Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from India's Total Sanitation Campaign

Dean Spears*

July 3, 2012

Abstract

Open defecation without a toilet or latrine is among the leading global threats to health, especially in India. Although it is well-known that modern sewage infrastructure improves health, it is unknown whether a sanitation program feasible for a low capacity, poor country government could be effective. This paper contributes the first causally identified estimates of effects of rural sanitation on health and human capital accumulation.

The Indian government's Total Sanitation Campaign reports building one household pit latrine per ten rural persons from 2001 to 2011. The program offered local governments a large *ex post* monetary incentive to eliminate open defecation. I use several complementary identification strategies to estimate the program's effect on children's health. First, I exploit variation in program timing, comparing children born in different years. Second, I study a long difference-in-differences in aggregate mortality. Third, I exploit a discontinuity designed into the monetary incentive.

Unlike many impact evaluations, this paper studies a full-scale program implemented by a large government bureaucracy with low administrative capacity. At the mean program intensity, infant mortality decreased by 4 per 1,000 and children's height increased by 0.2 standard deviations (similar to the cross-sectional difference associated with doubling household consumption per capita). These results suggest that, even in the context of governance constraints, incentivizing local leaders to promote technology adoption can be an effective strategy.

*dspears@princeton.edu. Sneha Lamba provided valuable research assistance. I appreciate conversations with staff at the Ministry of Drinking Water & Sanitation, UNICEF, and the World Bank WSP and with district and village officials. Thanks to Farzana Afridi, Radu Ban, Robert Chambers, Tom Classen, Janet Currie, Angus Deaton, Taryn Dinkelman, Josephine Duh, Sebastian Galiani, Jacob Goldin, Raymond Guiteras, Avinash Kishore, Susanna Makela, Doug Miller, Mushfiq Mobarak, Abhiroop Mukhopadhyay, John Papp, Manisha Shah, Yaniv Stopnitzky, Tom Vogl, Alix Zwane; to seminar participants at ISI Delhi, Princeton, the Gates Foundation, and the World Bank; and especially to Anne Case, Diane Coffey, Michael Geruso, and Jeff Hammer. Remaining errors are my own.

1 Introduction

According to joint UNICEF and WHO (2012) estimates for 2010, 15 percent of people in the world openly defecate without any toilet or latrine; 60 percent of these live in India. The global impact of poor sanitation on infant and child death and health is profound. Black et al. (2003) estimated that 10 million children under 5 die every year – 2.4 million of them in India – and that a fifth to a quarter of these deaths are due to diarrhea. Disease early in life also has lasting effects on the health and human capital of children who survive (Almond and Currie, 2011). Evidence from the history of now-rich countries has demonstrated that complete sanitation infrastructure – sewage pipes and septic tanks – importantly improves health outcomes. However, it is not plausible that these public investments will soon be implemented by the limited capacity states that govern many poor people. Therefore, it would be important to learn the effects of low-cost sanitation programs that could be implemented by poor country bureaucracies.

Well-identified evidence on the effectiveness of sanitation policy tools available in poor countries remains absent from the literature. Much of the policy focus within rural "water and sanitation" programs, and much of the econometric evidence, has been on water supply (Black and Fawcett, 2008). Yet, evidence on the health effects of programs to improve rural water supply is of mixed quality and results (Zwane and Kremer, 2007). The econometric "water and sanitation" literature has largely ignored low-cost strategies for excreta disposal – that is, adequately constructed and used household pit latrines – about which the literature lacks well-identified estimates of causal effects. If people living in poor countries are unlikely to receive sewage pipes, and if improving water supply alone does not eliminate the hazards and negative externalities of open defecation, could a rural sanitation program that is feasible to a poor country government improve human capital accumulation and health production?

This paper estimates effects of India's Total Sanitation Campaign (TSC) on infant mortality and on children's height. This campaign represents a large effort to improve rural

sanitation: over the approximately ten-year period studied, it reports building one latrine per 10 rural people in India; it spent $1.5 billion. The TSC was designed to improve upon perceived shortcomings of earlier programs: instead of emphasizing subsidies for building infrastructure, it included an *ex post* monetary incentive for local political leaders to eliminate open defecation and made use of village social structures. As Ravallion (2012) and others have observed, much of the causally credible evidence of program effectiveness in the literature may concern small programs or ideal policy conditions that may not generalize. The TSC was implemented by the Indian government at central and local levels; many of the estimates presented will be representative of rural India. Thus, this paper studies a full-scale program, and its benefits and costs reflect real implementation.

The TSC caused a decline in infant mortality: at the mean observed program intensity, IMR decreased by about 4 infants per 1,000. Additionally, children who lived their first year of life in years and districts with better sanitation grew taller by about 0.2 standard deviations, on average. This magnitude is comparable to the cross-sectional difference in height associated with doubling household consumption per capita. This adds further evidence of an effect of the disease environment after birth on subsequent height (Bozzoli et al., 2009) to the growing literature on the importance of early life health. The result suggests that poor sanitation could account for part of the widespread and extreme stunting among Indian children.

Estimating causal effects of infrastructure is always challenging, given possibly endogenous construction (Dinkelman, 2011); this paper combines converging evidence of causality from three identification strategies. The first and main results use individual-level data with year and district fixed effects. I match data on infants' survival of their first year of life to district-level administrative data on latrine construction each year. Relative to other children born in the same districts or in the same years, rural children exposed to better sanitation in their first year of life were more likely to survive infancy.

Several falsification tests of this first strategy are consistent with a causal effect. Because

the TSC is a rural program, urban children were not exposed to it: I find no "effect" of the TSC on urban children. Similarly, tests inspired by Granger causality rule out spurious effects of district trends: later latrines have no temporally backwards "effect" on the health of children born before they were constructed. As an additional credibility check, I show that the effect of the TSC is concentrated on post-neonatal mortality, which is sensitive to disease environments, not neonatal mortality in the first month of life, which is less so. Further, the effect is greater for children who ate food other than breast milk earlier in their lives, consistent with sanitation reducing fecal contamination of food and water, to which these children would be more exposed. Similarly, the effect of TSC latrines is greater in districts with higher population density, where contamination from open defecation is otherwise more likely. Applying this same identification strategy to another individual-level dataset, I find that children who lived their first year of life in district-years with more TSC latrines grew taller than other children born in different years or different districts.

The second set of estimates use a difference-in-differences identification scheme, applied to district-level census and related aggregate infant mortality data. These long-difference, between-district results replicate the within-district estimates. No "effect" of the TSC trend is seen in the decades before the program, nor is there an "effect" of other government programs happening at the same time.

The third strategy exploits a monetary prize offered by the Indian government to village officials for successfully implementing the program. The prize is discontinuously increasing in village size; the discontinuous incentive was devised solely for this program. I expect that political chairmen of villages with populations just above prize discontinuities have greater incentives to implement the program than chairmen of slightly smaller villages. Empirically, I find that districts with more villages just above the prize discontinuities experienced less infant mortality in data from after the program; conversely, districts with more villages just below the discontinuities experienced more infant mortality. Because these prizes can be captured by a small number of informed and socially powerful political actors within

villages, the response to these incentives is credible.

This paper contributes to the literature in four important ways. First, it offers the first causally identified evidence of the effects of on-site rural sanitation on health and human capital accumulation. Previous papers in the economics literature have focused on water supply or have inappropriately compared nearby households with and without latrines, ignoring externalities. This analysis uses existing, large-scale survey data sets, and benefits from the availability of multiple sources of variation in program intensity. Second, this paper analyzes full-scale implementation by the Indian government, rather than a pilot project. Despite severe governance constraints, the TSC prevented infant deaths effectively, on average, compared with other programs measured in the literature. Third, the results highlight a successful mechanism: the findings suggest that *ex post* incentives may motivate government performance in weak states and that incentivized local leaders can use social forces to promote technology adoption. Finally, this paper contributes to the growing literature on lasting consequences for human capital of early-life disease: poor sanitation both increases mortality and decreases height. This combination of short-run and long-run effects is consistent with current knowledge and hypotheses on human capital over the life course.

1.1 Studies of sanitation and health

At least since John Snow's investigation into the London cholera epidemic of 1853, researchers have documented a statistical effect of public infrastructure on diarrheal disease (Freedman, 1991). More recently, Watson (2006) studied heterogeneous timing of public health investments – including sewer connections and septic tanks – at U.S. Indian reservations, and found that a 10 percentage point increase in the fraction of homes receiving improved sanitation reduced American Indian infant mortality by 2.5 percent. Similarly, Cutler and Miller (2005) document a large effect of water filtration and chlorination on mortality in major U.S. cities in the early 20th century.

Despite these important studies in rich countries, this paper is the first of which I am

aware to present well-identified estimates of the effects of public investments in rural household latrines in a poor country.[1] In addition to its geography, this paper differs from related studies in three ways: its direct focus on excreta disposal, rather than water quality or quantity; its use of objective health outcomes as dependent variables; and its organization around open defecation as a public bad with negative externalities.

A recent literature in development economics studies improvements to water supply in poor countries. For example, Devoto et al. (2011) investigate randomized promotion of household water connections in urban Morocco; although household connections importantly reduced time spent collecting water, they had no effect on health, which the authors propose is due to the availability of public taps of comparable quality. In contrast, Meeks (2011) exploits heterogeneity in the rollout of community taps in rural Kyrgyzstan to document a decline in children's intestinal infections, among other effects. Kremer et al. (2011) find that protecting springs in rural Kenya reduces fecal contamination of water and child diarrhea.

However, as Black and Fawcett (2008) observe, there is ideally "no connection" between water supply and excreta disposal – logically as well as physically (8). Pit latrines, if used and constructed properly, can prevent feces from contaminating the water supply without piped water (Franceys et al., 1992, 43). More importantly, oral-fecal disease transmission is common by means other than water. I have observed people in rural India prepare food and touch babies' mouths with hands likely contaminated with child or adult feces (*cf.* Coffey, 2012). Moreover, open defecation is not limited to remote fields. I have observed human feces in and near homes, and have witnessed children defecating in and near homes and playing near open defecation areas. In some parts of rural India, so-called "dry latrines" are common: corners or concrete slabs in or near homes where people defecate on the ground or

[1]Earlier papers have studied rural sanitation (Esrey et al., 1991), but much of this literate predates recent econometric emphasis on "design-driven studies" (Angrist and Pischke, 2010) and may not be persuasive to many economists (Zwane and Kremer, 2007). For example, it is common in this literature to identify effects using the case-control method (*e.g.* Daniels et al., 1990). Barreto et al.'s (2007) recent study of urban sanitation in Salvador, Brazil reports a large decline in diarrhea morbidity, but primarily cites rates in the treated city before and after the program.

floor, in the anticipation of the feces' eventual removal by a low-caste "sweeper" (UNICEF, 2011). Unfortunately, sanitation – meaning safe excreta disposal *per se* – remains an often lacking input towards children's health in rural India.

This paper uses large samples to measure objective, "reduced form" health outcomes as dependent variables: mortality and height (*cf.* Schmidt and Cairncross, 2009). Most comparable research has studied mothers' reports of children's diarrhea.[2] However, reported morbidity can be systematically unreliable: in a survey experiment in India, Das et al. (2012) find that changing the recall period reverses the sign of the apparent health care - economic status gradient. Measuring *diarrhea* with respondent reports may be particularly difficult (Schmidt et al., 2011). For example, Zwane et al. (2011) show that households randomly selected to be surveyed more frequently report less child diarrhea. There is less concern about such bias in reports of child death or in height measured by surveyors. Additionally, Humphrey (2009) suggests that chronic but subclinical "environmental enteropathy" – a disorder caused by fecal contamination which increases the small intestine's permeability to pathogens while reducing nutrient absorption – could cause malnutrition, stunting, and cognitive deficits without necessarily manifesting as diarrhea (*see also* Petri et al., 2008; Mondal et al., 2011). Using mortality and height as dependent variables, rather than reported diarrhea morbidity, more completely captures the effects of open defecation while reducing errors in measurement.

Finally, the appropriate policy response to an infectious disease depends on the externalities it entails (Gersovitz and Hammer, 2004). As Miguel and Kremer (2004) discuss in the case of intestinal worms, estimating the effect of treating a single individual for an infectious disease with externalities may substantially underestimate the effect of treating a group. Similarly, because of externalities in disease transmission, open defecation has health effects beyond the households that practice it. Other evaluations of rural sanitation have focused

[2]The combination of my data and the details of this program do not permit identification of a causal effect of the TSC on reported diarrhea.

on differences between households that do and do not have latrines (Daniels et al., 1990; Esrey et al., 1992; Lee et al., 1997; Cheung, 1999; Kumar and Vollmer, 2012). In contrast, the independent variable in this paper varies at the district level, asking what the effect is of living in an area where a greater fraction of the people have access to household latrines.

1.2 India's Total Sanitation Campaign

The Total Sanitation Campaign is a one of seven "national flagship programs" of the Government of India.[3] The TSC encourages villages to become "open defecation free," largely through the construction and use of household pit latrines. The program was announced in 1999, but began building latrines in 2001, which I take as the beginning of the program, in some cases using 2001 census data for baseline figures.

The design of the TSC incorporated lessons from the poor record of India's earlier Centrally Sponsored Rural Sanitation Programme, launched in 1985 with a focus on construction of relatively expensive latrines (Black and Fawcett, 2008). The TSC emphasized low-cost latrines (approximately 1,500 to 2,000 rupees, or $30 to $40, at market exchange rates), and offered only a partial subsidy (60 to 90 percent of average prices), expecting beneficiaries or villages to contribute towards construction costs, especially in better-off households. Moreover, the TSC was planned around the known difficulties in encouraging poor households to engage in preventative health-promoting behavior (*e.g.* Dupas, 2011). In particular, it incorporated two plausible features that economists have found to usefully promote behavioral change in other programs: it offered an *ex post* incentive for achieving the desired outcome, and it made use of existing village social structure.

[3]For a detailed account of the implementation of the TSC, see the analysis and memoirs of Alok (2010) an Indian Administrative Service officer involved with the TSC from 2001 through 2008.

1.2.1 Why the TSC might have succeeded

In October of 2003, the Indian government announced the Clean Village Prize (NGP in Hindi initials), an incentive for villages (Gram Panchayats) that achieve "open defecation free" status. When a village's chairman decides the village is eligible, it submits an application to the Ministry of Rural Development, whose monitoring division dispatches a monitoring agency to verify that every household in the village is disposing of its feces safely (Alok, 2010, 287). If a village is approved, its chairman receives the prize from a political figure at a prestigious ceremony. The prize includes a monetary incentive.[4] This incentive is large for rural India: \$1,000 to \$10,000 per village at market exchange rates, \$3,400 to \$34,000 PPP, with a mean computed from this paper's data of 74 rupees per capita, or about \$5 PPP. The prize is discontinuously increasing in village size, in a way that will be detailed and exploited as an additional source of variation in program intensity in section 3.3.

An advantage of the NGP, relative to programs that focus on construction, is that it incentivizes the desired outcome, rather than one input (*cf.* Easterly, 2002). As Holmstrom and Milgrom (1991) describe, incentivizing an agent for performance on one input towards a complex goal could worsen the final product; thus one worries about "teaching to the test." Indeed, Glewwe et al. (2003) find that a program that rewarded teachers in Kenya based on students' test scores caused an increase in short-term exam preparation sessions, with little further effect. However, other studies have found that properly conditioned incentives can improve health programs in developing countries. For example, Banerjee et al. (2010) documented that parents in rural Rajasthan who were rewarded with lentils and metal plates were more likely to have their children immunized. Similarly Thornton (2008) found that many more experiment participants in Malawi sought their HIV test results among a group offered a small cash incentive.

Although the NGP incentive money is nominally intended to be spent on village devel-

[4]To be clear, throughout this paper, this monetary prize will be referred to as the "incentive;" sometimes Indian government and other documents describe the partial subsidy described in the paragraph above as the TSC's "incentive," but I follow economists' convention in calling this a "subsidy."

opment, there is much qualitative and quantitative evidence of local elite capture of such government funds in rural India. Thus, the TSC makes use of existing social structure, by giving the village chairman, often a socially powerful figure, an incentive to motivate the rest of the community.[5] As an illustrated government manual insists, "it is the duty of the elected representatives to convince the community members that every family must have a sanitary toilet" (Government of Assam and UNICEF, 2010, 3). Such informal social pressure can be powerful in rural communities, perhaps especially given social hierarchy in India. For example, Hoff et al. (2011) demonstrate that high-caste participants in economic games (unlike low-caste participants) are willing to punish defectors who violate cooperative social norms. Karlan (2007) documents that social connections in Peru enable the monitoring and enforcement required for joint-liability loans. Adoption of new technology in developing countries may be particularly dependent on social learning and influence (Besley and Case, 1993). Indeed, in the very context of India's TSC, Pattanayak et al. (2009) find in a randomized, controlled trial in Orissa that in villages receiving a "shaming" treatment[6] latrine ownership increased from 6% to 32%, but over the same period there was no increase in ownership in control villages.

1.2.2 Administrative records on TSC implementation

As its key independent variable, this paper uses administrative records on the implementation of the TSC collected at the district level by the government of India. In the period under study, India had about 600 districts, although some had no rural population. Individual-level

[5]One village chairman in Uttar Pradesh told us about how, when the village of another chairman he knew won the award, "he" (the chairman) got the prize. Similarly, when I asked about motivational messages painted in a village in Madhya Pradesh, a resident explained that the chairman was attempting to win the prize. Black and Fawcett (2008) interviewed a winning chairman in Andhra Pradesh, who had extended his term in office by making his wife the nominal chairman; he "set his heart on winning" the NGP and – after investing some village funds to further subsidize the poorest families' latrines – eventually collected a large reward in Delhi (119-120). For more qualitative evidence, see the supplementary appendix.

[6]Shaming included "a community walk aimed at drawing attention to poor hygiene and triggering a collective emotional response [and] determining and discussing the volume of faecal matter accumulated in a village" (581).

mortality results will use data on the rural populations of 553 districts with an average of about 300 observed infants each; district-level mortality results will use 280 districts because the 2010-11 Annual Health Survey only covered nine states.

An annual-frequency, district-level time series of TSC household latrine construction, collected for administrative purposes, is publicly available on the program's website at http://tsc.gov.in/. Figure 1 presents one depiction of the differential rollout of the TSC over time, with four state trends shown for illustration. An acceleration is visible after the announcement of the NGP in late 2003. Importantly, these latrine counts are as reported by the Indian government. Although sections 3.2.2 and 3.3 will demonstrate that the key results are robust to alternative measures of program intensity that are independent of the TSC, most of this paper estimates results of latrine coverage as reported by the TSC. This is likely both an overestimate and an underestimate of total program activity. It is an underestimate insofar as the TSC undertook activities beyond household latrine construction, such as social mobilization and in some cases school latrines. It is an overestimate if bureaucrats inflate construction counts (*cf.* Imbert and Papp, 2011). I assume that the true intensity of the TSC is an increasing function of household latrine construction. If so, final estimates of the mean, overall effect of the TSC on infant mortality may be reliable (*see* section 5).

A "process evaluation," rather than an "impact evaluation," would conclude that the TSC is far from successfully implemented throughout India. Indeed, this heterogeneous rollout is part of what permits this paper. Why has the TSC been implemented more intensively in some parts of India than others? There are surely many reasons – including the distribution of a district's villages around discontinuities in the NGP monetary prize, which will be exploited in section 3.3. However, implementers at the village, district, NGO, and federal level have all insisted to me that the key source of heterogeneity is the interest in this program of a district's village chairmen and, to a lesser extent, its District Magistrate, an administrative officer who is sometimes moved to another district without warning. Almost every estimate in this paper will use some form of district or village fixed effect, controlling

for fixed heterogeneity in the politics and other properties of places. Identification of the effect of rural sanitation will be driven by differences in the changes or timing of changes in latrine coverage due to the TSC, perhaps ultimately driven by the sum of the attention of many local government officials. For more detail about the TSC and quantitative and qualitative evidence on its implementation and the sources of heterogeneity in its intensity, please see the supplementary appendix.

1.3 Outline

This paper combines several sources of individual survey and district census data with TSC administrative records to estimate effects of the TSC using complementary identification strategies (see table 1). Table 2 presents summary statistics. In the decade from 2001 to 2010-11, when the TSC reported building almost 0.6 latrines per rural household on average, rural infant mortality fell by about 20 babies per 1,000 and urban IMR fell by about 9. The first identification strategy, presented in section 2, investigates this decline in IMR at the individual level. The third round of the District Level Household Survey contains data on the survival of infants born from 2004 through the survey date in 2007 or 2008, permitting comparison of a district's changes in infant mortality with its time-series of latrine construction. I construct longitudinal data from the cross-sectional birth history, allowing panel data methods.

Section 3 approximately replicates the within-district result with a difference-in-differences estimate at the district level. Census data from 2001 and earlier are used for baseline figures and falsification tests, but 2011 census infant mortality data are not yet available. Therefore, endline IMR data is gathered from the published bulletins of the 2010-11 Annual Household Survey. This survey was conducted by the government of India in nine relatively poor states in north India: Uttar Pradesh, Chhattisgarh, Madhya Pradesh, Uttarakhand, Jharkhand, Orissa, Bihar, Assam, and Rajasthan.[7] Section 3.2.2 presents a further replica-

[7]The Office of the Registrar General's documentation explains: "These nine states, which account for

tion, using measures of TSC intensity independently collected by UNICEF in 47 districts of Uttar Pradesh, India's largest state, and finds similar results. Section 3.3 uses the discontinuity in the NGP incentive to construct a measure of predicted variation in district-level TSC intensity based on village-level 2001 census data, measured before the announcement of the incentive. Relative to other districts cross-sectionally, districts with more villages just above the prize discontinuities suffered less infant mortality in 2010-11. Moreover, districts with many villages just below the prize discontinuities saw more infant mortality, and this opposite effect is approximately equal in absolute value.

Section 4 uses the India Human Development Survey to estimate an effect of early life sanitation on height-for-age, using the same individual-level identification strategy as section 2. Finally, section 5 discusses the magnitude of these estimates in the context of other well-identified estimates of effects on infant mortality and childhood stunting in the literature.

2 Infant mortality: Individual-level estimates

Are infants who are exposed to more on-site rural sanitation in their first year of life more likely to survive that year, compared with other infants born in the same district in different years, or in other districts in the same year? This section estimates that a difference in sanitation coverage equivalent to the India-wide mean implementation is associated with about 4 fewer infant deaths per 1,000. These estimates are produced using individual-level data from a large household survey that is representative of rural India and that permits a wide range of covariate controls.

about 48 percent of the total population in the country, are the high focus states in view of their relatively higher fertility and mortality indicators."

2.1 Empirical strategy

The District Level Household and Facility Survey is collected by the International Institute for Population Sciences in Mumbai.[8] The third round, in which 93 percent of interviews were conducted in 2008 and 7 percent in 2007, surveyed 1,000 to 1,500 households in every district of India, selected from 50 primary sampling units in each district.

The survey asked interviewed women about every pregnancy since January 1st, 2004. From this birth history, I create a longitudinal, repeated cross-section dataset of 198,287 infants born alive. Each infant is matched to the number of TSC latrines per rural person that had been built in her district by her first year of life, summing construction over previous years. Infants' months of birth are used to create a weighted average of the counts in the two calendar years that her first year spans. Thus, if a child were born in April of 2004, in 2004 there were 0.12 TSC latrines per capita in her district, and in 2005 there were 0.16, then her independent variable would be $0.12 \times 0.75 + 0.16 \times 0.25 = 0.13$. This number would, by construction, be the same for every child born in the same district in the same month. Results below are qualitatively robust to simply assigning each infant the count of latrines built by the calendar year of her birth.

The dependent variable is an indicator for death in the first year, multiplied by 1,000: it takes on the a value of 0 or 1,000. This scaling is to facilitate comparison with infant mortality rates published per 1,000 infants. This variable is used in a linear estimation of the probability of surviving to one year of age, how IMR is often explained. However, this longitudinal probability – in essence demographers' ${}_1q_0$ for a cohort of one infant – is slightly different from ordinary IMRs, computed as the count of infant deaths divided by the count of infant births in a year (Preston et al., 2000).

Effects of the TSC are estimated taking advantage of the differential timing of its inten-

[8] Questionnaires are online at http://www.rchiips.org/Questonaire.html and the data is available from IIPS to researchers.

sification across districts. Using district and year fixed effects, I estimate:

$$IMR_{idt} = \beta\, TSC_{dt} + \underbrace{D_i\gamma + H_i\zeta + M_i\mu + P_i\phi}_{\text{individual-level controls: } X_i\theta} + \alpha_t^{years} + \delta_d + \varepsilon_{idt}, \tag{1}$$

where IMR is the scaled survival indicator, TSC is TSC latrines per capita, i indexes infants, d indexes districts, and t indexes time of birth in month. The regression includes district fixed effects δ and year of birth fixed effects α. Results will additionally be shown with state $\times$ year fixed effects in place of year fixed effects and with village fixed effects in place of district fixed effects, as robustness checks. For a model verifying the appropriateness of this fixed effects strategy, please see the supplementary appendix.

Individual-level covariates X_i are added sequentially, in sets. Demographic variables, D_i, are the child's sex and birth order and an indicator for being part of a multiple birth. Household variables, H_i are indicators for housing type, for caste group, for having a nominally poverty-targeted government ration card, and for using clean cooking fuel and further include an asset index factor score included with the DLHS, entered linearly and as quintile indicators. Mother and childcare controls, M_i, are the child's mother's age at birth, the count of females in the household, and indicators that the mother correctly reports that one should increase fluids during diarrhea, that she ever went to school, that she can read a sentence, and that the baby was exclusively breastfed for any initial period. Finally, P_i captures exposure of the baby and mother to any other government programs that might confound this result: indicators for receiving any antenatal care, what the DLHS classifies as "full" antenatal care, and any postnatal care within two weeks of the birth, as well as indicators that the mother reports having heard or seen any message promoting institutional delivery and promoting immunization.

As Bertrand et al. (2004) recommend, standard errors are conservatively clustered by district (rather than, for example, district-years). With more than ten times the required 50 districts, large-sample standard errors are reliable (Cameron et al., 2008).

2.2 Results

Results are presented in Table 3. The estimates for rural infants from every specification are quantitatively similar to those in the column 1, the simplest case: the construction of TSC latrines is associated with a decline in rural infant mortality. Because of the construction of the independent variable, the coefficient is scaled as the effect on IMR of the TSC building one latrine per rural person; this should not be taken literally and was not the program's goal. The table also presents linearly[9] scaled effects: latrine construction equivalent to one pre-program standard deviation is associated with a decline in IMR of about 3 deaths per 1,000 and the effect at the mean level of TSC construction is an elimination of about 4 deaths per 1,000.[10] These effects remain quite similar with the full set of controls, as well as with state-year fixed effects or village fixed effects.[11]

In principle, these results could be spuriously driven by district trends in IMR uncorrelated with the controls but correlated with TSC construction. If so, one might expect a similar correlation with mortality among the district's urban children. However, if these results in fact reflect an effect of the TSC, there would be no urban effect, because urban infants were not exposed to this rural program. Columns 8 and 9 estimate the same models with urban children; the "effect" of the TSC is not statistically distinguishable from zero and has, if anything, a positive sign.

Gender is a salient division in Indian society, and many resources are more readily invested by some households in boys than in girls. However, open defecation is a public bad, and a neighbor's fecal pathogens might be expected to have a similar effect on boys and girls. Empirically, there is no interaction of infant sex and TSC intensity; the estimated interaction of TSC_i with an indicator for being a girl has a t-statistic of 0.55.

[9] If included, a quadratic term for TSC is not statisticaly significant ($t = -0.07$).

[10] If column 1 is estimated using logit, instead of this linear probability model, the coefficient of -2.92 translates into a similar decline in IMR of 4.28 per 1,000.

[11] Although not reported in the table due to possible over-controlling, if district-specific linear time trends are added to equation 1's full specification – that is, $\sum_d [\beta_d year_t \times \delta_d]$ – the result is essentially identical: a coefficient of -87.3 with a clustered standard error of 34.3.

Program intensity is measured per rural capita because what ultimately matters is whether every person's feces is disposed safely. The average household size varies throughout India, and could depend on whether multiple nuclear families (such as adult brothers living in a compound with their father) are counted as a household. Moreover, I have seen households with more than one government-provided latrine. With these caveats, I replicated the regression using latrines per household, rather than per capita. The results are similar: in place of the decline in IMR of 85.7 for one latrine per person, there is a decline of 15.0 for one per household ($s.e. = 6.7$), which is about one-fifth of the earlier estimate, consistent with 5 people per household. Finally, the DLHS includes survey weights, which could bias regression results (Deaton, 1997). Repeating the estimation without the recommended survey weights increases the estimated effect in absolute value very slightly from -85.7 to -86.4.

2.2.1 Parallel trends: Evidence from the 1990s from the DLHS-2

This section's identification strategy depends on the parallel trends assumption that there would be no relevant difference in the trends in infant mortality across districts, in the absence of the TSC. Although this assumption cannot be directly tested, I replicate the estimation of equation 1 using infant mortality data from the second round (rather than the third round) of the DLHS. In place of four years of infant mortality during the program, 2004-2007, I use as the dependent variable four years of infant mortality *before* the program, 1997-2000. The independent variables are kept the same: latrine construction from 2004-2007.

We would not expect an effect of latrines built in the 2000s on infant mortality in the 1990s. Any correlation, therefore, would be evidence against parallel trends – evidence that, in fact, districts were trending apart before the program, in a way correlated with TSC implementation. However, there is no correlation. The estimate analogous to column 1 of table 3 is -8.1, with a t-statistic of -0.14. Thus, there is no evidence that pre-program trends were not parallel. For more details, please see the supplementary appendix.

2.2.2 Mechanisms: Post-neonatal mortality, breastfeeding, & population density

Infant mortality is the sum of neonatal mortality and post-neonatal mortality: death in the first month and in months 2-12, respectively. Neonatal mortality is importantly influenced by health care and technology at the time of birth. Post-neonatal mortality is more related to exposure to factors such as the disease environment. Bozzoli et al. (2009), studying adult heights in Europe and the U.S. among people born from 1950 to 1980, find that declines in post-neonatal mortality are associated with increases in height, but neonatal mortality is not. If the results table 3 are indeed driven by the effect of the TSC on the disease environment, we would expect to see an effect on post-neonatal mortality, not neonatal mortality.

To test this, I constructed two new indicators that sum to IMR_i: an indicator for dying in the first month, and for dying in months 2-12. Like before, both of these indicators are multiplied by 1,000 for comparability with published rates. Because these data originate in mothers' recall, there is likely measurement error, so any separation of these two rates may be even sharper than what I find.

Table 4 presents the results of estimating the same regressions with these new dependent variables, including the same fixed effects and full set of controls. Comparing columns 1 and 3 with columns 7 and 8, it is clear that the effect of the TSC is concentrated on post-neonatal mortality, as expected, with no statistically significant effect on neonatal mortality.

Although there are many ways that rural Indian babies may be exposed to fecal-oral infection, breastfeeding would be partially protective (Brown, 2003). I constructed a variable from 0 to 1 reflecting the fraction of the first year after a baby's birth that remained after she first had food other than breast milk, according to the mother's report. This number would be 0 for a baby exclusively breastfed for this whole year, 1 for a baby given supplemental food on the day of its birth, and linear in between. As table 2 reports, this variable has a mean of 0.72, or introduction of complementary food after about three months.

Columns 4 and 5 include this variable, demeaned (to preserve comparability of the main effect) and interacted with TSC intensity. The protective effect of district-level TSC intensity is greatest for infants who are exposed earlier in life to non-breast milk food. Thus, consistently with a plausible mechanism, the TSC is most helpful for infants who would otherwise have been at the greatest risk.

In a similar verification of a plausible mechanism, TSC intensity would be expected to have a greater effect on health where population density is greater: these are the places where somebody else's open defecation might otherwise be most likely to cause infection. Root (1997) finds that regional variation in population density within Zimbabwe explains variation in child mortality, due to transmission of infectious disease. Similarly, Ali et al. (2002) show that high population is associated with greater cholera risk in a rural area of Bangladesh.

Therefore, in a mechanism check that is not reported in the table, I added an interaction between TSC intensity and district rural population density to the basic specification of equation 1. The coefficient on TSC intensity is essentially unchanged ($\hat{\beta} = -95.7$, $s.e. = 39.5$) and the interaction has the predicted negative sign (-0.019, $s.e. = 0.0073$). This means that a one standard deviation increase in rural population density is associated with about an 8 percent increase in the effect of TSC latrines. This is consistent with TSC latrines protecting infants from exposure to open defecation, which would have been more likely in higher-density populations.

2.2.3 Timing of the effect

If the TSC caused declines in infant mortality, so these results do not merely reflect correlated district time trends, we would expect the program intensity at the time of the baby's first year of life to matter more than at other times. In particular, causes should precede effects (Granger, 1969), and latrines constructed after a baby's first year of life should not impact its survival of that first year.

With this motivation, I re-estimate the regressions from the first columns of tables 3 and 4, using TSC latrine construction from other years as the independent variable. In particular, I use latrine counts from two years after the first year of birth and two years before the first year of birth. Because the TSC construction time series extends beyond the DLHS birth history in both directions, this does not entail a reduction of the sample. These regressions use district fixed effects; this will difference out the true TSC count in the infant's birth year, so with such a short time-series – four years from 2004 to 2007 – one would not expect an "effect" of this mistimed program intensity.

Figure 2 plots coefficient estimates for regressions of infant and post-neonatal mortality rates on TSC intensity two years before and two years after the infants' birth, as well as the estimate of the correctly timed effect from tables 3 and 4. In these regressions with district and year fixed effects, there is no statistically significant "effect" of the lead and lagged values of the program on infant mortality.[12] The pattern is clearest for post-neonatal mortality – where an effect of the TSC is *a priori* most likely to be visible – as the point estimates fall sharply away on either side of the estimate of the correctly timed effect.

3 Infant mortality: District-level estimates

Section 2 found that infants born in years with more TSC latrines were more likely to survive than infants born in other years in the same districts. Comparing across districts, did IMR decline by more during the 2000s, on average, in districts where more latrines per capita were constructed under the TSC? This section first applies a difference-in-differences strategy to long differences in district-level published census rates. Next, these results are approximately replicated for one state with data on TSC intensity from another source. Finally, I present evidence of a causal role of the TSC, using a discontinuity in the mapping from village

[12]The results are similar (only the coefficient on properly timed TSC intensity is statistically significant and the coefficients on the lead and lag are small) if the lead and lag independent variables are included with the properly timed independent variable in the same specification. Please see the supplementary appendix for details.

population to NGP monetary incentive.

3.1 Method

Indian districts are likely heterogeneous in ways that are both correlated with IMR and with TSC intensity. For example, columns 5 and 6 of table 5 report that TSC construction over the 2000s is positively associated with infant mortality in 2000 and 2010, which is consistent with the possibility that districts that already had better sanitation coverage probably also had low IMR and did not need to build as many more latrines during the TSC.

Therefore, my estimates will use district-level difference-in-differences – or, more precisely, collapsed first differences, which produce the same coefficient estimate with more conservative inference (Bertrand et al., 2004). The main specification is

$$\Delta IMR_d = \beta_0 + \beta_1 \Delta TSC\ \ latrines_d + \Delta X_d \theta + W_d \omega + \varepsilon_d, \tag{2}$$

where changes are from 2001 to 2010-11 and d indexes districts. ΔX is a vector of other demographic changes from census data (change in female literacy, overall literacy, population growth, and change in the child sex ratio). W is a vector of cross-sectional data from other sources that could be correlated with changes in IMR over this period. Only 280 districts will be used because 2010-11 IMR data is from the Annual Health Survey, which is collected by the same agency of the Indian government as the census, but is not part of the 2011 census. Note that the total number of TSC latrines constructed is the same as the change in TSC latrines over this period because none were built before 2001.

The controls W include latrine coverage before the program, as computed from census data by Ram and Shekhar (2006), to prevent spurious regression to the mean. As an indicator of district remoteness I include driving distance from the district capital to the state capital in kilometers.[13] From the 2001 census, I include the district's rural population, its count of

[13]The data were collected from Google Maps using the Stata tool presented by Ozimek and Miles (2011).

villages, and their interaction (perhaps a high population has different implications if people are concentrated in a few large villages).

3.2 Results

Estimates of equation 2 are presented in columns 1 through 4 of table 5, adding controls in stages. Taking the scaling of the dependent variable inappropriately literally, building one latrine per person would be associated with a decline in infant mortality of about 40 infants per 1,000, but as before, this extrapolates far outside of both the program's actual achievement and its goals. At the average level of TSC latrine construction in these 280 districts, the TSC caused a decline in infant mortality of about 4 per 1,000; a 2001 cross-district standard deviation in latrine coverage is associated with a decline of between one and one-and-a-half infants per 1,000. Adding the controls decreases the coefficient estimates a little and increases the standard errors a little, but does not importantly change the result.

The differencing is important: as columns 5 and 6 show, TSC intensity is positively associated with cross-sectional IMR. This suggests that the results are not merely reflective of better-off, healthier districts enjoying higher capacity governance and better program implementation. Columns 7 and 8 report opportunities to disprove the "parallel trends" assumption that enables difference-in-differences estimation. Although no test can prove what would have happened in the counterfactual case without the program, one might doubt causal identification if the program were correlated with pre-program trends. In fact, the TSC is uncorrelated with 1991 to 2001 and with 1981 to 1991 district-level changes in IMR; neither t-statistic exceeds one, and one of the two point estimates is positive.

3.2.1 Falsification tests

Table 6 presents three further opportunities to falsify the causal identification of the TSC's effect. Panel A repeats the regressions from table 5 with urban IMR as the dependent variable, in place of rural IMR. Because the TSC is a rural program, it would not be expected

to have much effect on urban children. Indeed, there is no apparent effect, and adding the controls changes the sign of the coefficient estimate.

As a placebo test, panel B estimates the "effect" of the JSY, a government incentive scheme introduced at a similar time and intended to encourage institutional delivery. The independent variable is the percentage of women in the district surveyed in the DLHS-3 who had a pregnancy during the reference period who reported receiving the JSY incentive money. These regressions should certainly not be taken as any type of well-identified evaluation of the JSY. However, naïvely regressing the change in IMR on this measure finds no correlation, suggesting that the result is not simply due to confounding variation in state capacity or government commitment to early life health.

Panel C conducts a similar placebo test using NREGA, a large government workfare program enacted in 2005. While it is quite implausible that NREGA would have had a large effect on infant mortality, its data are reported in a similar online government database as the TSC's (*see* Imbert and Papp (2011) about the NREGA data). Therefore, any district-level variation in record-keeping capacity or corruption might be correlated across NREGA and TSC data. The panel presents three regressions, using as independent variables the fraction of households who have a job card (a document used to get NREGA work), the fraction who have actually received NREGA work, and the fraction of those with job cards who have received NREGA work. All of these variables are taken from the government's online monitoring system. There are no statistically significant "effects" of NREGA on infant mortality.

3.2.2 UNICEF monitoring data

Thus far, all of the measures of TSC program intensity used have been computed from the TSC's own administrative records. This section uses an alternative data source. In March 2010, the Lucknow regional office of UNICEF organized an independent measure of TSC

activity in 47 districts of Uttar Pradesh, the largest and one of the poorest states of India.[14] The data are the result of a survey of an average of 134 households in each district. Five villages (*gram panchayats*) were randomly selected in each district, one from each of five strata by population size; within these, 10 percent of households were sampled from the voter list. Of the 6,443 households surveyed, the survey found that around 40 percent had a toilet or latrine.

Table 7 presents estimates of equation 2 using the UNICEF data as the independent variable. In this data, TSC intensity is measured as latrines per household, rather than per capita. As before, more TSC latrine construction is associated with a greater decline in rural infant mortality – an effect estimated even more precisely when the change controls are included. Columns marked "weight: sample" are weighted by the square root of each district's sample size. There is no apparent "effect" on urban IMR: none of the t statistics are greater than 1, two of the estimates are positive, and the sign changes when the controls are added.

3.3 The clean village prize

As section 1.2.1 described, village chairmen were motivated to fulfill the objectives of the TSC with an *ex post* monetary prize. The NGP incentive is a discontinuously increasing function of village population, as measured before the introduction of the prize. Small variations in population size across these discontinuities among otherwise similar villages could imply large differences in local leaders' motivation to achieve an open defecation free village. Districts with more villages that happen to be above the prize discontinuity would then, on average, see the TSC more intensively implemented.

This section constructs various predictors of district-level TSC intensity, based on village populations in the 2001 census and the NGP rules. This can be seen as a robustness check

[14] I am very grateful for the use of this data, but all estimations and conclusions are my own responsibility and do not necessarily reflect any position or judgment of UNICEF or its employees.

both of the data – using another measure of TSC intensity that is not based on administrative records – and of the inference of causality – using another source of heterogeneity in TSC intensity. This discontinuous incentive was devised solely for this program, and these cut-points are not associated with any other distribution scheme.

With ideal data, one would estimate a traditional regression discontinuity at the village level. However, village-level health outcome data that could be matched (with real village names) to population data are unavailable. Additionally, the TSC's village level administrative data on latrine construction are of low quality, and in some cases have not been updated for several years. The supplementary appendix contains qualitative evidence from field visits and quantitative evidence making some use of administrative data, both suggesting that the NGP was indeed implemented as described here and did motivate village chairmen more where effective incentives were greater.

3.3.1 Empirical strategy

Table 8 reprints the NGP rules, reporting the incentive offered in each population interval. Villages are not in competition: in principle, they could all qualify for the prize. The prize is large enough to motivate a leader who will personally capture much of it, but not nearly large enough to have a direct wealth effect on infant mortality. There are four points where the incentive discontinuously increases. Importantly, these incentives are functions of published census data, collected in 2001. Therefore, relevant village populations were fixed before the NGP was announced in 2003. In other contexts, researchers have worried about the validity of discontinuity-based causal identification if actors have an incentive to non-randomly manipulate their position relative to the discontinuity (*e.g.* Urquiola and Verhoogen, 2009; Barreca et al., 2011). Here, this is not a concern.[15]

Consider the incentives facing a village chairman who is trying to decide whether to

[15]Unsurprisingly, the density of 2001 village population is smooth through the discontinuities: 50.2 percent of the 11,347 villages with populations within 20 people of the nearest discontinuity are above the nearest discontinuity; this is not statistically significantly different from 50 percent ($F = 0.18$, $p = 0.67$).

invest in attempting to win the NGP. This comes at a cost, including any financial costs, such as of latrine construction or painting motivational messages on walls, and an opportunity cost of other ways he could focus his attention and political capital. In a stylized binary representation, he will attempt to win the NGP if

$$incentive(populaton) > \frac{cost(population)}{probability(population)}, \tag{3}$$

where the *incentive* is the NGP prize, the *cost* is the sunk cost of attempting to win, and the *probability* is the probability of winning, given that he pays the costs of attempting. Each of these could be a function of the village population: for example, more people require more latrines. Although there is no statistical data on the village chairman's cost and probability functions, it is reasonable to assume that the right-hand side of this inequality is a continuously increasing function of population. Because the incentive is discontinuously increasing, village chairmen will have the greatest motivation to attempt to win the prize in villages with populations slightly larger than the four population cut-points, and the least motivation in villages with populations just below the cut-points.

This incentive operates at the village level; however, IMR data is reported for districts. Therefore, I construct district-level variables to reflect the intensity of the incentive in each district, using 2001 census data on the population in each village.[16] Because the purpose is not to precisely estimate the effect of each rupee of incentive, what matters for causal identification here is the shape of the the mapping from a district's population distribution to its incentive variable. I use two complementary approaches.

The first approach is to compute the district's incentive per capita if every village in the district were to win the prize; for districts of identical overall population or population per village, this could be very different if people live in villages above or below the cut-points. This variable is positively associated with district TSC construction (a 100 rupee increase in

[16]There are a mean of 1010 villages per district.

a district's average incentive is linearly associated with a 0.12 percentage point increase in latrines built per capita, $t = 3.6$). However, it is not associated with 2001 infant mortality (t less than one and positive, not negative); nor with 2001 toilet coverage ($t = -1.3$). For more detail and regression tables, see the supplementary appendix.

The second approach is to calculate the fraction of the district's population who live in villages just above the cutpoints. For example, I compute the fraction of the population who live in villages with populations less than the population that is 30 percent of the distance from the greatest lower threshold to the least greater threshold. Thus, a village of population size 1,200 would count because it is between the 1,000 and 2,000 discontinuities and is within 30 percent of the 1,000 cut-point (1,300), but a village of size 3,000 would not count because it is between 2,000 and 4,000 and greater than 2,900, the population 30 percent of the distance between these two.

This method has two advantages. First, multiple bandwidths can be used as robustness checks. Second, this approach suggests a straightforward test of causal identification. Districts with many villages above the discontinuities would be expected to exhibit more TSC intensity, and districts with many villages below the discontinuities would exhibit less. If the association between these measures and health outcomes is because of the mechanisms I propose, then "flipping" the measure – counting the fraction of people living the same distance *below* each discontinuity – would produce an estimate similar in absolute magnitude but opposite in sign.[17] In other words, this method allows us to test whether districts with many people living in villages just above the discontinuity ultimately experienced better infant health, and whether districts with many people just below the discontinuity experienced worse infant health.

[17]Fujiwara (2011) implements a similar test, finding that the effect on Brazilian states of having many towns with populations above a discontinuity changes in sign across electoral cycles when the discontinuity had different implications.

3.3.2 Estimation

Table 9 presents estimates of district-level cross section regressions:

$$IMR_d^{2010-11} = \beta incentive_d + X_d\theta + \varepsilon_d, \tag{4}$$

where *incentive* is one of five operationalizations of the district aggregate incentive measure. Notice that the incentive is a cross-sectional variable computed from the 2001 census; because it is a district fixed-effect, it cannot be used in within-district regressions such as those in section 3.1. Also, only endline IMR is used as the dependent variable, instead of the difference since 2001. This is to avoid mechanical division bias because 2001 population is in the denominator of the average incentive measures. However, results are very similar (typically greater in absolute value) if this first-difference is used instead.

Each of the 20 estimates presented is from a separate regression of IMR on the incentive measure for that column, including the controls for that row: baseline IMR, baseline toilet coverage, and 2001 population per village. Including controls has small effects on the estimates, increasing them for some incentive measures, decreasing them for others. In every case the aggregated incentives are significantly associated with IMR with the predicted sign.[18] Although not reported in the table, there is no similar association with endline *urban* infant mortality: regressing urban IMR on the same incentive per capita and controls produces a t-statistic of -1.2.

Compare the estimates moving across columns 2 through 4, as the bandwidth narrows on the discontinuity, and then from column 4 to column 5, as the incentive measure flips from an indicator of high average incentives to an indicator of low average incentives. When the

[18] As a falsification test, I replicate the first regression in column 1 with an incorrect computation of incentive per capita computed with each population discontinuity displaced by 500 people. This false "incentive per capita" has a correlation of 0.956 with the true incentive per capita. However, it is uncorrelated with endline infant mortality, with a coefficient of -0.021 and a standard error of 0.087 (compare with the first row of the table); with all three controls (compare with the bottom row) the coefficient is -0.214 with a standard error of 0.194.

implication of the measure switches from high to low incentives, the sign of the coefficient changes from a negative association with infant mortality to a positive association. Moreover, once any controls are included, the point estimates in columns 4 and 5 are very similar in absolute value. The reported p-values verify that the estimates in columns 4 and 5 are different, which would not be the case under the null hypothesis of no effect. Therefore – using measures computed independently of the TSC's administrative records – in districts where the NGP discontinuity would predict high TSC intensity, IMR was lower after the program than in other districts; however, in districts where the NGP predicts low TSC intensity, IMR was greater than in other districts.

3.3.3 Open defecation in the 2011 census: The NGP as an instrument

The 2011 Indian census, independently from TSC administrative data collection, estimated the fraction of rural and urban households openly defecating. These data were collected at the "endline" period, relative to this paper's analysis of the TSC, and at about the same time as the Annual Health Survey infant mortality data used as the dependent variable of this section. The availability of these data suggests a further robustness check: estimating the effect of census open defecation on infant mortality at the district level, using the computed NGP prize per capita based on the discontinuity as an instrumental variable. The results in section 3.3.2 could then be considered the reduced form of this analysis which, again, would make no use of the TSC administrative records.

The full results are presented in detail in the supplementary appendix. First, the TSC administrative records on latrine construction are correlated with open defecation, but unsurprisingly these do not vary one-for-one, given measurement error. Second, a larger district average NGP prize per capita indeed predicts less open defecation ($t = 3.03$ in this first-stage regression). Finally, using the NGP prize per capita as an instrument, moving from 0% to 100% open defecation in the 2011 census is associated with an increase in rural infant mortality of 89 infant deaths per 1,000 ($s.e. = 41$). This figure approximately replicates the

fixed effects result of table 3, which reports a similar estimate of 85 using the DLHS-3 and TSC administrative records. There is no association if *urban* infant mortality is instead used as the dependent variable.

4 Height and stunting: Effects on human capital

This paper has presented evidence that the TSC reduced infant mortality. The early life disease environment also has an effect on children who survive (Checkley et al., 2008). A well-developed literature has documented an association between height and cognitive achievement in rich countries (Case and Paxson, 2008); this gradient is even steeper for Indian children (Spears, 2012). Do children who live in districts that had more TSC latrines during their first year of life grow taller than other children born in the same district in other years, or other districts in the same year? This section answers this question using height-for-age data from the IHDS, matched with the same district-level time series of latrine construction used in section 2.

4.1 Empirical strategy

The IHDS is a cross-section of about 40,000 households, representative of India at the country level. It was mainly conducted in 2005. For children under 5 years old at the time of the survey, it reports the month and year of each child's birth as well as measured height. Like with the DLHS, I use this cross-section to construct a repeated cross-section of children born in different years. The identification strategy is to compare heights among children 0-59 months old who were born at different times within a district and in different districts at the same time to see if those who live in places that had accumulated more TSC latrines by the first year of their lives were taller at the time of the survey.

The IHDS is used to estimate the regression

$$z_{idt} = \beta TSC_{dt} + (A_i \times sex_i)\,\Gamma + H_i\theta + M_i\vartheta + S_i\psi + \alpha_i^{years} + \delta_d + \varepsilon_{idt}, \tag{5}$$

where i indexes children, d indexes districts, t indexes time of birth in months, and z is the child's height-for-age z-score, computed using the WHO 2006 reference population. Most regressions use district and year fixed effects, but state $\times$ year fixed effects and village fixed effects will also be substituted in as robustness checks.

Almost all regressions include $A_i \times sex_i$, a set of 120 sex-specific age-in-month indicators. That is, these include an indicator for being an 8 month old boy, a 9 month old girl, and so on. This is the same level of disaggregation used to make height-for-age z-scores. Therefore, although there is evidence that well-off Indian children grow to international standards (Bhandari et al., 2002), this ensures the results are not driven by any difference between Indian children and the reference population, or by overall early-life Indian growth faltering. Additionally, it controls for differences across district-years in age structure.

The complete specification will also include household H and mother M control vectors. H includes consumption per capita as a quadratic polynomial, a set of eight caste and religion indicators, and household size. M contains the mother's age at marriage, the highest level of education by any female in the household, and an indicator for whether the mother correctly answered that children need more fluids during diarrhea.

Finally, information on older siblings S is included as another attempt to control for household-level heterogeneity. The IHDS also recorded height for children aged 8 to 11. All of these children were born several years before the TSC. About 25 percent of children in this sample have an older sibling aged 8 to 11 with height data. While this subsample is too small to estimate an effect of the TSC with any precision, some specifications will control for older sibling height, with missing values set to 0 and an indicator included for not having data on any older sibling's height.

4.2 Results

Table 10 reports estimation results.[19] Adding household controls, using village or state-year fixed effects, and controlling for older siblings' heights all change the coefficient estimate very little relative to its standard error.[20] Within this sample, 1,311 households have two children under 5 years old; estimating with household fixed effects (not reported in the table) changes the coefficient in column 1 only from 8.6 to 8.4, but on this small sample the standard error increases to 9.5. At its mean intensity, the TSC is associated with an increase in height of one-fifth of a reference standard deviation.[21] This is approximately the same magnitude as the cross-sectional difference in height associated with a doubling of consumption per capita, as computed by linearly regressing height on log consumption in this same sample, with district and year fixed effects. For comparison, from the first DHS-NFHS survey in India in 1992-3 to the most recent in 2005-6, the mean height for age z score for children under 3 increased by 0.4 from -1.9 to -1.5.

The last two columns change the sample. In column 8, the sample is restricted to children at least 2 years old. This is an important robustness check because z-scores measured at younger ages, when growth is fastest, are likely to have more measurement error. Moreover, Indian children's heights stop falling relative to the international reference population between 18 and 24 months, leveling out at about -2. While I have no reason to suspect that this early-life trend could explain these results, restricting the sample to ages where z-scores are

[19] The IHDS has sampling weights with a standard deviation approximately equal to their mean. Repeating the regressions without the weights decreases the coefficient estimate in specification 1 from 8.6 to 7.4 and increases the t-statistic from 2.25 to 2.29. Omitting weights in column 5, with H and M controls, decreases the coefficient estimate from 6.3 to 6.1 and increases the t-statistic from 1.72 to 1.96.

[20] Specifications with some controls but not the full set of controls are statistically significant at the 0.05 level with one-sided p-values but not two-sided p-values. A one-sided test is the most powerful test of this one-sided alternative hypothesis. As a robustness check I implemented the wild cluster bootstrap of the t statistic recommended by Cameron et al. (2008) for column 5, with H and M controls and clusters at the district level. The null hypothesis of no effect is rejected with a one-sided p-value of 0.04 and a two-sided p-value of 0.08.

[21] With the caveats stated in section 2, if the TSC independent variable is scaled as latrines per rural household, instead of per capita, the coefficient in column 1 is 2.17 (se 0.75) and in column 5 is 1.68 (se 0.72). At the mean *household* intensity the TSC is associated with an increase in height of 0.21 standard deviations.

not associated with age rules out the possibility. The coefficient estimate is slightly larger, but is not statistically significantly different from earlier estimates.

As a placebo test, column 9 changes the sample from rural to urban children, who were not exposed to the TSC. A similar coefficient would suggest that the results are driven by spurious district trends. However, the "effect" is very small for urban children, and even slightly negative.

As a final note, the possibility of mortality selection suggests that table 10 may present *underestimates* of the effect of the TSC on survivors' heights. Section 2 found that the TSC importantly increased infant survival. If the marginal infants are shorter than average, then their survival will bias the estimate of the effect on height downwards.

5 The size of the effect of the TSC

How large are the estimated effects of the TSC? If one-quarter of 2001 rural infant mortality of about 83 per 1,000 was due to fecal pathogens (Black et al., 2003), and if the TSC accounted for a decline of 4 as approximately estimated in tables 2 and 3, then the TSC has eliminated about one-fifth of baseline rural infant deaths due to fecal contamination.[22] There is a growing set of well-identified estimates of impacts on infant mortality. Like this paper, most of these use observational data, rather than experimentation, because detecting effects on mortality often requires a large sample.

Unlike many other studies and the canonical program evaluation case, the TSC was not a binary treatment. Both identification strategies, however, suggested that at its country-wide average intensity the program reduced rural IMR by about 4. Although the supplementary appendix details the comparisons in more detail, this effect is of a plausible magnitude, given the literature. It is somewhat smaller than the effect estimated by Galiani et al. (2005) of

[22]This is comparable to a similar rough computation by Kremer et al. (2011) of the effect on diarrhea mortality of spring protection in Kenya (p. 188, footnote 20), although they did not measure mortality directly, and estimate that spring protection reduces diarrhea morbidity by almost one-fourth.

privatization of municipal water supply in Argentina, where 90 percent of households studied had a toilet and 72 percent of municipalities had sewer connections; they find a decline in child mortality of 5.4, or about 8 percent. The effect of the TSC on IMR is about one-third of the convergence in the black-white gap in infant mortality in the rural U.S. south from 1965 to 1975 (Almond et al., forthcoming). The mean effect of the TSC on height-for-age of around one-fifth of a standard deviation is comparable to the effect documented by Barham (2012) of the Matlab Maternal and Child Health and Family Planning Program in Bangladesh, about 0.22 standard deviations.[23] The TSC's effect is smaller than the effect of an early-life nutritional supplement in the famous Guatemala INCAP experiment: 0.36 standard deviations for boys and 0.68 for girls at 36 months (computed from Martorell et al., 1979).

Although corruption is a common problem in low capacity governments, many academic impact evaluations study programs implemented by high capacity NGOs or motivated governments, potentially biasing estimates of effectiveness and complicating policy implications (Duflo et al., 2007; Coffey, 2011). As Ravallion (2012) explains, "a small program run by the committed staff of a good NGO may well work very differently to an ostensibly similar program applied at scale by a government or other NGO for which staff have different preferences and face new and different incentives" (110). Projects in developing countries often suffer from "missing expenditures": discrepancies between official project records and the actual resources used (Olken, 2007). Unlike some estimates of program impact, this paper's are inclusive of all heterogeneity of administration and losses to corruption, under actual implementation at scale.

Latrine construction figures are almost certainly inflated, due to bureaucrats' incentives

[23]The effect of the TSC is also similar to associations between developing country sanitation and children's health found in less cleanly identified studies. Lavy et al. (1996) find that community level "poor sanitation and water" is associated with a 0.18 standard deviation decrement in height for age z-scores. Checkley et al. (2004) report that children in peri-urban Peru "in households without adequate sewage disposal and with small [water] storage containers" are about 0.6 standard deviations (my normalization) shorter at 24 months than "children in households with sewage and with large storage containers."

to appear successful. This is unlikely to explain the finding of a positive effect of the TSC: corrupt reporting is unlikely to be *positively* correlated with the sort of governance or institutions that increase child health, and this would not explain the results with UNICEF data or the computed NGP incentives. Additionally, any inflated recording does not influence the average effectiveness calculations, if the goal narrowly but reasonably is to estimate the mean effectiveness of India's TSC: the coefficient is the average effect of reported TSC latrines (which is *exactly* what the government is able to create), and it is multiplied by the mean number of reported TSC latrines.

However, one may be interested in the mean effectiveness of *real* latrines, perhaps as built by a *hypothetical* program that does not inflate its numbers. Proofs in the supplementary appendix document that, if latrine counts are proportionately inflated, estimates of the overall impact of the TSC on IMR and height are valid for this purpose. If, instead, there are year-to-year changes in latrine count inflation that are not positively associated with year-to-year improvements in infant health, then estimates of the average effect of the TSC are biased downwards.

6 Conclusion

This paper has used several data sources and several empirical strategies to demonstrate a positive effect of on-site rural sanitation, as implemented by India's Total Sanitation Campaign, on children's health. Although latrines were not randomly distributed, an effect is seen on rural children using within-district and across-district variation; no "effect" is seen on urban children, before the program, or from other, placebo programs; the timing of the association between mortality and program intensity is consistent with a causal effect; and predictions based on discontinuities in a prize to local leaders are verified: districts where the average incentive to implement well is greater saw less infant mortality after the program but not before, districts where the average incentive is lower experienced more mortality.

Beyond the TSC, this paper shows that rural sanitation has an important contribution to make to children's health and that open defecation is a health hazard and public bad even in rural areas. The results suggest that policy and research should look beyond improving rural water supply alone, as important as this may sometimes be. These findings underscore the importance of chronic enteric infection to global poor children's health, and demonstrate the value of studying effects of fecal pathogens through reduced form health outcomes, rather than reported diarrhea alone. Finally, these results suggest that *ex post* incentives for achieving verifiable outcomes may be an effective way to motivate government agents in low capacity bureaucracies and that social influence can promote technology transfer.

This paper's results imply that as late as a decade ago, India had much room to improve its sanitation infrastructure, and may still. In India's 2005-06 DHS survey, 55.3 percent of households reported having no toilet or latrine.[24] This is slightly better than DHS reports for Namibia in 1992 (57.9), and much worse than Zimbabwe in 1994 (34.7), Zambia in 1992 (31.7), Uganda in 1995 (19.4), and Sierra Leone in 2008 (23.1). Deaton (2007) observes that it is puzzling that people in India are so much shorter, on average, than people in Africa, despite being richer, on average. Surely many factors contribute to this phenomenon.[25] Yet, according to a linear extrapolation of the estimates in section 4, bringing India to Sierra Leone's level of sanitation coverage would increase average height-for-age by roughly half a standard deviation, or about a quarter of the distance to the WHO international reference population. The effects of the TSC suggest that exceptionally widespread unsafe disposal of excreta in India could be part of the explanation for the puzzle of Indian height.

[24] An early release of figures computed from the 2011 census offers a similar figure (53.1 percent), implying that more Indian households have a cell phone than a toilet or latrine.

[25] For example, sanitation would not readily explain slower cohort-to-cohort growth in women's height than men's (Deaton, 2008).

References

Ali, Mohammad, Michael Emch, J.P. Donnay, Mohammad Yunus, and R.B. Sack. 2002. "The spatial epidemiology of cholera in an endemic area of Bangladesh." *Social Science and Medicine*, 55: 1015–1024.

Almond, Douglas and Janet Currie. 2011. "Chapter 15 Human Capital Development before Age Five." *Handbook of Labor Economics*, 4(B): 1315–1486.

Almond, Douglas, Kenneth Y. Chay, and Michael Greenstone. forthcoming. "Civil Rights, the War on Poverty, and Black-White Convergence in Infant Mortality in the Rural South and Mississippi." *American Economic Review*.

Alok, Kumar. 2010. *Squatting with Dignity: Lessons from India*, New Delhi: Sage.

Angrist, Joshua D and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3 – 30.

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. 2010. "Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives." *British Medical Journal*, 340: c2220.

Barham, Tania. 2012. "Enhancing Cognitive Functioning: Medium-Term Effects of a Health and Family Planning Program in Matlab." *American Economic Journal: Applied Economics*, 4(1): 245–73.

Barreca, Alan I., Melanie Guldi, Jason M. Lindo, and Glen R. Waddell. 2011. "Saving Babies? Revisiting the effect of very low birth weight classification." *Quarterly Journal of Economics*.

Barreto, Mauricio L, Bernd Genser, Agostino Strina, Maria Gloria Teixeira, Ana Marlucia O Assis, Rita F Rego, Carlos A Teles, Matildes S Prado, Darci N Santos Sheila M A Matos, Lenaldo A dos Santos, and Sandy Cairncross. 2007.

"Effect of city-wide sanitation programme on reduction in rate of childhood diarrhoea in northeast Brazil: assessment by two cohort studies." *The Lancet*, 370: 1622 – 1628.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics*: 249–275.

Besley, Timothy and Anne Case. 1993. "Modeling Techonolgy Adoption in Developing Countries." *American Economic Review: Papers and Proceedings*, 83(2): 396–402.

Bhandari, Nita, Rajiv Bahl, Sunita Taneja, Mercedes de Onis, and Maharaj K. Bhan. 2002. "Growth performance of affluent Indian children is similar to that in developed countries." *Bulletin of the World Health Organization*, 80(3): 189–195.

Black, Maggie and Ben Fawcett. 2008. *The Last Taboo: Opening the Door on the Global Sanitation Crisis*, London: Earthscan.

Black, Robert E, Saul S Morris, and Jennifer Bryce. 2003. "Where and why are 10 million children dying every year?" *The Lancet*, 361(9376): 2226 – 2234.

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-Domeque. 2009. "Adult Height and Childhood Disease." *Demography*, 46(4): 647 669.

Brown, Kenneth H.. 2003. "Diarrhea and Malnutrition." *Joural of Nutrition*: 328S–332S.

Cameron, A. Colin, Jonah. B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics*, 90(3): 414 – 427.

Case, Anne and Christina Paxson. 2008. "Stature and Status: Height, Ability, and Labor Market Outcomes." *Journal of Political Economy*, 116(3).

Checkley, William, Gillian Buckley, Robert H Gilman, Ana MO Assis, Richard L Guerrant, Saul S Morris, Kåre Mølbak, Palle Valentiner-Branth, Claudio F Lanata, Robert E Black, and The Childhood Malnutrition and Infection Network. 2008. "Multi-country analysis of the effects of diarrhoea on childhood stunting." *International Journal of Epidemiology*, 37: 816–830.

Checkley, William, Robert H Gilman, Robert E Black, Leonardo D Epstein, Lilia Cabrera, Charles R Sterling, and Lawrence H Moulton. 2004. "Effect of water and sanitation on childhood health in a poor Peruvian peri-urban community." *The Lancet*, 363(9403): 112–118.

Cheung, Y B. 1999. "The impact of water supplies and sanitation on growth in Chinese children." *Journal of the Royal Society for the Promotion of Health*, 119(2): 89–91.

Coffey, Diane. 2011. "Limits to Experimental Evaluation and Student Attendance in Rural Indian Schools: A Field Experiment." paper presented at annual meeting session 89, Population Association of America, Washington.

——— 2012. "Field notes from Sitapur, U.P.." excerpts available in blog posts, www.riceinstitute.org.

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

Daniels, D.L., S.N. Cousens, L.N. Makoae, and R.G. Feachem. 1990. "A case-control study of the impact of improved sanitation on diarrhoea morbidity in Lesotho." *Bulletin of the World Health Organization*, 68(4): 455–463.

Das, Jishnu, Jeffrey Hammer, and Carolina Sánchez-Paramo. 2012. "The impact of recall periods on reported morbidity and health seeking behavior." *Journal of Development Economics*, 98(1): 76–88.

Deaton, Angus S.. 1997. *Analysis of Household Surveys*: World Bank.

Deaton, Angus. 2007. "Height, health, and development." *PNAS*, 104(33): 13232–13237.

——— 2008. "Height, health, and inequality: the distribution of adult heights in India." *American Economic Review: Papers & Proceedings*, 98(2): 468–474.

Devoto, Florencia, Esther Duflo, Pascaline Dupas, William Pariente, and Vincent Pons. 2011. "Happiness on tap: Piped water adoption in urban Morocco." working paper 16933, NBER.

Dinkelman, Taryn. 2011. "The effects of rural electrification on employment: New evidence

from South Africa." *American Economic Review*, 101(7): 3078–3108.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Chapter 61: Using Randomization in Development Economics Research: A Toolkit." *Handbook of Development Econommics*, 4: 3895 – 3962.

Dupas, Pascaline. 2011. "Health Behavior in Developing Countries." *Annual Review of Economics*, 3: 425–449.

Easterly, William. 2002. "The cartel of good intentions: The problem of bureaucracy in foreign aid." *The Journal of Policy Reform*, 5(4): 223–250.

Esrey, Steven A., Jean-Pierre Habicht, and George Casella. 1992. "The Complementary Effect of Latrines and Increased Water Usage on the Growth of Infants in Rural Lesotho." *American Journal of Epidemiology*, 135(6): 659–666.

Esrey, S. A., J. B. Potash, L. Roberts, and C. Shiff. 1991. "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, schistosomiasis, and trachoma." *Bulletin of the World Health Organization*, 69(5): 609621.

Franceys, R., J. Pickford, and R. Reed. 1992. *A guide to the development of on-site sanitation*, Geneva: World Health Organization.

Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology*, 21: 291 – 313.

Fujiwara, Thomas. 2011. "Voting Technology, Political Responsiveness, and Infant Health: Evidence from Brazil." working paper, Princeton University.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsky. 2005. "Water for Life: The Impact of the Privatization of Water Services on Child Mortality." *Journal of Political Economy*, 113(1): 83–120.

Gersovitz, Mark and Jeffrey S. Hammer. 2004. "The Economical Control of Infectious Diseases." *The Economic Journal*, 114(January): 1–27.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2003. "Teacher Incentives." working paper 9671, NBER.

Government of Assam and UNICEF. 2010. *Community Movement for Sanitation and Dignity*: Ministry of Rural Development, Government of India.

Granger, Clive W. J.. 1969. "Investigating Causal Relation by Econometric and Cross-Sectional Methods." *Econometrica*, 37: 424–438.

Hoff, Karla, Mayuresh Kshetramade, and Ernst Fehr. 2011. "Caste and Punishment: the Legacy of Caste Culture in Norm Enforcement." *Economic Journal*, 121(556): F449–F475.

Holmstrom, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization*, 7(special issue): 24–52.

Humphrey, Jean H. 2009. "Child undernutrition, tropical enteropathy, toilets, and handwashing." *The Lancet*, 374: 1032 – 35.

Imbert, Clément and John Papp. 2011. "Estimating Leakages in India's Employment Guarantee." in Reetika Khera ed. *The Battle for Employment Guarantee*: Oxford.

Joint Monitoring Programme for Water Supply and Sanitation. 2012. *Progress on Drinking Water and Sanitation: 2012 Update*: WHO and UNICEF.

Karlan, Dean S.. 2007. "Social Connections and Group Banking." *Economic Journal*, 117(February): F52F84.

Kremer, Michael, Jessica Leino, Edward Miguel, and Alix Peterson Zwane. 2011. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions." *Quarterly Journal of Economics*, 126(1): 145–205.

Kumar, Santosh and Sebastian Vollmer. 2012. "Does access to improved sanitation reduce childhood diarrhea in rural India." *Health Economics*.

Lavy, Victor, John Strauss, Duncan Thomas, and Philippe de Vreyer. 1996. "Quality of health care, survival and health outcomes in Ghana." *Journal of Health Economics*, 15(3): 333–357.

Lee, Lung-fei, Mark R Rosenzweig, and Mark M Pitt. 1997. "The effects of improved

nutrition, sanitation, and water quality on child health in high-mortality populations." *Journal of Econometrics*, 77: 209–235.

Martorell, R., C. Yarbrough, R. E. Klein RE, and A. Lechtig. 1979. "Malnutrition, body size, and skeletal maturation: interrelationships and implications for catch-up growth." *Human Biology*, 51(3): 371–89.

Meeks, Robyn C.. 2011. "Water Works: The Economic Impact of Water Infrastructure." job market paper, Harvard University.

Miguel, Edward and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica*, 72(1): 159–217.

Mondal, Dinesh, Juliana Minak, Masud Alam, Yue Liu, Jing Dai, Poonum Korpe, Lei Liu, Rashidul Haque, and William A. Petri, Jr. 2011. "Contribution of Enteric Infection, Altered Intestinal Barrier Function, and Maternal Malnutrition to Infant Malnutrition in Bangladesh." *Clinical Infectious Diseases*.

Olken, Benjamin A.. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy*, 115(2): 200 – 249.

Ozimek, Adam and Daniel Miles. 2011. "Stata utilities for geocoding and generating travel time and travel distance information." *Stata Journal*, 11(1): 106–119.

Pattanayak, Subhrendu K, Jui-Chen Yang, Katherine L Dickinson, Christine Poulos, Sumeet R Patil, Ranjan K Mallick, Jonathan L Blitstein, and Purujit Praharaj. 2009. "Shame or subsidy revisited: social mobilization for sanitation in Orissa, India." *Bulletin of the World Health Organization*, 87: 580 – 587.

Petri, William A., Jr, Mark Miller, Henry J. Binder, Myron M. Levine, Rebecca Dillingham, and Richard L. Guerrant. 2008. "Enteric infections, diarrhea, and their impact on function and development." *Journal of Clinical Investigation*, 118(4): 1266–1290.

Preston, Samuel, Patrick Heuveline, and Michel Guillot. 2000. *Demography: Measuring and Modeling Population Processes*: Wiley-Blackwell.

Ram, F. and Chander Shekhar. 2006. *Ranking and Mapping of Districts Based on Socio-economic and Demographic Indicators*, Mumbai: International Institute of Poulation Sciences.

Ravallion, Martin. 2012. "Fighting Poverty One Experiment at a Time: A Review of Abhijit Banerjee and Esther Duflo's *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*." *Journal of Economic Literature*, 50(1): 103–114.

Root, Graham. 1997. "Population Density and Spatial Differentials in Child Mortality in Zimbabwe." *Social Science and Medicine*, 44(3): 413–421.

Schmidt, Wolf-Peter and Sandy Cairncross. 2009. "Household Water Treatment in Poor Populations: Is There Enough Evidence for Scaling up Now?" *Environmental Science and Technology*, 43(4): 986 – 992.

Schmidt, Wolf-Peter, Benjamin F Arnold, Sophie Boisson, Bernd Genser, Stephen P Luby, Mauricio L Barreto, Thomas Clasen, and Sandy Cairncross. 2011. "Epidemiological methods in diarrhoea studies – an update." *International Journal of Epidemiology*, 40(6): 1678–1692.

Spears, Dean. 2012. "Height and Cognitive Achievement among Indian Children." *Economics and Human Biology*, 10: 210–219.

Thornton, Rebecca L.. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review*, 98(5): 1829–63.

UNICEF. 2011. "Converting Dry Latrines in the District of Budaun, Uttar Pradesh: A Story of Commitment, Determination, and Dignity." report, UNICEF Office for Uttar Pradesh.

Urquiola, Miguel and Eric Verhoogen. 2009. "Class-Size Caps, Sorting, and the Regression-Discontinuity Design." *American Economic Review*, 99(1): 179–215.

Watson, Tara. 2006. "Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations." *Journal of Public Economics*, 90(8-9): 1537 – 1560.

Zwane, Alix Peterson and Michael Kremer. 2007. "What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review." *World Bank Research Observer*, 22(1): 1–24.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S. Karlan, Richard Hornbeck, Xavier Giné, Esther Duflo, Florencia Devoto, Bruno Crepon, and Abhijit Banerjee. 2011. "Being surveyed can change later behavior and related parameter estimates." *PNAS*, 108(5): 1821–1826.

Figure 1: Construction of TSC latrines, 2000-2011

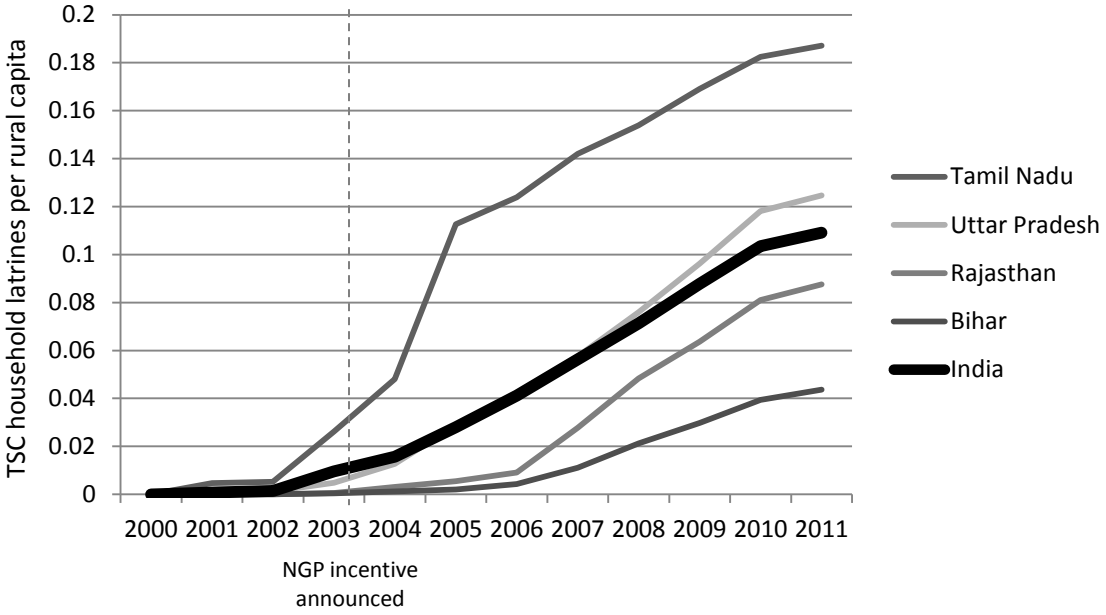
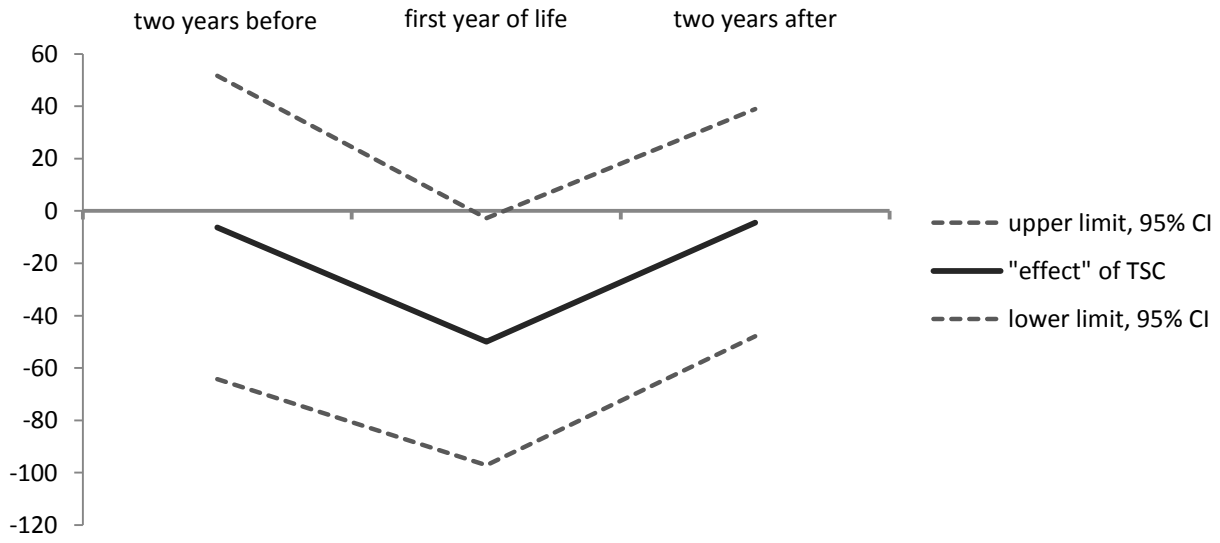


Figure 2: No efffects of time-displaced TSC latrine construction on mortality

(a) post-neonatal mortality rate



(b) infant mortality rate

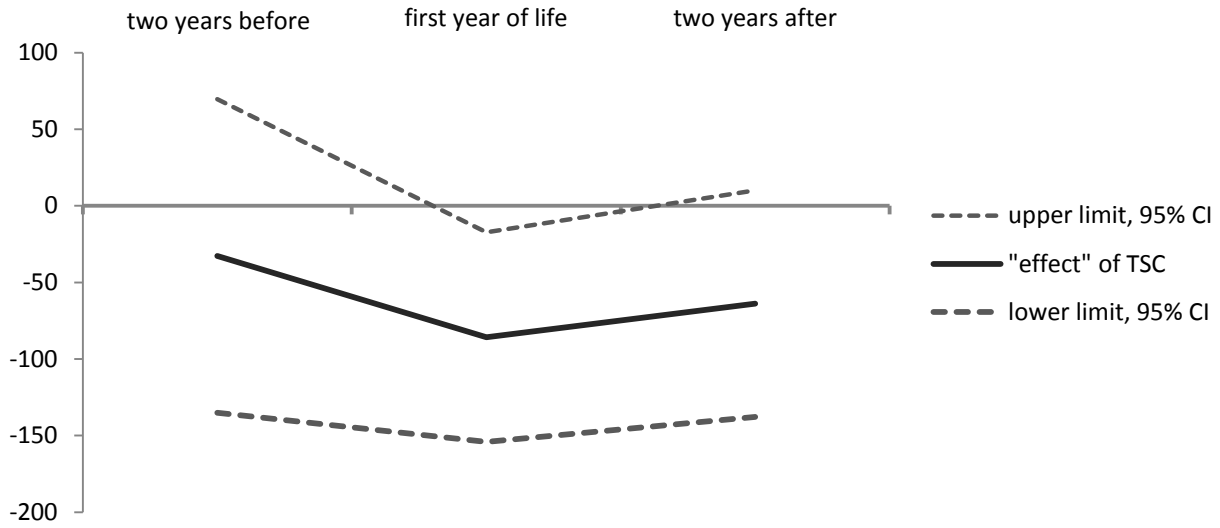


Table 1: Data sources

data source	when collected	observations	contribution
Total Sanitation Campaign (tsc.gov.in)	2001-present	district latrine construction	program intensity
District Level Household Survey (DLHS-3)	2007-08	198,287 infants	infant mortality
District Level Household Survey (DLHS-2)	2002-04	216,159 infants	pre-program parallel trends
India Human Development Survey (IHDS)	2004-05	10,409 children under 5	height for age
Annual Health Survey	2010-11	280 districts in 8 states	infant mortality rates
2001 Indian census	2001	280 districts (others not used)	infant mortality rates
			baseline district covariates
		635,141 villages	population for prize
2011 Indian census	2011	280 districts (others not used)	endline district covariates
			endline open defecation
UNICEF monitoring data	2010	47 districts in Uttar Pradesh	TSC intensity

Table 2: Summary statistics

	mean	standard error
District-level data, 280 districts		
rural IMR, 2010-11	63.22	0.85
rural IMR, 2001	83.42	0.92
urban IMR, 2010-11	50.73	1.19
urban IMR, 2001	61.98	0.68
TSC latrines per rural person, by 2010	0.101	0.0027
TSC latrines per rural household, by 2010	0.585	0.016
baseline household latrine coverage, 2001 (combined rural and urban)	0.263	0.011
NGP prize (rupees per capita)	73.59	0.49
Infants-level data, births 2004-2007		
TSC latrines per rural person, by 2004	0.014	0.00012
TSC latrines per rural person, by 2005	0.024	0.00013
TSC latrines per rural person, by 2006	0.037	0.00015
TSC latrines per rural person, by 2007	0.044	0.00029
female	0.477	0.0012
fraction of first year after non-breastmilk	0.720	0.00065
Child-level data, children under 5 in 2004-05		
TSC latrines per rural person, by 2004-5	0.023	0.00152
height-for-age z score, rural mean	-2.44	0.033
height-for-age z score, urban mean	-1.90	0.043
female	0.482	0.0060
consumption per capita, rural mean (Rs/month)	621	5.09

Table 3: Effects on TSC on individual-level infant mortality, DLHS-3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sample:	rural	rural	rural	rural	rural	rural	rural	urban	urban
	Indicator for died before 1 year old × 1,000 ($_1q_0$ or cohort IMR)								
TSC latrines per capita	-85.70*	-94.94**	-91.57**	-84.84*	-87.22*	-125.0**	-83.16*	43.55	69.11
	(34.82)	(34.28)	(34.14)	(34.21)	(34.21)	(45.36)	(37.98)	(65.82)	(63.15)
effect of program mean	-3.84	-4.25	-4.10	-3.80	-3.90	-5.60	-3.72		
effect of baseline std. dev.	-2.71	-3.00	-2.89	-2.68	-2.76	-3.90	-2.63		
district fixed effects	✓	✓	✓	✓	✓	✓		✓	✓
year fixed effects	✓	✓	✓	✓	✓		✓	✓	✓
sex, single birth, order		✓	✓	✓	✓	✓	✓		✓
household controls			✓	✓	✓	✓	✓		✓
mother & childcare controls				✓	✓	✓	✓		✓
health program controls					✓	✓	✓		✓
state × year fixed effects						✓			
village (PSU) fixed effects							✓		
n (infants)	164,795	164,795	164,762	164,762	164,762	164,762	164,762	33,492	33,482

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$.

Robust standard errors clustered by district in parentheses. "Effect of program mean" multiplies the coefficient estimate by the mean of program intensity in these districts in 2007.

Table 4: Mechanisms of the effect of the TSC on mortality, DLHS-3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
mortality rate:	post-neonatal mortality rate (months 2-12)					neonatal (month 1)	
TSC latrines	-50.01*	-48.12*	-51.51†	-47.17*	-48.21*	-35.69	-39.10
per capita	(24.04)	(24.33)	(29.60)	(23.55)	(23.63)	(28.74)	(27.25)
latrines × fraction after				-74.84*	-83.69*		
non-breast milk food				(32.48)	(32.99)		
fraction of first year after				-0.587	-2.734*		
non-breast milk food				(1.310)	(1.358)		
district fixed effects	✓	✓	✓	✓	✓	✓	✓
year fixed effects	✓	✓		✓	✓	✓	✓
state × year fixed effects			✓				
controls		✓	✓		✓		✓
n (infants)	164,795	164,762	164,762	164,795	164,762	164,762	164,795

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$.

Robust standard errors clustered by district in parentheses. "Fraction after non-breast milk food" is a constructed variable between 0 and 1: the fraction of the first year after a baby's birth that remained after she first had food other than breast milk, according to the mother's report; a larger number reflects earlier complementary feeding or weaning.

Table 5: Effects on district-level change in rural infant mortality

	(1)	(2)	(3)	(4)
	ΔIMR, 2010-11 minus 2001			
TSC latrines per capita	-44.87**	-40.84*	-35.97*	-35.56†
	(16.39)	(16.24)	(17.67)	(18.70)
effect of program mean	-4.55	-4.14	-3.65	-3.61
effect of baseline st. dev.	-1.42	-1.29	-1.14	-1.12
latrine coverage, 2001		0.184***	0.213***	0.189***
		(0.0418)	(0.0544)	(0.0528)
change in female literacy			0.152	0.191
			(0.282)	(0.273)
other change controls			✓	✓
population, 2001 census				2.52×10^{-6}
				(2.09×10^{-6})
village count, 2001 census				-0.000900
				(0.00233)
population × village count				1.22×10^{-9}
				(1.08×10^{-9})
distance to state capital				0.00909
				(0.00967)
constant	-15.65***	-20.89***	-23.70***	-26.95***
	(1.743)	(2.142)	(3.952)	(5.557)
n (districts)	280	280	280	280

	(5)	(6)	(7)	(8)
	IMR, 2010	IMR, 2001	ΔIMR, 2001-1991	ΔIMR, 1991-1981
TSC latrines per capita	66.31***	101.9***	-24.14	10.41
	(16.82)	(19.05)	(25.03)	(35.53)
all controls from column 4	✓	✓	✓	✓
n (districts)	280	280	280	258

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$. Heteroskedasticity robust standard errors in parentheses. "Effect of program mean" multiplies the coefficient estimate by the mean of program intensity in these districts in 2010.

Table 6: District-level falsification tests

	(1)	(2)	(3)
Panel A:"Effect" of TSC on change in urban IMR			
TSC latrines per capita	-20.62	38.15	
	(23.76)	(27.29)	
controls		✓	
n (districts)	197	197	
Panel B: "Effect" of JSY on change in rural IMR			
Fraction of pregnant	-15.82	13.35	
women who received JSY	(9.65)	(11.82)	
controls		✓	
n (districts)	280	280	
Panel C: "Effect" of NREGA on change in rural IMR			
Fraction of households who	-0.277		
received job cards	(2.687)		
Fraction of households who		-4.479	
received work		(4.105)	
Fraction receiveing work,			-0.887
among those with cards			(3.021)
controls	✓	✓	✓
n (districts)	280	280	280

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$. Heteroskedasticity robust standard errors in parentheses. Sample size n is smaller in Panel A because the census does not report rural IMR for many primarily urban districts. "controls" are the complete set of controls used in table 5.

Table 7: Effects on ΔIMR, UNICEF data on TSC intensity in 47 districts in Uttar Pradesh

	(1)	(2)	(3)	(4)	(5)	(6)
	change in rural IMR			change in urban IMR		
Δ h.h. latrine coverage	-18.08†	-17.41†	-23.33*	11.41	14.49	-9.119
due to TSC	(10.23)	(9.739)	(10.93)	(17.85)	(17.63)	(19.95)
effect of baseline std. dev.	-3.32	-3.19	-4.28			
Δ female literacy			-2.204*			-1.121
			(0.903)			(1.502)
Δ literacy			2.742*			5.018**
			(1.042)			(1.708)
Δ sex ratio			-0.0198			-0.337
			(0.233)			(0.347)
population growth			0.0892			-0.220
			(0.216)			(0.346)
constant	-17.14**	-17.81**	-14.86	-3.628	-4.859	-40.51
	(2.235)	(2.170)	(11.80)	(3.758)	(3.651)	(25.73)
weight	none	sample	none	none	sample	none
n (districts in UP)	47	47	47	36	36	36

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$.
Heteroskedasticity robust standard errors in parentheses. Change in latrine coverage is measured in latrines per household. Columns 2 and 5 are weighted by the square root of the district sample size.

Table 8: Clean village prize (NGP) incentive by village population

	village population in the 2001 census:				
	below 1,000	1,000 to 1,999	2,000 to 4,999	5,000 to 9,999	over 10,000
rupees:	50,000	100,000	200,000	400,000	500,000
dollars (market):	1,000	2,000	4,000	8,000	10,000
dollars (PPP):	3,400	6,800	13,600	27,200	34,000

Table 9: Effect of NGP incentive on endline (2010/11) rural IMR, Annual Health Survey

(1)	(2)	(3)	(4)	(5)			
independent variable based on the incentive:					controls included:		
incentive per capita	fraction within 50% above cut	fraction within 40% above cut	fraction within 30% above cut	fraction within 30% below cut	2001 IMR	2001 toilet coverage	population per village
-0.209	-28.72	-33.95	-42.22	100.2			
(0.129)	(14.44)	(17.41)	(22.75)	(28.52)			
			$p = 0.000$				
-0.285	-29.84	-31.61	-37.63	40.20	✓		
(0.094)	(10.54)	(12.95)	(16.84)	(17.36)			
			$p = 0.001$				
-0.277	-30.05	-32.22	-39.73	42.53	✓	✓	
(0.094)	(10.29)	(12.57)	(16.47)	(21.18)			
			$p = 0.002$				
-0.461	-32.22	-27.92	-30.36	38.72	✓	✓	✓
(0.199)	(15.40)	(16.66)	(18.77)	(20.44)			
			$p = 0.013$				
median effect on endline IMR of a one standard deviation difference:							
-2.31	-2.36	-3.18	-4.50	2.32			
		n =280 districts					

Each estimate is from a separate regression of 2010/11 IMR on that column's independent variable. Heteroskedasticity robust standard errors in parentheses. The p-values between columns 4 and 5 reflect a test of the null hypothesis that the coefficients on 30 percent above and below are equal. The "median" for each column is over the four rows.

Table 10: Effect of TSC latrines on height-for-age, IHDS 2004-5

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sample:	rural	rural	rural	rural	rural	rural	rural	rural	urban
ages included (months):	0-59	0-59	0-59	0-59	0-59	0-59	0-59	24-59	0-59
TSC household latrines	8.582*	6.640†	7.682†	6.323†	6.302†	6.216†	10.335*	11.58*	-0.836
per capita	(3.815)	(3.651)	(4.466)	(3.656)	(3.670)	(3.686)	(4.331)	(5.697)	(4.04)
effect of program mean	0.20	0.16	0.18	0.15	0.15	0.15	0.24		
effect of baseline std. dev.	0.27	0.21	0.24	0.20	0.20	0.20	0.33		
district fixed effects	✓	✓	✓	✓	✓	✓		✓	✓
year of birth fixed effects	✓	✓		✓	✓	✓	✓	✓	✓
age (mo.) × sex indicators		✓	✓	✓	✓	✓	✓	✓	✓
state × year fixed effects			✓						
household controls				✓	✓	✓	✓	✓	
mother controls					✓	✓	✓	✓	
sibling 8-11 height z						0.182***	0.195***	0.192***	
						(0.044)	(0051)	(0.050)	
no sibling 8-11						-0.536***	-0.602***	-0.592***	
						(0.119)	(0.138)	(0.122)	
village fixed effects							✓		
n (children)	7,299	7,299	7,299	7,299	7,299	7,299	7,299	4,802	3,110

Two-sided p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. One-sided p-values: † $p < 0.05$. Clustered standard errors in parentheses. "Effect of program mean" multiplies the coefficient estimate by the mean of program intensity in these districts in 2005.

Supplementary Appendix

Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from India's Total Sanitation Campaign

Contents

Quantitative evidence that the TSC occurred as described

This section presents household-level evidence from the IHDS that the TSC happened, meaning that places in which more latrines were supposed to have been built indeed saw more latrine construction, and that this is not just a spurious correlation with some omitted variable. Recall that the IHDS was conducted in 2005; therefore, the independent variable here is district cross-sectional TSC intensity in 2005. The IHDS contains a continuous measure of consumption per capita (rather than, say, merely an asset index or indicators for quintiles of an asset index as in the DHS and DLHS), which allows me to control for a quartic polynomial of consumption. The observations in these regressions are rural households; the IHDS is nationally representative.

Higher reported construction is correlated with less open defecation

Table A1 presents regressions of an indicator for open defecation (measured as reporting no access to a toilet or a latrine) on TSC intensity, with and without state fixed effects and a set of controls. District-level reported intensity of the TSC (measured as latrine construction) is associated with households being less likely to report not having access to a latrine or toilet. There is no similar effect on urban households: using the urban dependent variable, the sign flips from positive to negative when controls are added, and even when negative the *t*-statistic is less than one.

Table A1: The TSC is associated with less open defecation, IHDS

	(1)	(2)	(3)
	no household latrine or toilet, 2005		
TSC latrines per rural capita built by 2005	-0.541*	-0.646*	-0.586*
	(0.265)	(0.316)	(0.298)
state fixed effects		✓	✓
consumption[4]			$F = 81$
			$p = 0.000$
BPL card			0.0437***
			(0.0103)
caste & religion groups			$F = 35.5$
			$p = 0.000$
Constant	0.734***	0.736***	0.729***
	(0.0126)	(0.0108)	(0.0294)
n (rural households)	25,135	25,135	25,135

Clustered standard errors in parentheses. Two-sided *p*-values: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

TSC construction is not correlated with wealth or disease environment

Table A2 presents placebo regressions of other "outcomes" on TSC intensity. In the IHDS, TSC construction is not correlated with any of six other household-level inputs into children's health (Panel A), or any of six other household-level indicators of economic well-being (Panel B).

Table A2: The TSC was not correlated with wealth or the disease environment, IHDS

Panel A: TSC intensity was not associated with other features of children's disease environment

	(1)	(2)	(3)	(4)	(5)	(6)
dependent variable:	hand-wash with soap	diarrhea correct	water source in house	house has electricity	clean cooking fuel	mother's age at marriage
TSC latrines per rural person built by 2005	0.281	0.422	-0.686	0.431	-0.306	0.407
	(0.337)	(0.367)	(0.422)	(0.302)	(0.325)	(2.637)
state fixed effects	✓	✓	✓	✓	✓	✓
consumption[4]	✓	✓	✓	✓	✓	✓
caste & religion groups	✓	✓	✓	✓	✓	✓
Mean	0.324	0.430	0.402	0.612	0.260	16.65
n (rural households)	25274	25274	25274	25274	25274	20349

Panel B: Total Sanitation Campaign intensity was not associated with households' wealth

	(1)	(2)	(3)	(4)	(5)	(6)
dependent variable:	consumption per capita	ration card for poor	has table	has pressure cooker	has mixer	has clock
TSC latrines per rural person built by 2005	-47.46	0.0864	0.439	0.418	-0.0292	0.181
	(283.76)	(0.334)	(0.333)	(0.304)	(0.138)	(0.265)
state fixed effects	✓	✓	✓	✓	✓	✓
consumption[4]			✓	✓	✓	✓
caste & religion groups	✓	✓	✓	✓	✓	✓
Mean	712	0.417	0.568	0.250	0.109	0.788
n (rural households)	25274	25274	25245	25102	25274	25231

Clustered standard errors in parentheses. Two-sided *p*-values: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Qualitative evidence from the field on effects of the NGP incentive and sources of heterogeneity in TSC implementation

While conducting the research for and writing this paper in the fall of 2011 and spring of 2012, I predominately lived in a mainly rural part of Sitapur district, in the Indian state of Uttar Pradesh, with several visits to Delhi and some travel to other states in north India. This gave me opportunities to verify administrative claims about the TSC and to better understand heterogeneity in TSC implementation.

A village which won the prize

According to the TSC administrative records, exactly one village in my home block won the NGP. It won in 2008. Visiting this village would help verify the quality of the data: many villages here do not have many latrines, but the government records claim that this village does. If it did, and if people appeared to be using them, that would support the reliability of the data.

Indeed, when I arrived unannounced in early 2012, 4 years after the village won the prize, I found that there were about as many latrines as households. Figure A1 shows a latrine bearing the consecutive numbering that reveals them as a government project. They were ubiquitous, in rich and poor parts of the village. In the "downtown" part of the village, near the village council building, there was a shared sanitary complex, but other than some urinals on the men's side, it appeared to be a defunct tragedy of the commons.

Figure A1: Latrines, labeled and numbered

Perhaps more importantly, the latrines appeared to be in use. They typically had doors or a piece of cloth hung over the entryway (a good sign that a latrine is not in use is that somebody has repurposed the door for housing material). Most of them had a "mug," a cup for water that Indian people use for hygiene. Some of the better-off households had two latrines, and were using one for storage but also one as a latrine. Figure A2 shows the inside of a latrine, with visible indications of recent use.

Figure A2: A latrine in use

Why did this village attempt to win the prize? I was unable to talk with the village chairman[1] – or, more probably I actually did, but she would not admit it. Like in the story of a prizewinning village told by Black and Fawcett (2010), in this village a male pradhan had managed to hang onto power when his job cycled into reservation for a woman by making his wife the pradhan. When I went to the pradhan's house – painted a blue color signaling high caste – and asked for the pradhan, some women leaned out and explained that all of the men were away.

After returning from the village, I looked up its population in the 2001 census: 1,050 people. At just above the largest discontinuity (1,000), this translates into an incentive per capita of over 95 rupees, close to the maximum arithmetically possible under the scheme. It may be little surprise that this village invested in becoming open defecation free, given its large incentive to do so.

[1] Throughout this paper, I have used the term "village chairman" to refer to the person called the *pradhan* in Sitapur, Uttar Pradesh, and whom I have heard called the *sarpanch* in Gujarat, Madhya Pradesh, and Rajasthan.

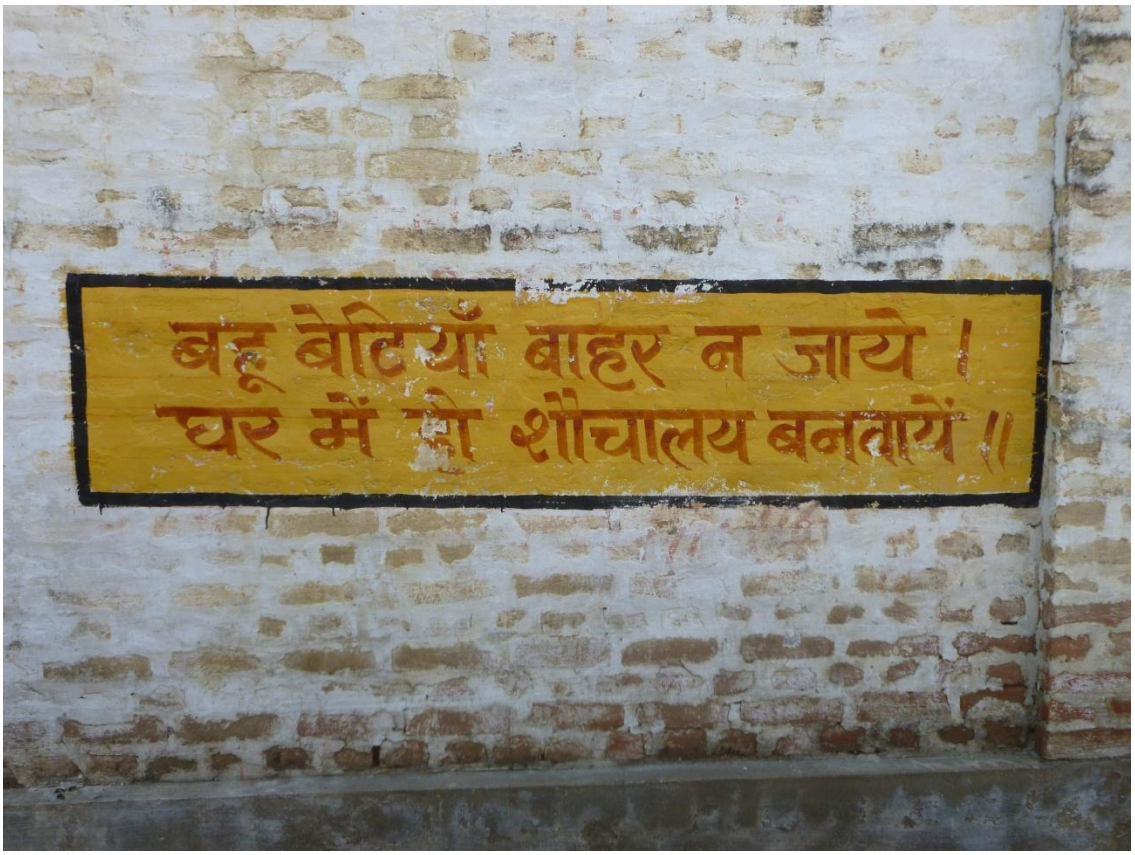


Figure A3: A motivational message

Figure A3 depicts a sign painted on a wall near a busy foot path in the village. Its Hindi message roughly translates into "don't let your daughters and daughters in law go outside – make a toilet in your house!" I have seen similar messages painted on walls in other villages in Uttar Pradesh and in Madhya Pradesh. This is an example of the TSC's emphasis on motivation to use latrines in addition to construction.

This slogan also illustrates another aspect of the TSC: its use of existing rural social structures, in this case, gender norms and hierarchy. Implied in the slogan is that a male householder is making the decision, and that he would want to keep young household women away from the public. The TSC has been unashamed to make use of conservativism in Indian culture. While this may, to some observers, have a cost in what it positively norms (Cialdini, 2003), if the program indeed improves children's health it may be worth it in a world of second-bests. As Bathran (2011) details in a review, Alok's (2010) memoirs and account of the TSC are full of apparently approving references to notions of purity and cleanliness associated with the Hindu caste system.

A village which did not win the prize

I was able to talk with the pradhan of another village, one which did not win the prize. Although most villages did not win the NGP, I selected this one to interview the pradhan because it was within a half-hour bicycle ride from my home outside of the district capital and it was not among the villages that my wife was studying in an unrelated project on the health of children and pregnant women.

On the way into the village I saw about three latrines: one a very fancy ventilation-improved latrine of the sort rarely seen in India with a solid brick superstructure, and another inside a triangular corner made by piling bricks a few feet out from the wall of a building. Unfortunately, there was an engagement party in the village, and when our meeting with the pradhan started, everybody crowded around to see the visitors, that is, us. So, the pradhan made a big point of saying that the government money for latrines would never be enough to build one for everybody. He was happy to tell me that in his village when people had to go, some used latrines, but most just "go outside."

The conversation turned to his proudest achievement in office: a dirt road with an impressive irrigation ditch. This presented an opportunity to get away from the crowds, and the pradhan was easily persuaded to head out by motorcycle and show off his road. It turned out that the project had been conducted with NREGA funds, and that the surrounding irrigated fields were essentially his. So, it appeared that elite capture of this particular public good was the way he had elected to use his term of office.

When I asked again, he said that he knew about the NGP. Indeed, he knew another pradhan who had won the prize and had gotten to travel to Delhi, and had even received one lakh rupees. However, he was not impressed or motivated by the incentive: his village was too large (with a 2001 population in the 3,000s [he mentioned 4,000], he was far from the 2,000 and 5,000 cutpoints, and faced a low incentive per capita of 65 rupees).

Moreover, somebody had apparently drawn a map badly, and one of the hamlets in his villages was on the other side of a creek.[2] That hamlet had been in a political fight with the rest of the village for at least a generation. Becoming "open defecation free" would require the cooperation of everybody, and neither side would be interested in helping the other win. Given the village's population relative to the incentive discontinuities, and given the haphazard lines on a district official's map, extracting an irrigation canal from NREGA was a more attractive use of his attention.

Interviews with district, NGO, and central government officials

Interviews with officials at other levels of the TSC's organization have consistently indicated that:

1. Heterogeneity in the TSC does not appear to originate in any top-down targeting, according to government goals or further heterogeneity that would threaten exogeneity. As a central government program, it has always been intended for everybody.
2. Heterogeneity in implementation intensity appears to importantly originate in the interests of local political actors, added up over many villages.

I met the District Magistrate, sometimes called the District Collector, a post to which an Indian Administrative Service officer is assigned relatively early in his career (at least, in undesirable districts such as this one). The District Magistrate – known as the "DM" – has a large amount of legal executive authority over the district, powers which are in practice limited by low state capacity and short tenures.

[2] A 20 minute bicycle ride around the creek reveals that this extra hamlet is in fact contiguous with another village! Arbitrary division of contiguous homes into *gram panchayats* appears to be common in this district.

We met in his large office in an old house, far from the commercial center of the district capital. I asked the DM if he had ever heard of the Total Sanitation Campaign, suspecting that the answer might be no. "Heard of it?" he leapt out from behind his desk to sit next to me. He had pulled out from under his desk a stack of booklets in Hindi: these had his picture on the cover with the district official responsible for village councils, and were full of detailed instructions for how a pradhan should implement the TSC, complete with diagrams for latrine construction. He proudly explained that he had written much of it himself. Like many IAS officers, he was originally trained as an engineer, a civil engineer in his case. "Distributed to everybody!" he beamed about the books.

Apparently, this DM had taken a special, personal interest in the TSC. He put together these booklets and summoned the district's pradhans to a series of special meetings to encourage them to comply, enough meetings so that he could interact with them all. (A pradhan whom I later met confirmed the DM's story). The DM explained to them the externalities of infectious disease: if you have a toilet and your neighbors do not, "the germs will not differentiate between them and you." Grinning, he recounted how he attempted to shame them, reminding them that they spend so much money on festivals as a matter of pride, but apparently do not have enough pride to keep their daughters in law from openly defecating.

All pradhans have special TSC accounts on which they can draw for the subsidy money (at least in this district), to organize construction with a contractors of their choice. As the DM told them "everything is there, only the willpower is needed." But, not all of the pradhans had implemented the program. When I asked, he guessed – now frowning – that maybe 20% in his district were trying. Strikingly, the chief executive of this district of over 4 million people felt ultimately powerless to do much more than advocate and exhort, even for a program so clearly important to him.

This meeting was on a Friday, and when I left we made rough plans to see each other again the next week, seemingly sincerely. But by Sunday, a new DM had been unexpectedly installed, and our old DM had been moved to another district. He had had the job for less than a year; he hadn't yet gotten around to redecorating his office. Apparently the new DM had been seen the day before interfering in an important state election in her district, and was shifted to ours, partially as a punishment, partially to block her meddling.

This episode illustrates the haphazard variation in political commitment to the implementation of the TSC. Although I do not know about the new DM's commitment to the program, the old DM likely would have been pessimistic. When I asked whether he knew of other DMs that were as enthusiastic about the TSC, he diplomatically but unmistakably allowed "I cannot say about what others are doing. I can only speak for me."

. . .

Throughout the time I was working on this project, I met several times with an employee responsible for TSC activities in a regional office of UNICEF. As he[3] described it, UNICEF is a "catalyst" for the TSC that aims to "instigate the system to move" and does this primarily by being "the technical support to the

[3] For clarity, I refer to the employee as "he"; to protect "his" anonymity I randomized the gender before writing this.

government." He showed me Power Point presentations that he had given to newly installed state government officials urging them to pay attention to the TSC.

The UNICEF employee was frank that the TSC was being implemented much more actively in some places than in others. He emphasized the need for "interpersonal communication" to persuade people to build and use latrines – the sort of influence that the TSC was trying to use in villages structures, although perhaps it was not always trying hard enough. He described the joint challenge of tradition and externality: "It's an age old problem. People are openly defecating. Well, I am not harmed."

He quite openly discussed corruption among the village chairmen and improper use of latrine superstructures. The office building had been constructed with an extra bathroom attached to his shared office. The bathroom was being used as a closet. When he went in to find pamphlets detailing the proper construction of latrines, he joked "we are learning from them" – meaning the villagers, whom I have seen store tools and food in their latrine superstructures.[4]

In his view, a key source of heterogeneity in implementation was "the P factor: the Pradhan factor," using the local term for the village chairmen. "Where the Pradhan was good, the opportunity [to implement the program] was good." He described that it is up to the Pradhan to sort out how the latrines will be constructed: for example, will the work for the whole village be contracted out together? Often, "Pradhan simply finds a mason." In short, "if the person is good he can make a difference. Lots of the program has depended on the Pradhan's influence."

In his experience, it was also the case that the NGP was a "motivating force." The prize was not announced until late 2003 and "once the award was started, the numbers increased like anything." He confirmed my impression that incentive money was largely in the hands of the Pradhan: "it is his choice whether he wants to give the incentive money to the people directly or whether he wants to get it himself. It is a different approach everywhere. Who should the money be given to? It is up to the Pradhan… There is corruption in the system."

District-level commitment also mattered, according to the UNICEF employee, but not as much. A key variable partially under district control is ensuring vacancies are filled in "motivator" positions, jobs for people who advocate compliance with the TSC.

He was more readily able to speculate on a process evaluation than an impact evaluation. Because he was very familiar with the implementation "bottlenecks" of the TSC, he was skeptical of the overall project. But he predicted it would have an effect on health in those places where it displaced "dry latrines" – what he called "the killer toilet." In an area for future research, he wondered whether improperly constructed latrines could, one day, contaminate groundwater.

[4] In one village, my wife saw latrine superstructures built with government funding being used to store bicycles or house goats. The village chairman appeared be building the latrines in stages: he hired one contractor to build the superstructures first, and another to construct the latrine pits second (which is technically feasible; they need not be right below). When she went back a few weeks later, the pits were indeed being dug.

Individual-level parallel trends: Evidence from the DLHS-2

The individual-level analysis of section 2 depends, like all fixed effects identification strategies, on the "parallel trends" assumption that there would have been no difference in the changes over time across districts, on average, correlated with the TSC implementation if there had been no TSC implementation. This is impossible to ever verify, but one thing that is possible is to look for "effects" of TSC latrines before the program.

Section 2 uses the third round of the DLHS. The second round of the DLHS (that is, a separate data set) includes birth history information on children born in 1997-2000, the four years before the TSC began building latrines, a length of time comparable to the four year stretch (2004-2007) used in section 2. To test for parallel trends before the program, I regressed 1997-2000 infant mortality (from the DLHS-2 this time) on 2004-2007 TSC latrine construction (the exact same independent variable used in section 2), with the same district and year fixed effects specification, anticipating finding no effect.

Table A3: The TSC was not correlated with individual-level IMR before the program, DLHS-2

	(1)	(2)	(3)	(4)
sample:	rural	rural	rural	urban
	individual-level IMR × 1,000 (1997-00)			
TSC latrines per	-8.171	-20.35	-18.58	95.57
capita (2004-07)	(59.37)	(63.90)	(63.43)	(65.63)
demography controls		✓	✓	
household controls			✓	
n (infants)	216159	216159	216157	76889

All columns include district and year fixed effects.

Clustered standard errors in parentheses. Two-sided *p*-values: + $p<0.10$,* $p<0.05$, ** $p<0.01$.

Any statistically significant correlation would be evidence against parallel trends, but nothing is statistically significant even at the 10% level. The rural coefficient estimate is approximately 10% the size of that in section 2.

The NGP incentive: village-level "first stage"

With ideal data, I would estimate a village-level regression discontinuity: a few years after the incentive is implemented, is infant mortality lower in villages with populations just above prize discontinuities than in villages with populations just below prize discontinuities? This would require village-level health outcome data, matchable with real village names to 2001 census data. The second round of the DLHS, which occurred before the NGP, makes village names public, but the third round, used in this paper, does not. I traveled to Mumbai to request this data from IIPS and was unsuccessful.

Section 3.3 of the paper regresses district-level endline IMR on district-level aggregates of the village incentives. This is motivated by the village-level discontinuity. If I do not have village-level health outcomes, could I at least verify at the village level that the discontinuity impacts TSC latrine construction?

This exercise faces data constraints, as well. Matching the TSC village-level data to 2001 census population data is very difficult: it must be done by hand, using village names, without knowing what block the village is in. Village names are often spelled differently, or are altogether different (for example, a surveyor accidentally wrote down the name of a hamlet instead, or Hindus and Muslims may use different names for the same place).

Even if this procedure were easy, it may not be very valuable. The village construction data is not a panel; there is one number recorded per village, which is replaced when a new number is eventually entered. However, it is not a cross-section either, because it has been updated irregularly: in 2011 for some villages, not since 2008 for others. The relevant ministry in Delhi reported having no record of historical village construction beyond this administrative database. A high ranking official in the central government unit responsible for the TSC told me about as directly as one could expect not to trust this data much. One reason is that it is, in principle, input by block officials, not district officials; blocks are subdivisions of districts in the Indian government and are typically very low capacity.

. . .

With all of these caveats, is there any evidence of a first-stage effect of the incentive on construction at the village level? I approached this question in two complementary ways, constructing two datasets. The results are presented in tables A4 and A5.

To construct the data set used in table A4, I randomly drew a sample of villages from the 2001 census, among villages with populations within 100 people of the discontinuities. 460 of these were able to be matched to TSC data, based on the village name. The advantages of this data are that its sample is focused on the discontinuity; it uses real 2001 census population data; and it is from all states. The disadvantages are that that the dataset is very small; almost no controls are available; and there is no reason to believe that matchability of village names is ignorably "random."

To construct the data set used in table A5, I matched TSC data to the TSC's own baseline survey for 50 districts. Entire blocks (sub-district units) were randomly selected, to be representative of the 280 districts used in the district-level analysis. Unfortunately, the TSC's baseline survey recorded the count of households, not populations, so I estimated each village's population by multiplying the count of households in each village by the average household size for that district in the 2001 census. This

estimation would be expected to attenuate the estimated effect of the incentive (some villages will be placed on the wrong side of the discontinuity).

With either dataset, I find that a higher incentive predicts more latrines being built, whether this is operationalized as OLS latrines per capita, or a poisson count with population exposure (or zero-inflated poisson), including with district fixed effects. The table A5 data from the TSC allows several controls, including for population itself and for the baseline fraction of households with latrines (suggesting that this is not just some sort of convergence to the mean).

Unfortunately, I do not have nearly enough data to compute a technical regression discontinuity. However, I can say this: the most densely populated discontinuity in the prize is around 1000 people. Among the 79 randomly-selected villages with populations estimated to be between 1000 and 1200, 7 (or 9%) of them won the "open defecation free" prize; among the 64 villages with populations from 800 to 1000, none of them did.

Table A4: Village construction data matched to population data matched from 2001 census, sampled around prize discontinuities

	(1)	(2)	(3)	(4)	(5)
model:	OLS	OLS	Poisson	poisson	zero-inflated poisson
dependent variable:	TSC latrines per capita		count of TSC household latrines		
incentive per capita	0.000395+	0.000337+	0.00751**	0.00690**	0.00617*
	(0.000208)	(0.000192)	(0.00268)	(0.00244)	(0.00247)
state fixed effects		✓		✓	
exposure			population	population	population
constant	0.0559**		-3.118**		-2.798**
	(0.0133)		(0.177)		(0.159)
inflation constant					-1.181**
					(0.110)
n (villages)	460	460	460	460	460

Heteroskedasticity-robust standard errors. Two-sided *p*-values: + $p<0.1$, * $p<0.05$, ** $p<0.01$.

Data were originally sampled from all India from the 2001 census, stratified by state, but the final dataset includes only villages that could be matched.

The coefficient estimates are not as dissimilar across OLS and poisson models as they might seem. A 10-rupee increase in the incentive would predict a difference in TSC intensity of about 0.4 of a percentage point, according to the OLS results. It would predict a difference of exp(0.07) or about 7 percent, which would be about 0.3 of a percentage point at 2005 average construction levels.

Table A5: Population data estimated from TSC baseline survey, representative of 280 districts used in discontinuity regressions in section 3.3

	(1)	(2)	(3)	(4)	(5)	(6)
model:	OLS	OLS	OLS	OLS	poisson	z-i poisson
dependent variable:	TSC household latrines built per capita				count of TSC latrines	
incentive per capita	0.000565**	0.000211*	0.000138+	0.000161*	0.00844**	0.00807**
	(0.000158)	(0.0000859)	(0.0000688)	(0.0000626)	(0.00172)	(0.00164)
village population		-0.00000484**	-0.00000526**			
		(0.00000165)	(0.00000141)			
baseline fraction of			-0.0161		0.192	0.0677
hh with latrines			(0.0169)		(0.238)	(0.236)
household size			-0.0200**		-0.319**	-0.272**
(district estimate)			(0.00519)		(0.0953)	(0.0938)
data entry dummies			✓	✓	✓	✓
district fixed effects				✓		
exposure					population	population
constant	0.0246*	0.0637**	0.162**		-2.354**	-2.308**
	(0.0102)	(0.0114)	(0.0346)		(0.567)	(0.541)
inflation constant						-2.682**
						(0.290)
clusters (districts)	50	50	50	50	50	50
n (villages)	2311	2311	2311	2311	2311	2311

Standard errors clustered at the district level. Two-sided *p*-values: + $p<0.1$, * $p<0.05$, ** $p<0.01$.
"Data entry dummies" are dummies for 2008 through 2011, the year that the village's data was last updated.

Reassuringly, the point estimates from the two tables – computed with different samples of villages, methods, and data sources – are not statistically significantly different from each other:

- Column 2 A3 to column 1 A4 (smaller to largest): $t = 0.92$.
- Column 1 A3 to column 3 A4 (larger to smallest): $t = 1.1$.

The NGP incentive: district-level falsification tests

Section 3.3 demonstrates that the more a discontinuity in the prize structure increases the average NGP incentive in a village, the lower its endline infant mortality. This is found with or without a range of controls, and with several implementations of the district-level incentive measure.

Although no test can ever prove that an independent variable is exogenous, it would be cause for concern if the incentive district aggregates were clearly correlated with other factors known to influence infant mortality. This section uses the IHDS data on rural households in the 280 districts used in the district-level analysis. I use a set of dependent variables: consumption per capita, asset count, literacy, whether the mother had an antenatal checkup for her birth (a smaller sample because not all women had pregnancies in the reference period), whether the household received PDS gain (perhaps a measure of state capacity or activity), and whether the mother in the "eligible women" module correctly answers that children should be given more fluids. The independent variable is the district average incentive, used in column 1 of table 9.

Table A6: The average NGP incentive in a district does not predict other variables, IHDS

	(1) consumption per capita	(2) asset count	(3) any literate	(4) antenatal checkup	(5) received PDS grain	(6) diarrhea correct
Panel A: Without population controls						
incentive	-2.844	-0.0226	0.000405	0.00449	0.00476*	-0.00498
	(2.053)	(0.0287)	(0.00194)	(0.00257)	(0.00192)	(0.00331)
constant	807.3***	9.422***	0.648***	0.225	-0.209	0.920***
	(148.1)	(2.088)	(0.142)	(0.186)	(0.137)	(0.244)
n	10875	10885	10885	3927	10885	8763
Panel B: With population structure controls						
incentive	-0.800	-0.0472	-0.00455	0.00262	0.00432	-0.00840
	(4.150)	(0.0338)	(0.00252)	(0.00431)	(0.00444)	(0.00521)
controls	✓	✓	✓	✓	✓	✓
constant	738.9*	10.52***	1.080***	0.384	-0.249	1.099**
	(297.3)	(2.571)	(0.198)	(0.303)	(0.309)	(0.415)
n	10875	10885	10885	3927	10885	8763

Clustered standard errors in parentheses. All columns use state fixed effects. * p<0.05, ** p<0.01, *** p<0.001. The sample is rural households in the 280 districts used in section 3 of the paper.

One coefficient is statistically significant, but only without controls. There is a 46 percent chance of seeing at least one coefficient statistically significant at the 95% level from 12 nonsense regressions.

The NGP as an instrument for 2011 census open defecation

The 2011 Indian census measured sanitation coverage for rural and urban parts of Indian districts. A first use of this data is to assess the quality of the TSC administrative data. A one-unit increase in TSC reported latrines is associated with a 0.25 unit decrease in *rural* open defecation in the 2011 census (t = -2.14, controlling for state fixed effects, district size, and baseline coverage), consistent with both a signal and some measurement error in the TSC statistics, but is not associated with *urban* open defecation (t = 1.34).

A further use of the 2011 census data is to verify that the district-level NGP variable from section 3.3 is, indeed, associated with less open defecation in the 2011 census. Table A7 reports this analysis, with various sets of controls. The expected negative effect is found. The coefficient estimates are larger in absolute value than those in tables A4 and A5, which is, again, consistent with measurement error in the official TSC data.

Table A7: Instrumentation first stage: the prize predicts endline open defecation

	(1)	(2)	(3)
	rural open defecation, 2011 census		
NGP prize per capita	-0.00282**	-0.00802***	-0.00292*
	(0.00090)	(0.00124)	(0.00131)
baseline coverage	-0.00874***	-0.00823***	-0.00640***
	(0.00037)	(0.00038)	(0.00072)
2001 IMR		0.00184***	0.00172*
		(0.00044)	(0.00067)
population per village		-0.000074***	-0.000035+
		(0.000015)	(0.000018)
state FEs (9 states)			✓
n (districts)	268	268	268

As the title of table A7 hints, it can also be interpreted as the first stage in a regression of 2011 infant mortality on 2011 open defecation as reported by the census, with the district-level NGP variable as an instrumental variable for open defecation. With a t statistic of -3.03 on the NGP incentive, a weak instrument is probably not a concern.

***y*: dependent variable**	***x*: independent variable**	***z*: instrumental variable**
2010-11 AHS infant mortality	2011 census open defecation	computed NGP prize per capita

Open defecation would be expected to be positively associated with infant mortality. Should we expect the IV estimate to be greater than or less than the OLS estimate? On the one hand, the IV estimate could be larger if census open defecation is measured with error, as it almost surely is. On the other hand, the IV estimate could be smaller if confounding factors positively correlated with open defecation are also positively correlated with infant mortality.

Table A8: Instrumental variable estimate of the effect of open defecation in infant mortality

	(1) OLS	(2) IV	(3) IV
	endline (2011) rural infant mortality		
rural open defecation (2011)	29.3*	89.3*	57.4*
	(12.2)	(41.1)	(25.3)
		instrument: NGP prize per capita	
baseline coverage	0.167	0.685+	0.528*
	(0.118)	(0.357)	(0.204)
2001 IMR			0.442***
			(0.081)
population per village			0.0016
			(0.0011)
n (districts)	268	268	268

Table A8 presents the results. None of the estimates are statistically significantly different from one another. Most striking is that the main IV estimate in column 2 – moving linearly from 0 to 1 open defecation would be associated with an increase in IMR of 89 – is quite similar to the main estimates of the paper reached using altogether non-overlapping data sources: the fixed effects estimate of about 85 infant deaths per 1,000 from table 3.

As always, this is a local average treatment effect of the effect of open defecation on IMR for districts moved by the NGP, which is perhaps more opaque than a simple regression, but is also perhaps of policy relevance. Finally, as a falsification test, there is no detected effect if urban infant mortality is instead substituted as the dependent variable (t = 1.08).

Effect sizes: comparisons with the literature

How do the estimates in this paper compare with effects of other phenomena on the same outcomes, as estimated in the econometric and related literature? In general, the effects found of the TSC are of plausible magnitude, on the order of what other studies find, and smaller than some cases.

Height for age: the median Indian child is at -2				
effect	what is it an effect of	source	econ	age
0.2	*Total Sanitation Campaign, India (0 to mean)*	*this paper*		*0 - 4 years*
0.2	*Total Sanitation Campaign, India (effect of one pre-program standard deviation)*	*this paper*		*0 - 4 years*
-.18	community "poor sanitation and water" in Ghana	Lavy, Strauss, Thomas, & de Vreyer (1996)	✓	up to 11 years
0.22	Matlab Maternal and Child Health and Family Planning Program in Bangladesh (vaccinations, maternal care)	Barham (AEJ: Applied 2012)	✓	8-14 years
0.3	micronutrient supplementation on <12 months in Mexico (>12 months no effect)	Rivera, et al (2001)		8-14 months
-.33	"inadequate sanitation" in Peru	Checkley, et al (2004)		24 months
0.35	zinc supplementation meta-study	Brown, et al (2002)		
0.44	having a "village midwife" (1 year of training) in Indonesia	Frankenburg, Suriastini, & Thomas (2005)	✓	1 - 8 years
0.67	INCAP: Atole rather than Fresco, Guatemala	Habicht, et al (1995)		3 years
1.16	woman receives South African pension, effect on girl child	Duflo (WBER, 2003)	✓	0 - 4 years

Early-life mortality rates: an average here is in the 50s in India				
effect	what is it an effect of	source	econ	rate
-3 to -4	*Total Sanitation Campaign, India (0 to mean)*	*this paper*		0-1 (IMR)
-2 to -3	*Total Sanitation Campaign, India (effect of one pre-program standard deviation)*	*this paper*		0-1 (IMR)
2.4	difference of 1 ppm of CO pollution in California	Currie & Neidell (QJE 2005)	✓	0-1 (IMR)
~ 4 or 5	air pollution from 1997 Indonesian wildfires	Jayachandran (JHR 2009)	✓	0-2
-5.4	privatizing water in Argentina	Galiani, *et al* (JPE 2005)	✓	0-4 (CMR)
-6 to -13	effect on blacks of integration of a hospital in Mississippi	Almond, Chay, & Greenstone (AER, forthcoming)	✓	1-12m. (PNNMR)
-10.7	crossing "very low birth weight" threshold in US (1.5 kg), thus triggering extra medical care [*see also* Barreca, *et al* 2011]	Almond, *et al* (QJE 2010)	✓	0-1 (IMR)

A model justifying the use of fixed effects

The central independent variable of the paper is not district sanitation coverage but the change in district sanitation coverage due to the program, since some unobserved prior period. Could this property of the data interact problematically with the use of district fixed effects? This section addresses this question and finds that fixed effects are appropriate.

Individual *i* lives in village *d* through years indexed *t*. We can observe (some features of) the individual's health h_{ivt}. We cannot observe a latent property of the individual: d_{ivt}, or disease of the bowels.

At time *t*, sanitation coverage in the village[5] is s_{vt}, a fraction between 0 and 1, inclusive. The key problem under consideration is that the researcher cannot observe s_{vt}, but instead does observe increments in *s* from an initial period *t** which is before dataset. Thus we observe Δs_{vt}, where

$$s_{vt} = s_{vt^*} + \Delta s_{vt}.$$

I'll approach this note in two sections. In the first, I will assume sanitation causes current d_{ivt}, which maps contemporaneously into *h*; this lets me isolate any issues related to the fact that what I observe is *changes* in the stock of sanitation, rather than the stock of sanitation itself. In the second I'll consider further complications due to any time-persistent enteropathy, and whether lagged sanitation matters.

Contemporaneous disease

Let there be some function *d* such that

$$d_{ivt} = d(s_{vt}; \theta_{ivt}),$$

where θ is a vector of properties of the individual. Many people who think about sanitation appear to believe that *d*(·) is non-linear, perhaps convex (so the last toilet is the most helpful) or s-shaped (so intermediate toilets are more helpful). However, there is no good evidence about this either way.

In this section we're imaging that contemporaneous poor sanitation translates into contemporaneous observable health, so $h_{ivt} = f(d_{ivt}; \theta_{ivt})$.

Let's assume that *d*(·) and *f*(·) are both linear functions, and that θ does not interact with the variables of interest. Then we can write

$$d_{ivt} = \alpha_0 + \alpha_1 s_{vt} + \alpha_2 \theta_{ivt} + \varepsilon^d_{ivt}.$$

$$h_{ivt} = \beta_0 + \beta_1 d_{ivt} + \beta_2 \theta_{ivt} + \varepsilon^h_{ivt}.$$

$$h_{ivt} = (\beta_0 + \beta_1\alpha_0) + \beta_1\alpha_1 s_{vt} + (\beta_1\alpha_2 + \beta_2)\theta_{ivt} + \left(\beta_1\varepsilon^d_{ivt} + \varepsilon^h_{ivt}\right).$$

$$h_{ivt} = \gamma_0 + \gamma_1 s_{vt} + \gamma_2 \theta_{ivt} + \varepsilon_{ivt}.$$

So, there is a linear relationship, as well, between h_{ivt} and s_{vt}. Substituting in our data restriction:

$$h_{ivt} = \gamma_0 + \gamma_1(s_{vt^*} + \Delta s_{vt}) + \gamma_2 \theta_{ivt} + \varepsilon_{ivt}.$$

[5] In the actual paper, the relevant geographical unit is the district, not the village, but I used up the "d" symbol on "disease."

$$h_{ivt} = \gamma_0 + \gamma_1 \Delta s_{vt} + \gamma_2 \theta_{ivt} + \gamma_1 s_{vt^*} + \varepsilon_{ivt}.$$

Of course, $\gamma_1 s_{vt^*}$ is exactly a village fixed effect (it has no *t* or *i*), so we can simply estimate

$$h_{ivt} = \gamma_0 + \gamma_1 \Delta s_{vt} + \gamma_2 \theta_{ivt} + \varphi_v + \varepsilon_{ivt},$$

which will allow a consistent estimate of the product $\beta_1 \alpha_1$ despite only having data about sanitation changes.

The two linearity assumptions were crucial here: without them $\gamma_1 s_{vt^*}$ could not have been linearly separated out as a linear fixed effect. Of course, this is generically true of fixed effects (*e.g.* Angrist & Pischke 2009, *p*. 222). So, only observing *new* sanitation is no problem if we assume linearity, which fixed effects implicitly always does.

Persistent enteropathy

"Enteropathy" simply means bowel disease. In the paper I refer to "chronic tropical enteropathy," also called "environmental enteropathy," a changing of the tissue of the inside of the small intestine in such a way as to let in more disease and take in less nutrition. This is caused by inflammation: too many immune-system cells respond to a very high concentration of ingested bacteria.

The significance of the term "chronic" and the point of my discussing it in this paper is that enteropathy could be harming children's health and human capital accumulation even without causing diarrhea. As a recent article in *Nature* summarizes, "the malabsorption associated with environmental enteropathy is often subtle in children, manifesting itself clinically only as stunting due to chronic undernutrition" (Kau, *et al.* 2011). This fact is econometrically important because it means that diarrhea morbidity (survey-reported or otherwise) is not a good proxy for the effects of open defecation on children's health, and perhaps especially not for the effects that economists would be most interested in (development of human capital). However, diarrhea morbidity is considered easy to measure with surveys, and it is the central outcome variable in the econometric literature on water and sanitation.

I do not have data on diarrhea that could be matched with a credible identification strategy for the TSC. So, I used reduced-form health outcomes: death, height, cognition. Environmental enteropathy could explain why I might find effects of sanitation even if some papers studying diarrhea outcomes do not.

Could enteropathy threaten the consistency or identification of estimates in the TSC paper? There are at least two possible ways that enteropathy might matter for my specifications:

1. **The function *f* mapping *d* to *h* is nonlinear.** Earlier I noted that some people think *d* is non-linear: the disease environment is a complicated function of sanitation coverage. Enteropathy suggests that *f* is non-linear, as well: too many germs in the intestine overwhelm a key barrier, leading to a discontinuous immune response. Non-linearity is a problem for fixed effects.
2. ***d* is persistent over time.** Could the tissue changes involved in environmental enteropathy be a persistent disease stock, even when the sanitation environment changes?

Let's focus on better understanding the second problem, since the previous sub-section demonstrated the importance of the first. To begin, it is not clear that environmental enteropathy *is* persistent in this sense. Recall that "chronic" here refers to "not acutely symptomatic," not to "long-lasting." There is

little evidence here, but what there is suggests that recovery is possible after changing the sanitation environment:

- Haghighi, Wolf, and Durie (1997) find that Peace Corps volunteers returning home with what they call "tropical sprue" (our environmental enteropathy) recover in the U.S.
- Rastogi, *et al* (1999) performed detailed medical tests on 47 Indian children with diarrhea, in order to sort them according to their underlying disease. The 46.8% found to have "tropical enteropathy" all "improved with broad spectrum antimicrobial therapy." Unfortunately, this does not reveal what would happen if they were removed to a safe environment but not treated with antibiotics.
- A series of papers by Proos and coauthors find that Indian infants and children adopted to Sweden come with a typical range of poor-country infections, but subsequently experience catch-up growth, although not quite to average healthy levels.

Nevertheless, consider persistence in *d* such that

$$d_{ivt} = \delta d_{ivt-1} + \alpha_1 s_{vt} + \varepsilon^d_{ivt},$$

with $\delta \in [0,1)$ so, with 0 the time of a person's birth,

$$d_{ivt} = \alpha_1 \sum_{\tau=0}^{t} \delta^{t-\tau} s_{v\tau} + \sum_{\tau=0}^{t} \delta^{t-\tau} \varepsilon^d_{iv\tau}.$$

Plugging back in we have

$$h_{ivt} = \beta_0 + \beta_1 \alpha_1 \sum_{\tau=0}^{t} \delta^{t-\tau} s_{v\tau} + \beta_1 \sum_{\tau=0}^{t} \delta^{t-\tau} \varepsilon^d_{iv\tau} + \beta_2 \theta_{ivt} + \varepsilon^h_{ivt}.$$

Assume, as is the case in my data, that $0 > t^*$, so the baby is born during the span of the sanitation data. Then

$$h_{ivt} = \beta_0 + \beta_1 \alpha_1 \sum_{\tau=0}^{t} \delta^{t-\tau} \Delta s_{v\tau} + \beta_1 \alpha_1 \sum_{\tau=0}^{t} \delta^{t-\tau} s_{vt^*} + \beta_1 \sum_{\tau=0}^{t} \delta^{t-\tau} \varepsilon^d_{iv\tau} + \beta_2 \theta_{ivt} + \varepsilon^h_{ivt}$$

This could be problematic in at least two ways, as our independent variable Δs_{vt} is potentially correlated with lagged values of itself as well as s_{vt^*}.

Note, however, that even if this model were correct, there is probably still no problem for the IMR estimates. My data is annual-frequency, and IMR is death within the first year. Environmental enteropathy is acquired through fecal-oral contamination. Babies do not orally consume fecal pathogens before they are born.[6] Therefore, $d_{iv0} = 0$ and the fixed-effects equations for infant mortality correctly assess the probability of surviving the first year as a function of the bowel disease acquired in the first year, which *is the same as* the stock of bowel disease in the first year.

[6] Although there is evidence that mothers' malnutrition can hurt babies *in utero*, I know of no evidence of fecal pathogens introduced to a mother orally making way to a baby's intestines, nor is there a reasonable mechanism by which this would happen. The placenta is a barrier against transmission of many microbes – maternal-fetal transmission of, for example, HIV (a virus much smaller than, for example, *E. coli*) is well documented, but my reading is that congenital infections typically target the fetal bloodstream.

Empirical investigations of the importance of non-linearity and lags

The preceding models suggest that we might worry about the estimates in the TSC paper if we believed that (a) the effects were importantly non-linear, or (b) lagged values of sanitation coverage were important for current levels of health. This section looks for evidence of these possibilities.

Non-linearity

As I report in the paper, in none of the regressions is a quadratic term on my TSC variable significant. For example, in the main IMR regression with the DLHS-3 a quadratic term has a *t* statistic of 0.07, and a quadratic and a cubic term together have a *F* of 0.05.

To make a graph for visual inspection of non-linearity, I did the double-demeaning (year and district) "by hand" in stata (rather than using xtreg) and made an lpoly. Visually, there is little evidence of non-linearity:

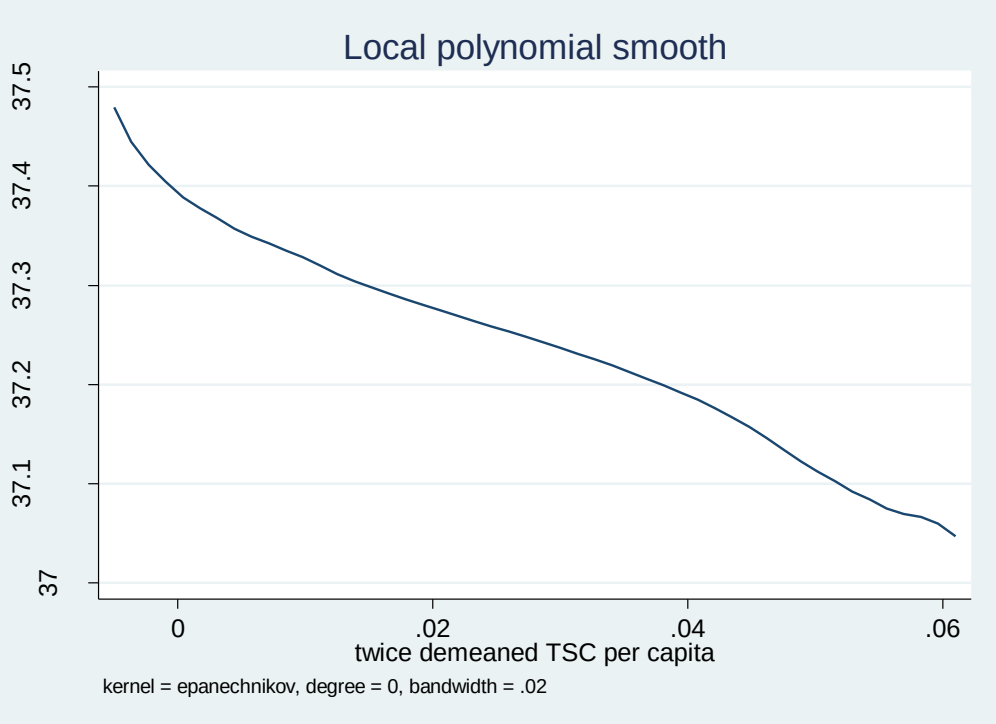


Figure A4: No evidence of much non-linearity

Lag or lead sanitation

The possibility of persistent effects of enteropathy suggests estimating the effects of time-displaced sanitation coverage. This is done in the paper with the displaced independent variables entered separately. Are they statistically significant when entered together? Table A9 adds leads and lags to the main DLHS results:

Table A9: Granger causality type estimates with leads and lags in one equation, DLHS-3

	(1)	(2)	(3)	(4)
dependent variable:	IMR	IMR	PNNMR	PNNMR
TSC at t = 0	-100.2*	-101.6*	-68.03*	-80.13**
	(42.05)	(45.34)	(28.39)	(30.24)
TSC at t = -2 (lagged)	-5.473	-5.101	8.288	11.36
	(57.55)	(57.81)	(32.47)	(32.28)
TSC at t = +2 (lead)		4.634		38.34
		(48.21)		(27.53)
year FEs	✓	✓	✓	✓
district FEs	✓	✓	✓	✓
N	145943	145943	145943	145943

These results give no reason to believe that the model should include lag or lead sanitation. This is consistent with a model of an effect on current health of current levels of sanitation. Note that, if anything, the point estimates are slightly larger than in the TSC paper.

Average effectiveness with errors in productivity reports: proofs

How quantitatively robust are estimates of the program's average effect to errors in bureaucratic reports of latrine construction, if we are interested in knowing the effect not of "reported TSC latrines," but of "real, constructed latrines"? Here, the question is how inflation or errors in the counts of latrines constructed will impact the estimates of the overall average effect of the TSC (that is, IMR down by about 4; height for age up by about one-fifth). Importantly, unlike common concerns about omitted variable bias, none of what is presented here challenges the robust *sign* (- for IMR, + for height) found for the effect of on-site rural sanitation.

When productivity is proportionately inflated: no ultimate problem

This is the case of the incentivized bureaucrats who always and predictably report more than they actually accomplished, a perhaps plausible case in the Indian context.

Let x be actual construction, and $\tilde{x}$ be reported TSC construction. Reports are inflated by a factor $\iota > 0$ (presumably but not necessarily greater than 1), so that $\tilde{x} = \iota x$. [7]

Let x have a true, fixed effect β on some outcome y, such that

$$y = \beta x + \varepsilon,$$

where ε is an error term uncorrelated with x.

If I use OLS to regress y on $\tilde{x}$, I will find, in expectation,

$$\mathrm{E}[\hat{\beta}] = \frac{cov(y, \tilde{x})}{var(\tilde{x})} = \frac{\mathrm{E}[\beta \iota x^2]}{\mathrm{E}[\iota^2 x^2]} = \frac{\beta}{\iota},$$

which will be an attenuated (biased towards 0) estimate in the likely case that $\iota > 1$.

What does this imply for the estimate of the overall effect? This is computed by multiplying $\hat{\beta}$ by $\bar{\tilde{x}}$, awkward notation for the sample mean of $\tilde{x}$.

$$\mathrm{E}[\hat{\beta}\ \bar{\tilde{x}}] = \left(\frac{\beta}{\iota}\right) \iota \mathrm{E}[x] = \beta \mu_x,$$

where μ_x is the expectation of the true mean of x. Notice that ι cancels, and we are left with the correct answer: $\beta \mu_x$.

Therefore, if productivity is inflated in this way, the coefficient estimate is attenuated, but the estimate of the overall effect of the program at its mean remains correct.

[7] I could have included an additive constant here, such that $\tilde{x} = \iota_0 + \iota_1 x$, but this would have been immediately differenced out by the fixed effects used in the paper.

When productivity contains random additive errors: more complicated

A case of more complicated errors in the administrative data produces a more complicated result. Now let

$$\tilde{x} = x + \iota$$

and, as before,

$$y = \beta x + \varepsilon,$$

with $\sigma_{x\varepsilon} = 0$. Then we will have

$$\begin{aligned} \mathrm{E}[\hat{\beta}] &= \frac{cov(x+\iota, \beta x + \varepsilon)}{var(x+\iota)} \\ &= \frac{\beta\sigma_{xx} + \beta\sigma_{x\iota} + \sigma_{\iota\varepsilon}}{\sigma_{xx} + \sigma_{\iota\iota} - 2\sigma_{x\iota}} \\ &= \beta\left(\frac{\sigma_{xx} + \sigma_{x\iota}}{\sigma_{xx} + \sigma_{\iota\iota} - 2\sigma_{x\iota}}\right) + \frac{\sigma_{\iota\varepsilon}}{\sigma_{xx} + \sigma_{\iota\iota} - 2\sigma_{x\iota}}. \end{aligned}$$

Classical measurement error. If ι is merely classical measurement error, mean zero and uncorrelated with anything else, then this reduces to the standard case where the estimate is attenuated due to $\sigma_{\iota\iota} > 0$. It is not implausible that Indian government record-keeping adds noise. Alok (2010) describes administrative assistants entering data by hand from paper forms to Excel spreadsheets, and details the case of one data entry operator in the central government office who was uncomfortable with the online software and insisted on entering data first from paper forms and then to the internet-based administrative records. I have visited the central government office responsible for this data and – while I have quite a bit of respect for the large size of their administrative undertaking and accomplishment relative to the small number people who appeared to be responsible for it – I noted that, for example, it took an employee some time to find a spreadsheet on his computer that his boss had asked him to show me.

If classical measurement error is the case, because $\mathrm{E}[\tilde{x}] = \mathrm{E}[x]$, the estimate of the overall impact of the program would be too small.

Corrupt inflation. Perhaps, instead, ι represents corruption, and has a positive mean: bureaucrats additively inflate the productivity in the districts under their supervision in administrative records. Sorting out the implications of this includes making further assumptions. In this case of low state capacity and corruption, I find it plausible to assume:

- $\sigma_{x\iota} < 0$: District-years that actually built more latrines will additively inflate their construction count by less. This could be the case if there are two "types" of district administration (those that try to do good and those that try to look good) or if there are institutional or psychological target counts of reported latrines, that administrators fulfill in fact or in report.
- $\sigma_{\iota\varepsilon} < 0$: ε is a measure of all of the factors unobserved and unrelated to the Total Sanitation Campaign that impact children's health. Much of the omitted variable bias in econometrics comes from the fact that good things tend to go together and bad things tend to go together. Along these lines, this assumption is tantamount to believing that district-years in which the

unobserved factors that promote children's health are greater are also those in which corrupt record-keeping is lesser, on average.

If both of these assumptions are correct, then $\mathrm{E}[\hat{\beta}] < \beta$, and my coefficient estimates are underestimates of the true effects of the TSC. However, if it is also the case that $\mathrm{E}[\tilde{x}] > \mathrm{E}[x]$ (and it is certainly possible for ι to have these correlations with ε and x with or without having a positive mean), then it is not possible to sign the bias of their product, and therefore in the estimates of the effect at the mean.

References cited only in supplementary appendix

Almond, Douglas, *et al*. (2010). "Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns." *Quarterly Journal of Economics,* (May).

Bathran, Ravichandaran. (2011). "Indian Sanitation." *Economic and Political Weekly*, *XLVI*(51: December 17), 34-37.

Cialdini, Robert B. (2003). "Crafting normative messages to protect the environment." *Current Directions in Psychological Science, 12*, 105-109.

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

Duflo, Esther. (2003). "Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa." *World Bank Economic Review, 17*(1), 1-25.

Frankenberg, Elizabeth, Wayan Suriastini, & Duncan Thomas. (2005). "Can expanding access to basic healthcare improve children's health status? Lessons from Indonesia's 'midwife in the village' programme." *Population Studies*, *59*(1), 5-19.

Jayachandran, Seema. (2009). "Air quality and early life mortality: Evidence from Indonesia's wildfires." *Journal of Human Resources*, *44* (4), 916-954.

Kau, Andrew L., *et al*. (2011). "Human nutrition, the gut microbiome and the immune system." *Nature*, 474, 327–336.