	(1.322)	(727.4)	(0.446)	(0.252)	(0.0501)
Observations	2,756	3,630	4,135	4,151	4,181
Dep var mean	3.046	27,858	28.47	8.140	18.00
Station FE	Yes	Yes	Yes	Yes	Yes
State x Year FE	Yes	Yes	Yes	Yes	Yes

* $p < 0.1$, ** $p < .05$, *** $p < 0.01$

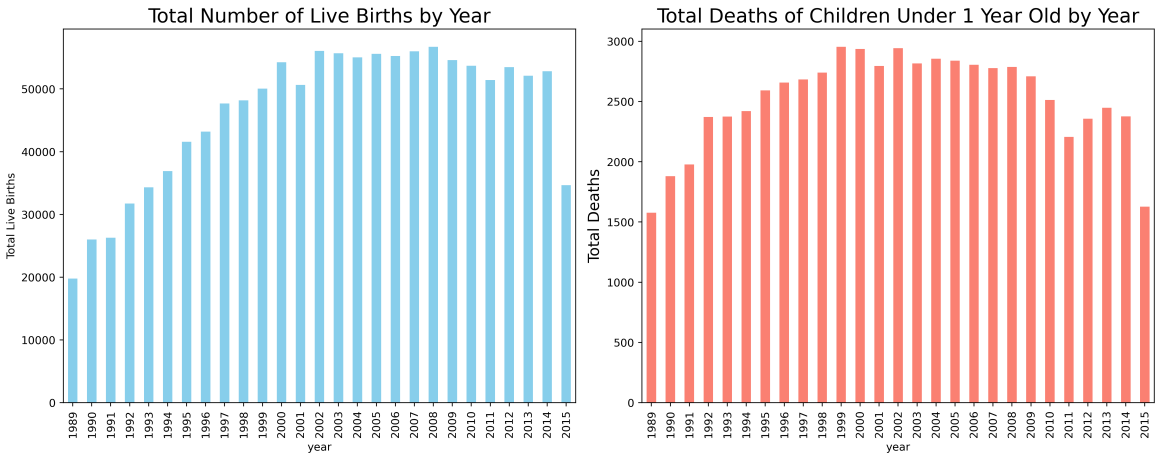
Notes: As Table 3, but the upstream basin is capped at a 100 km radius, serving as a robustness check for the main results. "High Clay" refers to an indicator for water quality monitoring stations with soil clay levels above the median in their upstream catchment basins (see Section 3.7). Standard errors (in parentheses) are clustered at the water station level. Column 1 reports coefficients for nitrate pollution, which is recorded as maximum nitrate-nitrite levels (mg/L). Columns 2 – 5 report placebo water quality indicators such as maximum fecal coliform (values divided by 100), maximum temperature, maximum dissolved oxygen, and maximum pH.

Figure D.1. Nitrogen Use by District Clay Content



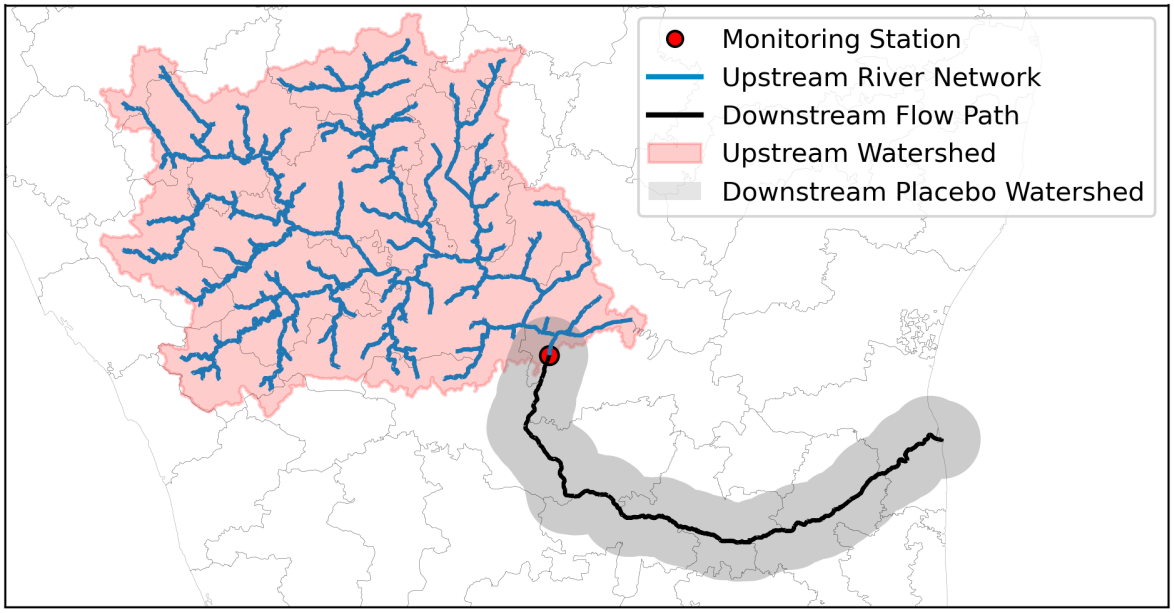
Notes: This figure displays the results of two regressions of district-level nitrogen usage on year fixed effects, estimated separately for high- and low-clay districts. Regressions also includes district fixed effects. 95% confidence intervals displayed are cluster-robust at the district level.

Figure D.2. Total infant births and deaths by year from the 2015/16 DHS survey



Notes: This figure shows the total number of live births and infant deaths within one year of birth as reported by women aged 15-49 in India's 2015/16 DHS survey. Upward trends through about the year 2000 reflect sample selection as women over age 49 are not surveyed. Reported births and infant deaths fully stabilize by the start of our sample period in 2007.

Figure D.3. Illustration of downstream river path



Notes: This figure shows an example downstream flow path (in black) for the highlighted water monitoring station (red circle), which is used to construct the "downstream clay" placebo treatment described in the main text. The flow path consists of the entire river segment from the focal water station down to its termination at a coastline. The associated "watershed" over which soil clay content is calculated consists of a 10km buffer on both sides of the downstream river segment (shown in grey).