Essays in the Economics of Sanitation and Human Capital in Developing Countries

Dean Spears

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Economics
Adviser: Anne C. Case

June 2013

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

Abstract

This thesis is a collection of three essays in the empirical economics of sanitation, open defecation, and child well-being in developing countries.

Physical height is an important economic variable reflecting health and human capital. Puzzlingly, however, differences in average height across developing countries are not well explained by differences in wealth. In particular, children in India are shorter, on average, than children in Africa who are poorer, on average. This paradox has been called "the Asian enigma" and has received much attention from economists. Chapter one provides the first documentation of a quantitatively important international gradient between child height and sanitation. I apply three complementary empirical strategies to identify the association between sanitation and child height: country-level regressions across 140 country-years in 65 developing countries; within-country analysis of differences over time within Indian districts; and econometric decomposition of the India-Africa height difference in child-level data. Open defecation, which is exceptionally widespread in India, can account for much or all of the excess stunting in India.

Chapter two studies the Indian government's Total Sanitation Campaign, which offered local government agents a large *ex post* monetary incentive to eliminate open defecation. I use two strategies to estimate the program's effect on children's health: first, heterogeneity in the timing of program implementation across districts, and second, a discontinuity in the monetary incentive to village governments. On average, the program caused a decrease in infant mortality and an increase in children's height. Importantly, this paper studies a full-scale program implemented by a large bureaucracy with limited capacity. In the context of governance constraints, incentivizing local government agents can be effective.

Chapter three is coauthored with Jeffrey Hammer. We study a randomized controlled trial of a village-level sanitation program, implemented in one district by the government of Maharashtra. The program caused a large but plausible average increase in child height. Unusually, the original World Bank evaluation team also collected data in districts where

the government planned but ultimately did not conduct an experiment, permitting us to analyze how the set of villages eligible for randomization into the treatment group might shape research findings.

Acknowledgements

I have received so much help, and I have more gratitude than I can accommodate here. Teachers and friends have invested in teaching me first about big picture political economy of development; then about choice theory, lab experiments, and the behavioral economics of poverty, cognitive control, and apparent impatience (including a year with a cohort of psychology students); and finally about the economics of health and early-life human capital accumulation in developing countries. Thank you all.

Each chapter involves its own debts. For chapter one, I am grateful for helpful suggestions from and discussion with Harold Alderman, Luis Andres, Radu Ban, Robert Chambers, Juan Costain, Jean Drèze, Marianne Fay, Ariel Fiszbein, Lawrence Haddad, Derek Headey, Peter Lanjouw, J.V. Meenakshi, Doug Miller, John Newman; from seminar participants at Princeton, the Delhi School of Economics, Tufts, IFPRI, the Public Health Foundation of India, LBSNAA, a Gates Foundation workshop in Cambodia, and the World Bank (in DC, Delhi, and Cambodia); and from its discussion on the blogs Development Horizons (by Lawrence Haddad), Development Impact (at the World Bank and thanks to Jed Friedman), and Ideas for India.

For chapter two, I appreciate conversations with staff at India's Ministry of Drinking Water & Sanitation, UNICEF, and the World Bank WSP and with district, block, and village officials. Thanks to Farzana Afridi, Radu Ban, Abhijit Banerjee, Robert Chambers, Tom Classen, Oliver Cumming, Janet Currie, Taryn Dinkelman, Josephine Duh, Sebastian Galiani, Jacob Goldin, Raymond Guiteras, Susanna Makela, Doug Miller, Mushfiq Mobarak, Abhiroop Mukhopadhyay, John Papp, Manisha Shah, Rohini Somanathan, Yaniv Stopnitzky, Tom Vogl, Alix Zwane; to seminar participants at NEUDC, the Gates Foundation, ISI Delhi, LBSNAA, NCAER's India Policy Forum, Princeton, and the World Bank. Sneha Lamba provided able research assistance.

Chapter three is coauthored with Jeffrey Hammer. We acknowledge with thanks financial support for this project from the World Bank Water and Sanitation Program. We are

grateful for and applaud the vision of the leaders in the Maharashtra government who championed this project. We appreciate helpful comments from Luis Andres, Bertha Briceño, Juan Costain, Oliver Cumming, Paul Gertler, Soma Ghosh Moulik, Rinku Murgai, and seminar participants at the World Bank in Washington, DC and in Delhi and at World Water Week in Stockholm. The views and conclusions presented in the chapter are our own.

I am grateful for the financial support of the economics department. I have benefited from Princeton's fellowship in social policy, and thank Katherine Newman, Sara McLanahan, Devah Pager, and Nancy Turco and my "JDP" classmates and teachers for the many benefits of that program. The specific research in this dissertation was supported by the Center for Health and Wellbeing's program on grand challenges of infectious disease, and I thank Janet Currie, Kristina Graff, and Susan Rizzo. I have been glad to be part of the Research Program for Development Studies, and have much for which to thank Lillian Anderson. At Princeton, I have also received research funding from the Center for Economic Policy Studies, the Princeton Laboratory in Experimental Social Science, and the Global Network on Inequality.

Friends in India have made this work possible and fun, and I especially thank those at the Delhi School of Economics: Abhijit Banerji, Ashwini Deshpande, Deepti Goel, and for all sorts of help Surjeet Singh. Reetika Khera has been a friend who made it her project to make us at home – often including in her own home – and set much of what enables our work In India in motion.

I have learned much from my friends at Princeton, including Josephine Duh, Jacob Goldin, Avinash Kishore, Nicholas Lawson, John Papp, and Abby Sussman. Mike Geruso deserves special thanks for generous and ongoing tutoring.

Many teachers at Princeton have taught me so much (and please see the dedication on the next page). Here I want to especially thank Angus Deaton, Jeff Hammer, and Sam Schulhofer-Wohl. They also taught me about being an economist, and some of my favorite

memories of Princeton involve learning from this crowd at Monday and Wednesday lunches. In an academic discipline not known for its friendliness, I think the third floor of Wallace stands out as a supportive place.

Anne Case is my adviser and my teacher. I realized a long time ago that many economists are knowledgeable and skilled, but Anne is also wise. I mentally carry her work as a model of doing so much with such clarity and simplicity. Over the years, she has accepted many duties and remained a committed teacher. Anne profoundly cares about her students. I am thankful that she agreed to be my adviser, and for so much since. She gave me the greatest treasure, for which I am grateful every day: the training which gives the opportunity to do this work. Thank you, Anne.

I am lucky to have such loving parents Donald and Janet Spears, and am grateful, too for their support and teaching. My dad was my first economics teacher. No acknowledgment is adequate for Diane Coffey. She hears about every regression first and last, and has read every word many times, including the ones that are not here anymore. We have sincere difficulty sorting out which ideas were whose when this is required; we work so much together now that it may be silly to try. I don't know what this dissertation would be about if not for Diane's being in my life, but it would lack much of whatever good is here. I look forward to the collaborations to come.

Finally, quite a lot of people had their height measured, and had their babies' length measured, and had their days interrupted, and answered personal questions, and tried to remember who was born exactly when, all so that we could compute these statistics. I appreciate it.

For the children I write about, and to my teachers, in deep gratitude: Anne Case and R.B., C.J.C., J.C., J.D., A.D., T.D., S.F., C.K., J.G., K.G., J.G., F.G., J.H., A.H., J.L., S.M., U.M., C.P., A.-J.P., K.R., S.S.-W., N.T.-B. & T.V.

Contents

List of Tables

List of Figures

Chapter 1

How much international variation in child height can sanitation explain?

1.1 Introduction

Physical height is a topic of expanding interest to economists (Steckel, 2009), in large part because it is an important correlate of human capital and health and is a predictor of economic productivity (Currie, 2009). Despite this attention, an important puzzle persists: international differences in height across present day developing countries are not well explained by differences in economic well-being (Deaton, 2007). In particular, people in India are shorter, on average, than people in Africa, despite the fact that Indians are also richer, on average, a fact that has been labeled the "Asian enigma" (Ramalingaswami et al., 1996).

One candidate explanation which has received relatively little attention in economists' recent investigations of the puzzle of Indian stunting (*e.g.* Deaton, 2007; Tarozzi, 2008; Jayachandran and Pande, 2012; Panagariya, 2012) is sanitation. Medical research documents that chronic childhood environmental exposure to fecal germs can be an important cause of stunting (Humphrey, 2009). Sanitation coverage is exceptionally poor in India, where over

half of households defecate openly without using a toilet or latrine, a much larger fraction than in other countries with similar income.

According to joint UNICEF and WHO (2012) estimates for 2010, 15 percent of people in the world, and 19 percent of people in developing countries, openly defecate without using any toilet or latrine. The primary contribution of this chapter is to document that much of the variation in child height among developing countries can be explained by differences in rates of open defecation. Sanitation robustly explains variation in stunting, even after accounting for GDP and other dimensions of heterogeneous economic development. The two other chapters in this dissertation, concentrating on internal validity, have documented the existence of a causal effect of sanitation on child height; in contrast, this chapter assesses the global statistical importance of sanitation quantitatively, using descriptive regressions and econometric decomposition techniques. In particular, differences in open defecation are sufficient to statistically explain much or all of the difference in average height between Indian and African children. These results suggest that open defecation is a policy priority of first-order importance.

This chapter makes several contribution to the literature. First, to my knowledge, it offers the first documentation of a quantitatively important cross-country gradient between sanitation and child human capital. Although the association between income and health has been widely studied within and across developing countries, the importance of sanitation has received much less attention. Moreover, I show that sanitation predicts child height even conditional on income. Controlling for GDP, the difference between Nigeria's 26 percent open defecation rate and India's 55 percent is associated with an increase in child height approximately equivalent to quadrupling GDP per capita.

Second, this chapter documents an interaction between sanitation and population density, consistent with a mechanism in which open defecation harms human capital through exposure to environmental germs. The number of people defecating openly per square kilometer linearly explains 65 percent of international variation in child height. This finding

clarifies the policy case for sanitation as a public good. Third, it contributes to a resolution of the puzzle of the "Asian enigma" of Indian stunting, which has received much recent attention from economists. Finally, the conclusions offer a reminder that height, often refereed to as an indicator of "malnutrition," broadly reflects early-life *net* nutrition, including losses due to disease.

Three sections of the chapter contribute complementary analyses of the relationship between height and open defecation, each focusing on a different dimension of heterogeneity. Section 1.2 studies country-year average sanitation and child heights; here, each observation is a collapsed DHS survey. Open defecation is particularly harmful to children's health where population density is high, creating a special risk of stunting in India. Section 1.3 compares children within India, introducing district fixed effects to repeated cross-section data constructed out of two rounds of India's National Family and Health Survey, in order to study differences within districts over time. Section 1.4 considers whether the India-Africa height gap can be explained by heterogeneity in village-level open defecation rates, using individual-level data on child heights and decomposition analysis in the spirit of Oaxaca-Blinder. Reweighting Indian data to match the sanitation of an African sample counterfactually increases the height of Indian children by more than the India-Africa gap. All three approaches find a similar and quantitatively important association between height and sanitation. Finally, a concluding section 1.5 considers whether estimates of of the association between height and sanitation in this chapter and from the literature are sufficient to account for the India-Africa gap.

1.1.1 Open defecation causes stunting

A growing literature in economics documents that physical height has its origins in early life health (*e.g.* Case and Paxson, 2008), especially in poor countries where environmental threats to health are more important than they are in rich countries, relative to genetics (Martorell et al., 1977; Spears, 2012). Two existing literatures indicate that early-life

exposure to fecal germs in the environment reduces children's subsequent height. First, medical and epidemiological literatures are accumulating persuasive documentation of the mechanisms linking open defecation to poor health and early life human capital accumulation.[1] Humphrey (2009) presents evidence that chronic but subclinical "environmental enteropathy" – a disorder caused by repeated fecal contamination which increases the small intestine's permeability to pathogens while reducing nutrient absorption – could cause malnutrition, stunting, and cognitive deficits, even without necessarily manifesting as diarrhea (*see also* Petri et al., 2008; Mondal et al., 2011; Korpe and Petri, 2012). Relatedly, Checkley et al. (2008) use detailed longitudinal data to study an association between childhood diarrhea and subsequent height.

Second, chapter two of this dissertation finds an effect of a government sanitation program in rural India. Similarly, chapter three reports a randomized field experiment in Maharashtra, in which children living in villages randomly assigned to a treatment group that received sanitation motivation and subsidized latrine construction grew taller than children in control villages. In subsequent work that acknowledges the priority of the chapters of this dissertation, Kov et al. (2013) document that increases in sanitation coverage over time in Cambodia are associated with increases in child height. Section 1.5.1 considers the estimates of these causally well-identified studies in the context of this chapter's results.

1.1.2 Open defecation is common in India

Of the 1.1 billion people who defecate in the open, nearly 60 percent live in India, which means they make up more than half of the population of India (Joint Monitoring Programme for Water Supply and Sanitation, 2012). These large numbers are roughly corrob-

[1]Perhaps the recent paper most complementary to this one is Fink et al.'s (2011) regression of an indicator for child stunting on variables including sanitation in 172 pooled DHS surveys. However, one key difference is that they focus on within-country height-sanitation correlations: all regressions include DHS survey fixed effects for country-years. In economics, Headey (2013) studies the association between child height and sanitation across countries in a recent paper, but focuses on the fraction with flush toilets, overlooking that pit latrines can dispose of feces safely.

orated by the Indian government's 2011 census, which found that 53.1 percent of all Indian households – and 69.3 percent of rural households – "usually" do not use any kind of toilet or latrine. In the 2005-6 National Family Health Survey, India's Demographic and Health Survey, 55.3 percent of all Indian households reported defecating openly, a number which rose to 74 percent among rural households.

These statistics are striking for several reasons. First, open defecation is much more common in India than it is in many countries in Africa where, on average, poorer people live. UNICEF and the WHO estimate that in 2010, 25 percent of people in sub-Saharan Africa openly defecated. In the largest three sub-Saharan countries – Nigeria, Ethiopia, and the Democratic Republic of the Congo – in their most recent DHS surveys, 31.1, 38.3, and 12.1 percent of households report defecating openly.

Second, despite GDP growth in India, open defecation has not rapidly declined in India over the past two decades. In the DHS, where 55.3 percent of Indian households defecated openly in 2005-06, 63.7 did in the earlier 1998 survey round, and 69.7 did in 1992. This is particularly true for poor people: the joint UNICEF and WHO report concludes that "the poorest 40 percent of the population in Southern Asia have barely benefited from improvements in sanitation." In 2010, 86 percent of the poorest quintile of South Asians defecated openly.

Of course, open defecation, even if very important, is certainly not the only factor shaping child height. This paper complements other recent research documenting the effects of social inequality on child health and human capital in India, especially by gender and including within households (Jayachandran and Pande, 2012). For example, Coffey et al. (2013) find that in rural joint Indian households, children of older brothers are taller, on average, than children of younger brothers, despite sharing a household environment. They interpret this as partially a reflection of the intrahousehold social status of the children's mothers. All of these mechanisms could be – and likely are – simultaneously active: even

children at the top of social divides are almost certainly shorter, on average, than they otherwise would be, due to the open defecation of their neighbors.

Therefore, it is already well-known that open defecation is bad for children's health; that early-life disease leads to lasting stunting; and that open defecation is exceptionally widespread in India. The contribution of this paper is to quantitatively assess the importance of the link among these facts. The results indicate that sanitation is a statistically important predictor of differences in the heights of children across developing countries and can explain differences of interest to economists and of significance to human development.

1.2 Evidence from country means: 140 DHS surveys

Across countries, observed in different years, how much of the variation in child height is explained by variation in open defecation? This section uses 140 DHS surveys, each collapsed into a single observation, to show that sanitation alone explains more than half of the variation across country-years.

The analysis proceeds in several steps. First, section 1.2.2 documents that, across country means, height is associated with open defecation, with little change after controlling for GDP. The next part of this section uses country fixed effects and replication on sub-samples of world regions to show that no geographic or genetic differences are responsible for the result. Then, this section verifies that other dimensions of infrastructure or well-being do not similarly predict child height. The following part observes that children would be more exposed to fecal pathogens where population is more dense, and finds that open defecation interacts with population density. The last part of this section documents that the association between height and open defecation is steeper among older children, consistent with an unfolding effect of accumulating exposure. Finally, section 1.2.3 considers the average height difference between children in South Asian and Sub-Saharan African countries, and shows that much of this gap is accounted for by sanitation.

1.2.1 Data

All data used in this paper are publicly available free of charge on the internet. Demographic and Health Surveys (DHS) are large, nationally representative surveys conducted in poor and middle-income countries. DHS surveys are collected and organized to be internationally comparable. In some countries, several rounds of DHS data have been collected; in others only one or two. I use every DHS survey which recorded household sanitation and measured child height.[2] This creates a maximum sample of 140 country-years and 65 countries, ranging in frequency from 26 countries that appear in the survey once to 10 that appear in four separate DHS surveys. The earliest survey in the dataset was collected in Pakistan in 1990; the most recent are from 2010.

I match data from other sources to the collapsed DHS surveys. GDP per capita and population are taken from the Penn World Tables. "Polity" and "Democracy" scores of democratization are taken from the Polity IV database. A measure of calorie availability produced by the World Food Program is used in some specifications. All other variables are from DHS surveys.

Using these data, the basic regression I estimate is

$$height_{cy} = \beta open\ \ defecation_{cy} + \alpha_c + \gamma_y + X_{cy}\theta + \varepsilon_{cy}, \tag{1.1}$$

where observations are collapsed DHS surveys, c indexes countries, and y indexes years. *Open defecation* is a fraction from 0 to 1 of the population reporting open defecation without using a toilet or latrine.[3] *Height* is the average height of children under 5 or children under 3, used in separate regressions. As robustness checks, results are replicated with country

[2] I use published summary statistics available online at www.measuredhs.com. DHS surveys do not include rich countries, such as the U.S. One important omission is China, where there has not been a DHS survey; see section 1.2.2.

[3] For example, in India's NFHS-3, the survey asks "What kind of toilet facility do members of your household usually use," with the relevant answer "No facility/uses open space or field." This importantly distinguishes latrine use from latrine ownership.

fixed effects α_c, year fixed effects γ_y, and time-varying controls X_{cy}, including the log of GDP per capita, all of which are added separately in stages. Standard errors are clustered by 65 countries.

1.2.2 Regression results

Figure 1.1 depicts the main result of this section: a negative association between open defecation and child height is visible across country years both for children under 3 and children under 5. Regression lines are plotted with and without weighting by country population. The three largest circles are India's National Family and Health Surveys (only one large circle appears in panel (b) because only the 2005 survey measured height of children up to age 5). Average height in India is, indeed, low. However, the fact that the Indian observations are on the regression line – and not special outliers – is an initial suggestion that sanitation might help resolve the "Asian enigma" of Indian height.

Table 1.1 reports estimates of regression 1.1 and will be referenced throughout this section of the chapter. The main estimate of a linear decrease in height of 1.24 standard deviations associated with changing the fraction openly defecating from 0 to 1 is qualitatively similar to chapter 2's estimates of 1.15 to 1.59, where effects of sanitation are identified using heterogeneity in the implementation of an Indian government program. In column 1, sanitation alone linearly explains 54 percent of the country-year variation of children's height in DHS surveys. Because sanitation and height are both improving over time, column 2 adds year fixed effects; the point estimates slightly increase and standard errors decrease, suggesting that the result is not an artifact of time trends.

Does the significance of open defecation merely reflect general economic development? Column 3 adds a control for GDP per capita; the coefficient on sanitation remains similar, which is consistent with Deaton's (2007) observation that income does not explain cross-

country height differences.[4] This is reflected in panel (a) of figure 1.2, which plots the residuals after regressing height of children under 3 years old on the log of GDP against the residuals after regressing open defecation on the log of GDP. The association remains, and the R^2 is similar: sanitation linearly explains 54.2 percent of variation in child height, and the sanitation residual explains 53.9 percent of the variation in the height residual.

Panel (b) of figure 1.2 adds average height and exposure to open defecation in wealth subsets of India's 2005 DHS to the basic plot of country mean height and sanitation. Included in published DHS data is a classification of households into wealth quintiles based on asset ownership. Average height of children within these groups is plotted against the rate of open defecation among all households in the primary sampling unit where they live, that is, the local open defecation to which they are exposed. Additionally, I follow Tarozzi (2008) in identifying an elite top 2.5 percent of the Indian population: children who live in urban homes with flush toilets that they do not share with other households; whose mothers are literate and have been to secondary school; and whose families have electricity, a radio, a refrigerator, and a motorcycle or car. Even these relatively rich children are shorter than healthy norms; this is expected, because 7 percent of the households living near even these rich children defecate openly. Indeed, the graph shows that their stunted height is approximately what would be predicted given the open defecation in their environment. More broadly, the association between height and sanitation among these wealth groups is close to the the international trend computed from country means. Exposure to nearby open defecation linearly explains 99.5 percent of the variation in child height across the five asset quintiles.

[4]GDP per capita statistically significantly interacts with open defecation to predict height: $\widehat{height}_{cy} = -1.42 - 1.18\, open\ \ defecation_{cy} + 0.14 \ln(GDP)_{cy} - 0.59\, open\ \ defecation_{cy} \times \ln(GDP)_{cy}$, where open defecation and GDP are demeaned, the coefficients on open defecation and the interaction are statistically significant at the 0.01 level, and GDP is statistically significant at the 0.05 level. Thus, the slope on ln(GDP) would be 0.36 with no open defecation, but only 0.038 at India's 2005 level of open defecation, consistent with the low apparent effect of recent Indian economic growth on stunting. Although it is difficult to interpret this result causally, one possibility is that private health inputs such as food do less to promote child height in a very threatening disease environment; I thank Angus Deaton for this suggestion.

A geographic or genetic artifact?

Perhaps people who live in certain countries or regions tend to be tall or short, and this is coincidentally correlated with open defecation. Is the result driven by certain countries or regions, or fixed differences such as genetics?

Figure 1.3 presents initial evidence against this possibility. The sample is restricted to countries with more than one DHS observation, and country means across collapsed DHS surveys are subtracted from the height and sanitation survey averages. The figure plots the difference in a country-year's height from that country's mean across DHS surveys against the difference in sanitation. The slope is similar to the undifferenced plot. Moreover, panel (b) continues to demonstrate an association despite not including any data from India.

Returning to table 1.1, column 4 adds country fixed effects.[5] A control is also added for the average height of mothers of measured children; this is in response to a possibility observed by Deaton and Drèze (2009) that Indian stunting is not caused by current nutritional deprivation or sanitary conditions, but is instead an effect of historical conditions that stunted the growth of women who are now mothers, restricting children's uterine growth. DHS surveys are categorized into six global regions;[6] column 5 adds six region-specific linear time trends $\sum_r \delta_r year_y$, to rule out that the effect is driven by spurious changes in specific parts of the world. Neither of these additions importantly change the estimate of the coefficient, although adding so many controls increases the standard errors.

Table 1.2 further confirms that no one region is responsible for the results. The association between height and sanitation is replicated in regressions that omit each of the six world regions in turn. The coefficient near 1 notably remains when South Asian observations are omitted, again suggesting that the result is not merely reflecting India.

[5]Computing estimates using first-differencing rather than fixed effects finds very similar results; I thank Luis Andres and Derek Headey for this suggestion.

[6]The regions are sub-Saharan Africa, South Asia, the Middle East & North Africa, Central Asia, East & Southeast Asia, and Latin America.

A statistical coincidence?: Omitted and placebo variables

Across rich and poor places, good conditions are often found together, and problems are often found in places with other problems. Would *any* measure of infrastructure, governance, or welfare be as correlated with height as is sanitation?

Column 6 of table 1.1 adds time-varying controls. These include female literacy, which is an important predictor of child welfare, and accessibility of water supply, all also from the DHS.[7] Development outcomes are often attributed to good "institutions;" the set of controls includes the polity and autocracy scores from the Polity IV database. With these controls added, the association between height and sanitation is essentially unchanged.[8]

Table 1.3 isolates each of these alternative independent variables in turn. The top part of the table, Panel A, shows that many of these variables are indeed correlated with child height, as would be expected. For example, children are about one standard deviation taller in places with 100 percent full immunization than in places where no children are fully immunized.[9] However, as shown in panel B, none of these "placebo" predictors matter for child height, conditional on sanitation and GDP. In particular, conditional on sanitation and GDP, child height is not associated with other types of infrastructure (electrification, water), democratic governance, female literacy, full immunization, or nutritional measures such as food availability, or the breastfeeding rate[10].

[7]Despite the frequency of undifferentiated references to "water and sanitation," improving water supply and reducing open defecation have very different effects on child health and other outcomes and should not be conflated (Black and Fawcett, 2008).

[8]If controls for the fraction of infants ever breastfed, the fraction of infants breastfed within the first day, and the fraction of infants fed "other liquids" in the past 24 hours are further added as measures of the quality of infant nutrition, the coefficient on open defecation in panel A with all controls becomes *larger* in absolute value, -1.59, with a standard error of 0.82. This is not statistically significantly larger than the other estimates.

[9]I thank J.V. Meenakshi for suggesting this control.

[10]Breastfeeding is an especially important variable because India has high levels of open defecation, short children, and poor breastfeeding. In a regression with open defecation, country fixed effects, log of GDP, the fraction of children ever breastfed, and the fraction of children breastfed on the first day, neither breastfeeding variable is statistically significant (t of 0.23 and 0.6 respectively), but open defecation has a similar coefficient of -0.849, statistically significant at the two-sided 0.1 level ($n = 139$). If open defecators per square kilometer is used in place of the open defecation fraction, again it is statistically significant ($t = -6$) and the breastfeeding variables are not.

China: An out-of-sample prediction

No DHS survey has been conducted in China. Therefore, Chinese data is not used in estimating any of the regression equations in this paper. However, Liu et al. (2012) report a height z-score estimate of -0.79 for Chinese children, using the China Health and Nutrition Survey and averaging over the years 1989 through 2006. According to WHO-UNICEF joint monitoring program statistics, over this time period about 4 percent of the Chinese population practiced open defecation. Therefore, relative to other developing countries such as India, Chinese children are exposed to relatively little open defecation, and are relatively tall.

How does this estimate compare with the prediction of the simple linear model in figure 1.1 or in column 1 of table 1.1? The model predicts an average Chinese height-for-age of -0.98. This absolute residual of 0.19 would rank 43rd among the 140 absolute residuals from the model, suggesting that the model accurately matches the data with this out-of-sample prediction. If this one observation from China is added to the model, the coefficient *increases* in absolute value from -1.24 to -1.37, and the t-statistic from -13 to -16.

Mechanism: Interaction with population density

If open defecation is, indeed, stunting children's growth by causing chronic enteric infection, then height outcomes should be consistent with this mechanism. In particular, children who are more likely to be exposed to other people's fecal pathogens due to higher population density should suffer from larger effects of open defecation. For example, Ali et al. (2002) show that higher population density is associated with greater cholera risk in a rural area of Bangladesh, and in chapter 2 I find a greater effect of India's Total Sanitation Campaign in districts with higher population density.

To test this conjecture, I construct a crude measure of "open defecators per square kilometer": the product of population density per square kilometer times the fraction of people reporting open defecation. Figure 1.4 reveals that this measure of exposure to fecal

pathogens (in logs, due to wide variation in population density) visibly predicts average child height. The regression in panel (a) of the figure explains 65 percent of variation in child height. Notably, India occupies the bottom-right corner of the graph, with high rates of open defecation and very high population density.

Does population density add predictive power beyond open defecation alone? The final column of table 1.1 adds an interaction between open defecation and population density. The interaction term is statistically significant, and the interaction and population density are jointly significant with, an $F_{2,63}$ statistic of 4.5 ($p = 0.0149$) in column (a) and $F_{2,57}$ statistic of 4.5 ($p = 0.0155$) in column (b).

A further implication of this mechanism is that open defecation will have a steeper association with child height in urban places than in rural places (Bateman and Smith, 1991; Bateman et al., 1993). Table 1.4 investigates this using two additional collapsed datasets, one containing only the urban observations in each DHS survey and one containing only the rural observations. Although GDP per capita is not available for urban and rural parts of countries, urban and rural women's height controls can similarly be computed from the DHS. In all cases, the urban coefficient on open defecation is greater than the whole-country coefficient and the rural coefficient is smaller. Hausman tests (reported under the open defecation coefficients in columns 5 through 7) verify that urban coefficients are larger than rural coefficients from the corresponding specifications.

Mechanism: A gradient that steepens with age

Height-for-age z scores are computed by age-in-months so that, in principle, the heights of children of different ages can be pooled and compared. If international reference charts were genetically or otherwise inappropriate for some countries, we might expect a consistent gap across children of different ages, analogous to a country fixed effect. However, stunting in India and elsewhere develops over time: children's mean z-scores fall relative to the norm until about 24 months of age, where they flatten out. This is consistent with early-life

health deprivation causing a steepening "gradient" between health and economic status, more steeply negatively sloped as children age (Case et al., 2002). If the association between height and sanitation were indeed the unfolding result of accumulating exposure to fecal pathogens, then it is plausible that the association would become steeper over the first two years of life, at a rate that flattens out.

Figure 1.5 plots the coefficients from estimating the basic equation 1.1 separately for collapsed means of children in four age groups: 0-5 months, 6-11 months, 12-23 months, and 24-35 months. Thus, as in the rest of this section of the paper, each coefficient is computed in a regression with 140 country means, but now these height means only include children in subsets of the age range. The independent variable – country-wide open defecation – is the same in each regression.

Two conclusions are visible in the figure. First, the gradient indeed steepens in age, at a rate that flattens. Second, the mean height of Indian children in the 2005 NFHS-3 is plotted for reference. The curve has a similar shape to the age pattern of the coefficients. This suggests that a fixed exposure to open defecation could be scaled into a similar shape as the Indian height deficit by an increasing association between sanitation and height.

1.2.3 The gap between South Asia and sub-Saharan Africa

Although people in South Asia are, on average, richer than people in sub-Saharan Africa, children in South Asia are shorter, on average, and open defecation is much more common there. How much of the South Asia–Africa gap can sanitation statistically explain, at the level of country averages?

Table 1.5 estimates regressions in the form of equation 1.1, with the sample restricted to countries in South Asia and sub-Saharan Africa and with an indicator variable added for data from South Asia. Of the 140 DHS surveys in figure 1.1, 11 are from South Asia

and 78 are from sub-Saharan Africa. In these data, children in South Asia are, on average, about one-third of a height-for-age standard deviation shorter.[11]

How do further controls change the estimate of this South Asia indicator? Merely linearly controlling for open defecation reduces the gap by 30 percent from 0.360 to 0.253. Controlling, instead, for the number of people openly defecating per square kilometer (the product of population density and the open defecation rate, column 4) reduces the coefficient by 83 percent to 0.061. Column 5 verifies that this result is not merely a misleading effect of population density, controlling for which increases the gap.

Pairs of columns 6-7 and 8-9 demonstrate the statistical robustness of the explanatory power of the density of open defecation. After controlling for the log of GDP per capita, adding a further control for open defecators per square kilometer explains 73 percent of the (larger) remaining gap. The density of open defecation reduces by 92 percent the height gap after controlling for both log of GDP and year fixed effects. Sanitation initially appears to explain much of the Africa-South Asia gap in child height. Section 1.4.3 considers the decomposition of this difference in more detail, using child-level height data.

1.3 Evidence from differences within Indian districts

How much of the change over time in Indian children's height is accounted for by the increase over time in sanitation coverage? One challenge to answering this question well is that, unfortunately, improvements in height and sanitation in India have both been slow.

[11] Jayachandran and Pande (2012), using individual-level DHS data from Africa and South Asia, suggest that first-born South Asian children are *taller* than first-born African children. Although it is almost certainly true that first-born Indian children are advantaged, I do not happen to see this particular reversal in the country-level dataset studied here. If country means are computed using only first-born children, I find that South Asian children are 0.22 standard deviations shorter ($s.e. = 0.05$), a reduction but not an elimination of the 0.36 gap in table 1.5. In this sample of country means of first-borns, the gap falls to 0.15 with a control for open defecation and to 0.08 with a control for open defecation per square kilometer. In the full sample of country-level means of first-borns, analogously to column 1 of table 1.1, moving from an open defecation rate of 0 to 1 is linearly associated with a decline in height for age of 1.11 standard deviations. Importantly, however, it can simultaneously be the case both that resources within India are disproportionately provided to first-borns *and* that children of all ranks within India are shorter than they otherwise would be due to the epidemiological environment.

As an illustration, in its 2005-6 DHS, 55.3 percent of Indian households reported open defecation, and the mean child was 1.9 standard deviations below the reference mean; this combination is almost identical to neighboring Pakistan's in its 1990-1 DHS 15 years earlier, when 53.1 percent of households did not use a toilet or latrine and the mean height for age was 2 standard deviations below the mean. This section studies change over time within India by constructing a panel of districts out of India's 1992-3 and 1998-9 DHS surveys.

1.3.1 Data and empirical strategy

The National Family and Health Surveys (NFHS) are India's implementation of DHS surveys. This section analyzes a district-level panel constructed out of the NFHS-1 and NFHS-2.[12] Districts are political subdivisions of states. Some districts merged or split between survey rounds, so households in the survey are matched to a constructed "district" that may be a coarser partition than actual district boundaries. In particular, a primary sampling unit (PSU) is assigned to a constructed district such that all splits and merges are assigned to the coarser partition, creating the finest partition such that each PSU is in the same constructed district as all PSUs which would have shared a district with it in either period. Thus if there were two districts A and B in the first round, which split before the second round into A', B', and C (a new district containing part of A and of B), then all of A, B, A', B', and C would be a single constructed district, although splits this complicated are rare.[13]

The empirical strategy of this section is to compare the heights of rural children under 3 years old in the NFHS rounds 1 and 2, using district fixed effects.[14] In particular, I regress child height on the fraction of households reporting open defecation at two levels of aggre-

[12]The third and most recent NFHS does not include district identifiers. I thank Fred Arnold for assistance with district identifiers required to construct this panel.

[13]The NFHS was not constructed to reach all districts, so households are only included in the sample if they are members of districts that appear in both survey rounds, to permit district fixed effects.

[14]Although district fixed effects are used, the NFHS did not survey the same villages in the two survey rounds; thus there remains an important cross-sectional component to the heterogeneity studied.

gation: districts and villages (or more precisely, rural primary sampling units). Because open defecation has negative externalities on other households, it is necessary to test for effects of community-wide sanitation coverage, rather than simply comparing households that do and do not have latrines; section 1.4.1 considers the econometric implications of these negative externalities in more detail.

Therefore, the regression specification is:

$$height_{idvt} = \beta_1 open\ defecation^{d}_{dt} + \beta_2 open\ defecation^{v}_{dvt} + \alpha_d + \gamma_t + X_{idvt}\theta + A_{idvt}\vartheta + \varepsilon_{idvt}, \quad (1.2)$$

where i indexes individual children, d are districts, v are villages (rural PSUs), and t are survey rounds 1 and 2. The dependent variable, $height_{idvt}$ is the height of child i in standard deviations, scaled according to the WHO 2006 reference chart. As recommended by the WHO, outliers are excluded with z-scores less than -6 or greater than 6. The independent variables $open\ defecation^{d}_{dt}$ and $open\ defecation^{v}_{dvt}$ are computed fractions 0 to 1 of households reporting open defecation in the child's district and village, respectively. Fixed effects α_d and γ_t are included for districts and survey rounds. The vector A_{idvt} is a set of 72 indicators for age-in-month by sex, one for each month of age for boys and for girls.[15] Controls X_{idvt} are at the household or child level: electrification, water supply, household size, indictors for being Hindu or Muslim, a full set of birth order indicators interacted with the relationship of the child's mother to the head of the household, twinship indicators, and month-of-birth indicators.[16] Results are presented with and without controls and fixed effects to verify robustness. The mean PSU studied here contains 10 children under 3 used in these regressions.

[15]Panagariya (2012) has recently argued that height-for-age z score reference charts are inappropriate for Indian children; because age-in-months-by-sex is the level of disaggregation used to create height-for-age scores, these controls fully and flexibly account for any deviation between the mean height of Indian children and the reference charts.

[16]If the survey rounds were conducted in different places in different times of year, different children would be under 36 months old. Month of birth is correlated with early-life human capital inputs (*cf.* Doblhammer and Vaupel, 2001, about developed countries).

1.3.2 Regression results

Table 1.6 presents results from estimations of equation 1.2, with simple OLS in Panel A and district fixed effects in Panel B. Districts which saw greater differences in sanitation also present greater differences in child height.

District-level open defecation rates do not statistically significantly predict child height once village-level open defecation is included, which plausibly suggests that villages are much nearer than districts (which are much larger than villages) to capturing the geographic extent of sanitation externalities. Village-level open defecation predicts child height with or without district fixed effects and with or without individual controls.[17] The coefficient on village open defecation is smallest in absolute value with fixed effects and controls,[18] although it is not statistically significantly different from the other estimates; this could reflect the well-known attenuating bias of fixed effects, if much of the important variation in sanitation that is causing variation in height has been captured by other controls, leaving noise remaining.[19]

Consistency of fixed effects estimates, which subtract level differences across groups, depends on a properly linear specification. Column 4 demonstrates that a quadratic term for village-level open defecation is not statistically significant, and indeed changes signs with and without district fixed effects. Potential non-linear relationships between village sanitation coverage and child height will be considered in more detail in section 1.4.1.

[17] If, instead of omitting observations with height-for-age z-score beyond ± 6, a cut-off of ± 10 is used, then results are very similar. For example, the coefficient in column 1 of panel A becomes -0.768 (0.222); the smallest coefficient in absolute value, column 3 in Panel B, becomes -0.292 (0.138). If the log of height in centimeters is used as the dependent variable instead of the z score, moving from 0 percent to 100 percent open defecation is associated with an approximately 2 percent decrease in height ($t \approx 4$, analogously to column 2 of panel B).

[18] Would any village-level (instead of household-level) asset or indicator of well-being have the same effect as sanitation? Adding village electrification and water averages to the most controlled regression, column 3 of panel B, changes the point estimate on open defecation only slightly, from -0.35 to -0.33 ($s.e. = 0.12$); these two village level variables have t-statistics of 1.15 and -0.59, respectively, with a joint F-statistic of 0.73.

[19] For readers concerned about this possibility, regressing height on village-level open defecation with no district or time fixed effects produces an estimate of -0.700 ($t \approx 4.5$) and of -0.501 ($t \approx 4.8$) with all the non-fixed-effect controls.

1.4 Evidence from pooled Indian and African surveys

Can differences in village-level sanitation coverage explain the difference in height between rural children in India and in sub-Saharan Africa? This section addresses this question using pooled child-level data from the rural parts of nine DHS surveys: India's 2005-6 NFHS-3 and eight surveys from Africa in the 2000s.[20] In particular, the DHS surveys nearest 2005 (and balanced before and after) were selected from the five largest African countries available;[21] the included countries account for 46 percent of the 2012 population of sub-Sahara Africa.

The argument of this section proceeds in several stages, building to a statistical decomposition of the India-Africa height difference, in the sense of Oaxaca-Blinder. First, section 1.4.1 verifies an association between village-level sanitation and height within the two regions. In particular, this section assesses the linearity of the relationship (assumed by some decomposition techniques), and notes that a village-level effect implies the presence of negative externalities. Then, section 1.4.2 considers a paradox implied by Deaton's (2007) finding that height is not strongly associated with GDP: the within-region association between open defecation and well-being has a different slope from the across-region association. Finally, section 1.4.3 proceeds with the decomposition, applying linear and non-parametric approaches to explain the India-Africa gap.

1.4.1 Effects of village sanitation: A negative externality

As a first step towards explaining the height difference between Indian and African children, this section verifies that village-level open defecation predicts children's height within each

[20] Here I again follow the WHO recommendation of dropping observations with height-for-age z-scores more than 6 standard deviations from the mean.

[21] This excludes South Africa, where height has not been measured in a DHS survey. Beyond this data availability constraint, this exclusion may be appropriate due to South Africa's unique history and demography; its exceptionally high sanitation coverage (11.6 percent open defecation in 1998) would make it a positive outlier even in the African sample. The eight African DHS surveys used are: DRC 2007, Ethiopia 2000 and 2005, Kenya 2005 and 2008, Nigeria 2003 and 2008, and Tanzania 2004. DHS sampling weights are used throughout.

region. As in section 1.3, I use household-level DHS data to find the fraction of households in a PSU reporting open defecation, which I treat as an estimate of village-level open defecation.

Thus, separately for each region r (Africa and India), I estimate:

$$\begin{aligned} height_{ivcr} &= \beta_1 open\ defecation^{v}_{vcr} + \beta_3 open\ defecation^{2}_{vcr} + \beta_3 open\ defecation^{i}_{ivcr} + \\ &\quad \alpha_c + X_{ivcr}\theta + A_{ivcr}\vartheta + \varepsilon_{ivcr}, \end{aligned} \tag{1.3}$$

where i indexes individual children under 5, v are villages (rural PSUs), c are country-years (DHS surveys) in Africa and states in India, and r are regions (India or Africa). The dependent variable, $height_{ivcr}$ is the height of child i in standard deviations, scaled according to the WHO 2006 reference chart. The independent variable $open\ defecation^{v}_{vcr}$ is the computed fraction 0 to 1 of households reporting open defecation in the child's village (again, implemented as rural primary sampling unit), with a quadratic term included in some specifications. Household-level open defecation $open\ defecation^{i}_{ivcr}$ is an indicator, 0 or 1, for the child's household. Including both household and village-level open defecation tests whether one household's open defecation involves negative externalities for other households.[22] In other words, is it only a household's own sanitation that matters, or do other households' sanitation matter, even controlling for one's own?

Fixed effects α_c are included for some specifications. As before, the vector A_{ivcr} is a set of 72 indicators for age-in-month by sex, one for each month of age for boys and for girls. Finally, X_{ivcr} is a vector of child or household level controls: indicators for household dirt floor; access to piped water; electrification; and ownership of a TV, bicycle, motorcycle, and clean cooking fuel; and the child's mother's literacy, knowledge of oral rehydration, age at first birth, count of children ever born, and relationship to the head of the household. These

[22] Günther and Fink's (2010) working paper version of Fink et al. (2011) conducts a similar analysis, regressing diarrhea and child mortality on household and cluster-mean water and sanitation variables.

controls help ensure that any correlation between height and open defecation is unlikely to reflect mere wealth differences.

Regression results

Figure 1.6 plots, separately for the African and Indian samples, the local polynomial regressions of child height on village open defecation, separating households that do and do not defecate openly. The figures make clear the distinct private and social benefits of sanitation. The private benefit is the vertical distance between the two lines; thus, in an average Indian village, children in households that do not openly defecate are about half of a standard deviation taller than children in households that do, although at least some of this difference reflects wealth and other heterogeneity. The social benefit – a negative externality on other households – is visible in the downward slope of the regression lines: children living in villages with less open defecation overall are taller, on average. Of course, some fraction of these correlations also reflects omitted variable bias. The dashed vertical lines show that open defecation is much more common in the Indian than in the African data. Note that, comparing across continents only children in households that do not practice open defecation, at all levels of village open defecation children in households that do not practice open defecation are shorter in Africa than in India.

Table 1.7 verifies the statistical significance of these results and estimates regression equation 1.3. In both samples, there is a clear association between child height and village-level sanitation. Especially in the Indian sample, the estimate changes little when controls are added. The coefficient on household-level sanitation is less robust: in the Indian sample it becomes much smaller when household and child controls are added, and in the African sample it loses statistical significance.

Linear effects on height?

So far, this paper has largely studied linear regression. A non-linear relationship between sanitation and height could be important for two reasons: first, fixed effects regression could be inconsistent; and second, a linear Blinder-Oaxaca decomposition could be inappropriate. Returning to figure 1.6, the relationship between sanitation and height appears approximately linear in the Indian data, but may not be among openly-defecating households in the African data.

Non-linearity can be tested by adding a quadratic term. Already, in table 1.6, a quadratic term was not statistically significant in the first two Indian DHS surveys. Panel A finds again that, in the Indian NFHS-3, there is no evidence for a quadratic term. In contrast, Panel B does find a quadratic term in the African sample. In light of the evidence for the importance of open defecation per square kilometer presented in section 1.2.2, one possible explanation for this negative quadratic term is that population density is relatively low in these African countries, so open defecation is not as important for health until there is more of it; unfortunately, geographic data such as population density is not generally available at the DHS PSU level.

1.4.2 A paradox: International differences in well-being

Deaton (2007) finds that international differences in height are not well explained by differences in GDP or child mortality. How could this be, given that poor sanitation increases infant mortality (*see* chapter 2), and richer people are more likely to have toilets or latrines?

Figure 1.7 suggests that this puzzle is an example of Simpson's Paradox: within separate subsets of a larger sample, the relationship between two variables can be very different from the relationship between the two variables in the larger, pooled sample.[23] In particular, the relationship in the pooled data also depends upon the relationship among group

[23]The difficulties involved in inferring relationships about individuals from group average data are also sometimes referred to as the problem of "ecological inference."

means. Consider a large dataset partitioned into subsets indexed $s \in S$. Let $\hat{\beta}$ be the OLS coefficient of y on x in the whole, pooled dataset, and $\hat{\beta}_s$ the OLS coefficient found when the data are restricted to subset s. Further, let $\hat{\beta}_b$ be the "between" regression coefficient found by regressing subsample means $\bar{y}_s$ on $\bar{x}_s$, weighted by the number of observations in each subsample. Then

$$\hat{\beta} = \sum_{s \in S} \lambda_s \hat{\beta}_s + \lambda_b \hat{\beta}_b, \tag{1.4}$$

where the weights λ_s are the fractions of the total sum of squares in each subsample s and λ_b is the fraction of the sum of squares from the subsample means. Therefore, if the between coefficient is very different from the within coefficients, the pooled coefficient computed from the entire dataset could also be quite different from the within-subsample slopes.

Figure 1.7 plots within-region, between, and pooled slopes to clarify this paradoxical case. Within both the Indian and African subsamples, more village-level open defecation is, indeed, associated with more infant mortality and less wealth. However, India has more open defecation, lower infant mortality, and more wealth, represented by the plotted circles. Therefore, the pooled regressions are essentially flat, potentially misleadingly showing no association between open defecation and infant mortality or a count of household assets included in the DHS – consistently with Deaton's original result.

1.4.3 Decomposing the gap

Decomposition methods estimate the fraction of a difference between two groups that is statistically explained by differences in other variables (Fortin et al., 2011). Decomposition techniques are commonly applied to wage inequality in labor economics and to demographic rates. Like any econometric analysis of observational data, whether decomposition results have a causal interpretation depends on the context and the sources of variation in independent variables. This section estimates the fraction of the India-Africa height gap statistically "explained" by differences in village-level sanitation, a main result of the pa-

per. Note that local area open defecation is used, rather than simply an indicator for a household open defecation, because of local externalities of sanitation.

Methods of decomposition

Three methods of decomposition are used. The first is a straightforward application of regression, as in table 1.5. I regress

$$height_{ivs} = \alpha India_s + \beta open\ defecation^{v}_{vs} + X_{ivs}\theta + \varepsilon_{is}, \tag{1.5}$$

where $height_{ivs}$ is the height-for-age z score of child i in village v in sample s, either India or Africa. The coefficient of interest is α, on an indicator variable that the child lives in India. The econometric question is by how much adding a control for village level open defecation shifts $\hat{\alpha}$ in the positive direction. This is essentially identical to the pooled Blinder-Oaxaca decomposition with an indicator for group membership recommended by Jann (2008). The statistical significance of the change in $\hat{\alpha}$ is evaluated with a Hausman χ^2 test. Various sets of control variables X_{ivs} are added in turn, which will, in general, change both $\hat{\alpha}$ and $\hat{\beta}$.

The second method is a weighted two-way Blinder (1973)-Oaxaca (1973) decomposition, using a Stata implementation by Jann (2008). In particular, having seen in section 1.4.2 that open defecation has different correlations within and across the Indian and African samples, I implement Reimers's (1983) recommendation to first estimate

$$height_{ivs} = \beta_s open\ defecation_{vs} + X_{ivs}\theta_s + \varepsilon_{is}, \tag{1.6}$$

separately for each sample s, and then compute the difference in height "explained" by open defecation as

$$\left(0.5\hat{\beta}_{\text{India}} + 0.5\hat{\beta}_{\text{Africa}}\right)\left(\overline{open\ defecation_{v,\text{Africa}}} - \overline{open\ defecation_{v,\text{India}}}\right), \tag{1.7}$$

creating a counterfactual "effect" of sanitation by weighting equally the within-sample slopes.

The third method is a non-linear decomposition, which computes a new mean for the Indian sample after reweighting to match the African sample's distribution of a set of observable independent variables.[24] In particular, the approach is to construct a counterfactual mean height of Indian children. First, partition both samples into groups $g \in G(X)$, which share values or ranges of values of a set of covariates X (which could include measures of open defecation). Next, for each group, compute $f(g\,|s)$, the empirical density of sample $s \in \{\text{India}, \text{Africa}\}$ in group g, using sampling weights. Finally, compute the counterfactual mean

$$\tilde{h}_{\text{India}} = \sum_{g \in G(X)} \sum_{i \in g} \frac{f(g\,|\text{Africa})}{f(g\,|\text{India})} w_i h_i, \tag{1.8}$$

where w_i is the sampling weight of observation i and h_i is the height-for-age z-score of child i in the Indian sample. The unexplained gap is then $\tilde{h}_{\text{India}} - \bar{h}_{\text{Africa}}$. The basic set of reweighting variables used is village and household open defecation, split into 20 groups corresponding to 10 levels of village open defecation for households that do and do not openly defecate.

Decomposition results

Table 1.8 presents the decomposition results. Panel A reports the change in the OLS coefficient on a dummy variable for India when a linear control for village-level open defecation is included, as in table 1.5. Panel B reports the change in the unexplained difference when open defecation is added to a weighted Blinder-Oaxaca decomposition. Panel C presents counterfactual differences from non-parametrically weighting the Indian sample to match the distribution of village and household open defecation in the African sample.

[24] Geruso (2012) recently applied this approach to compute the fraction of the U.S. black-white life expectancy gap that can be explained by a group of socioeconomic variables.

Columns 1 and 2 present the basic result: the simple sample mean with no controls, with and without adjustment for sanitation. Village-level open defecation linearly explains 99 percent of the India-Africa gap.[25] In the Blinder-Oaxaca and non-parametric decompositions, sanitation explains more than 100 percent of the gap; this "overshooting" is plausible because, in addition to having worse sanitation, Indian households are richer, on average, than African households.

The next four pairs of columns similarly find that open defecation explains much of the India-Africa gap after controls are added. Specifically, columns 3 and 4 control for demographic variables before decomposing the remaining gap: sex-specific birth order indicators and an indicator for single or multiple birth.[26] Columns 5 and 6 first control for a vector of socioeconomic controls, the same controls used earlier in equation 1.3. Columns 7 and 8 control for a village-level estimate of infant mortality, computed from mothers' reported birth history. Finally, columns 9 and 10 control for mothers' height. Because Indian mothers, like Indian children, are short, Indian children are taller than African children, conditional on their mothers' height. However, this correlation is unlikely to be quantitatively importantly causal,[27] and the counterfactual increase in Indian height from

[25]Household-level open defecation, used instead of village-level, explains 68 percent of the gap, a reminder of the importance of disease externalities.

[26]Jayachandran and Pande (2012) note that birth order is a predictor of child height in India (see footnote 11). For example, in the sample of rural children under 5 used here, I find that first children are 0.063 ($p = 0.044$) standard deviations taller than second children. However, this gap falls to 0.034 ($p = 0.275$) if controls for village and household open defecation are included (jointly significant with $F_2 \approx 95$), suggesting that children born into larger households may be more likely to be exposed to environmental fecal pathogens, although various important forms of intra-household discrimination surely exist, as well (*cf.* Jeffery et al., 1989).

Jayachandran and Pande (2012) also find, in their pooled DHS sample, that Indian first-borns are *taller* than African first-borns. In the individual-level sample studied here, Indian first-borns are 0.019 standard deviations shorter (a much smaller gap than in the full sample, but still negative); controlling for open defecation, they are 0.133 standard deviations taller, a 0.15 increase similar to those when sanitation is controlled for in table 1.8. If, for example, IMR is controlled for, Indian first-borns are 0.132 standard deviations shorter than African first-borns, which increases by 0.16 to Indians being 0.025 taller.

[27]Is it possible that historical conditions that restricted Indian mothers' size, but have now improved, are importantly restricting the fetal growth of their children? This question is difficult to answer in part because there are at least five reasons mothers' and children's height would be correlated: (1) mother's genetics, (2) assortative mating and father's genetics, (3) correlation of the child's early-life environment with the mother's early life environment, (4) endogenous effects of mothers' early life environments on their adult ability to care for their children (including *in utero* and through marriage markets), and finally (5)

matching African sanitation continues to exceed the original, simple gap to be explained, even after adjustment for mothers' height.

The *level* of the India-Africa height gap depends on the particular set of controls added before sanitation is accounted for. However, the counterfactual *change* in height upon accounting for open defecation is strikingly similar across specifications. In particular, the flexible non-parametric decomposition in Panel C might best accommodate any shape of the height-sanitation association. In all cases, non-parametrically matching the African distribution of open defecation increases the counterfactual Indian mean height by more than the 0.142 standard deviation simple difference in means.

1.5 Discussion: How much of the Africa-India gap does sanitation explain?

Several dimensions of variation in open defecation quantitatively similarly predict variation in child height: heterogeneity in aggregated country means, changes within Indian districts, and variation across village-level averages. Moreover, chapters 2 and 3 document causally well-identified estimates of effects of sanitation on height. Finally, in the sense of econometric decomposition, exceptionally widespread open defecation can explain much of exceptional Indian stunting. So, how much taller would Indian children be if they enjoyed better sanitation and less exposure to fecal pathogens?

intrauterine growth restriction directly caused by the historically determined aspects of a mother's size. Although all five would be reflected in a simple regression of child size on mother's size, only the last mechanism would allow mothers' stunting to itself cause present-day children's stunting. Evidence against this – that the mother-child height correlation is relatively *low* in India – is discussed in the working paper version of this chapter.

1.5.1 A linear thought experiment

To answer the question above would require knowing a "true" average causal effect of open defecation on Indian children – an unknown number that would likely average over much heterogeneity. However, one can envision possible answers by comparing a range of estimates of the association between height and open defecation, each reflecting its own particular context and combination of internal and external validity.

Children in rural India are, on average, 0.142 standard deviations shorter than children in the rural African sample in section 1.4; open defecation is 31.6 percentage points more common in India. Therefore, imagining a linear causal effect of sanitation β^t (t for "true"), the fraction of the rural India-Africa height gap that open defecation rates would explain would be whatever fraction β^t is of a 0.45 ($= 0.142 \div 0.316$) standard deviation increase in height resulting from moving from 100 percent to 0 percent open defecation. Moreover, children in India, where 55 percent of households openly defecated in the DHS, are about 2 standard deviations shorter than the reference mean. So, whatever fraction β^t is of 3.6 would be the fraction of the India-U.S. gap explained.

Table 1.9 collects estimates of the linear association between height and open defecation from this paper and others. As in section 1.4.3, "explaining" over 100 percent of the gap is plausible because wealth differences predict that Indian children should be taller than African children. Unsurprisingly, the instrumental variable treatment-on-the-treated estimate in chapter 3's small experimental sample has a large confidence interval. Both this estimate and chapter 2's may overstate the direct effect of latrines *per se* because the programs studied also promoted use of existing latrines.

Like many regression estimates of the effects of inputs on human capital, some of these may be biased upwards. Collectively, however, they suggest that a linear approximation to the "true" average causal effect of village-level sanitation coverage on Indian children's height is likely to be a large fraction of 0.45 (ignoring the additional explanatory power of

population density). If so, then sanitation could explain much or all of the difference in heights between Indian and African children.

1.5.2 Conclusion: Stunting, "malnutrition," and externalities

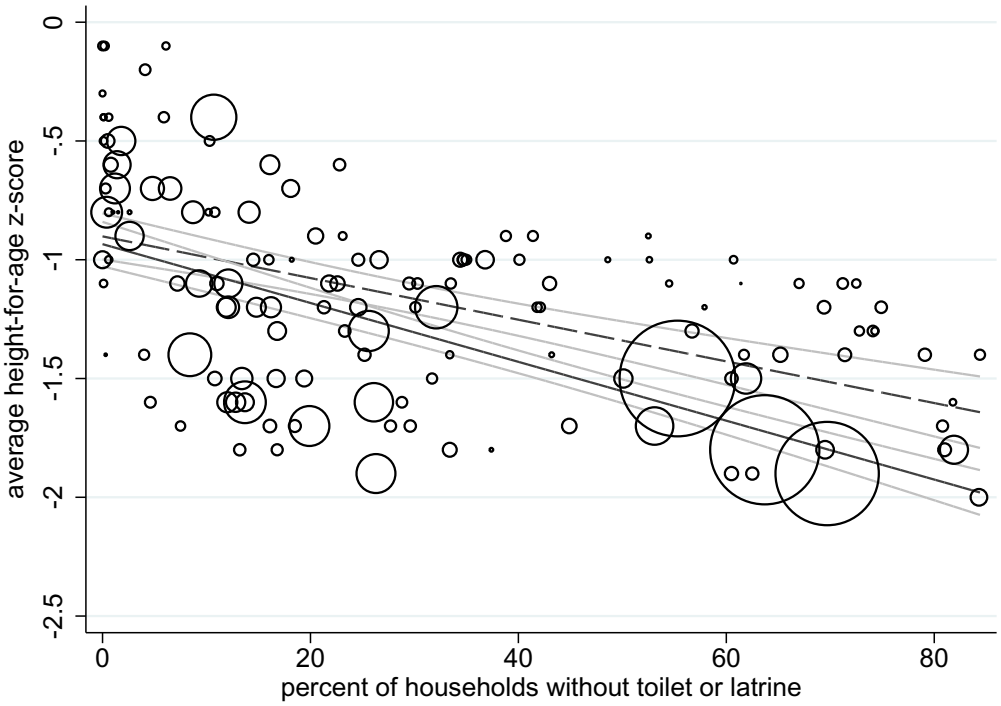
Section 1.4.1 presented evidence that one household's open defecation imposes negative externalities on its neighbors. Village-level externalities are important for at least two reasons. First, negative externalities are a classic rationale in public economics for government intervention: if households do not consider the effect of their own open defecation on other people, they will be too reluctant to switch to using a latrine. Second, statistical approaches that only study private resources will be unable to fully explain heterogeneity in height. For example, Tarozzi (2008) finds that even Indian children in the richest households in the NFHS (that is, with the most assets) are still shorter than international reference norms. Panagariya (2012) interprets this result to suggest that international norms are incorrect for Indian children, because even children with "elite or privileged" household health inputs are stunted.[28] Yet, this interpretation ignores externalities: many of the asset-rich households in the NFHS are exposed to a disease environment created by the open defecation of other households. Bhandari et al. (2002) study the height of Indian children living in Green Park, a single affluent neighborhood in South Delhi where there is essentially no open defecation; these children grow to international reference heights.

Although a child's low height-for-age is often called "malnutrition," Waterlow (2011) has advocated instead using "the term 'stunted,' which is purely descriptive and does not prejudge the question of whether or not the growth deficit is really the result of malnutrition," often narrowly interpreted as food, especially in policy debates. Early-life disease – and especially chronic disease due to fecal pathogens in the environment – appears to be another important determinant of height. If so, determining whether open defecation is an

[28] Tarozzi does recognize that his approach does not capture important effects of "the epidemiological environment, with its impact on infections" (463).

importantly binding constraint on Indian children's height may be a step towards a policy response able to resolve this Asian enigma.

(a) children born in the last 3 years



(b) children born in the last 5 years

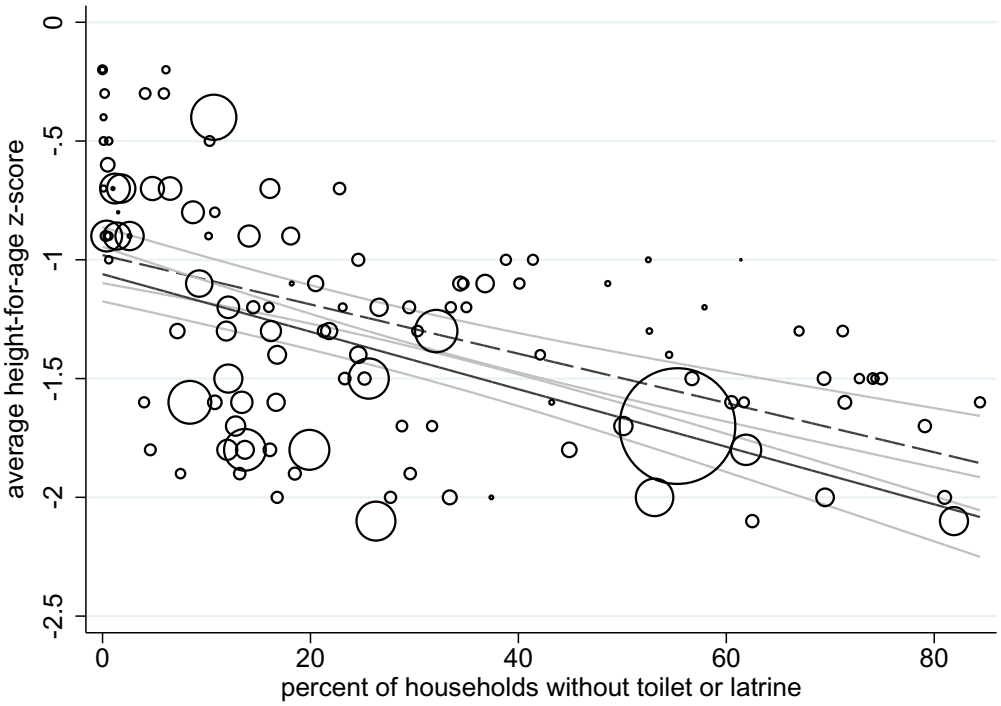
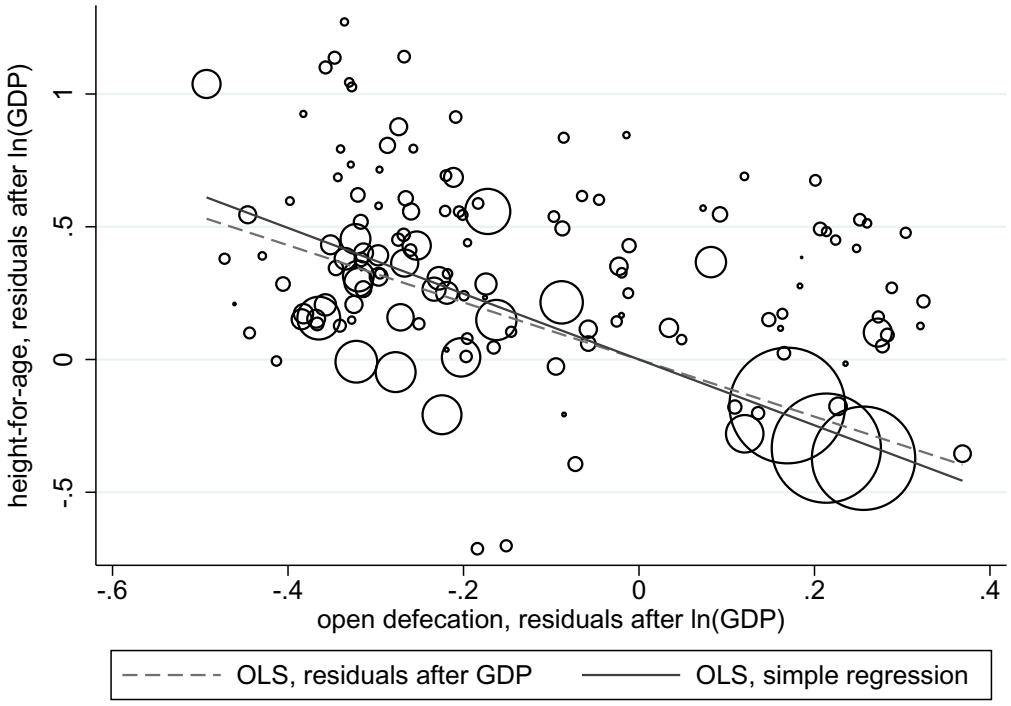


Figure 1.1: Open defecation predicts child height, DHS survey round country-years

Solid OLS regression lines weight by country population; dashed lines are unweighted.

(a) residuals after ln(GDP)



(b) country means and subsets of India (2005) by wealth

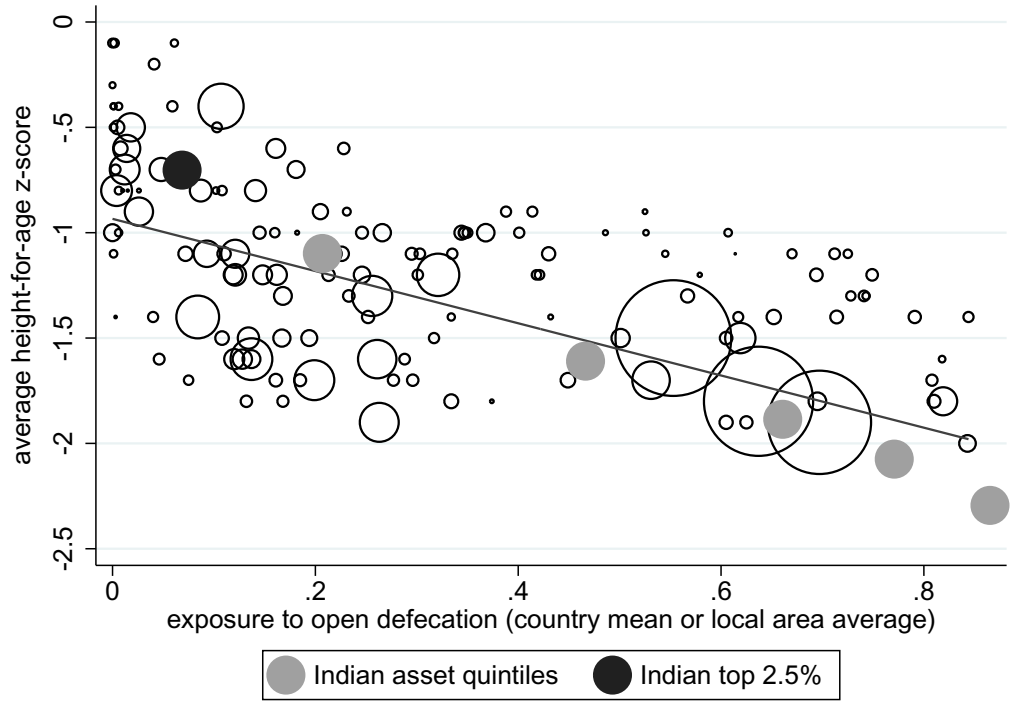
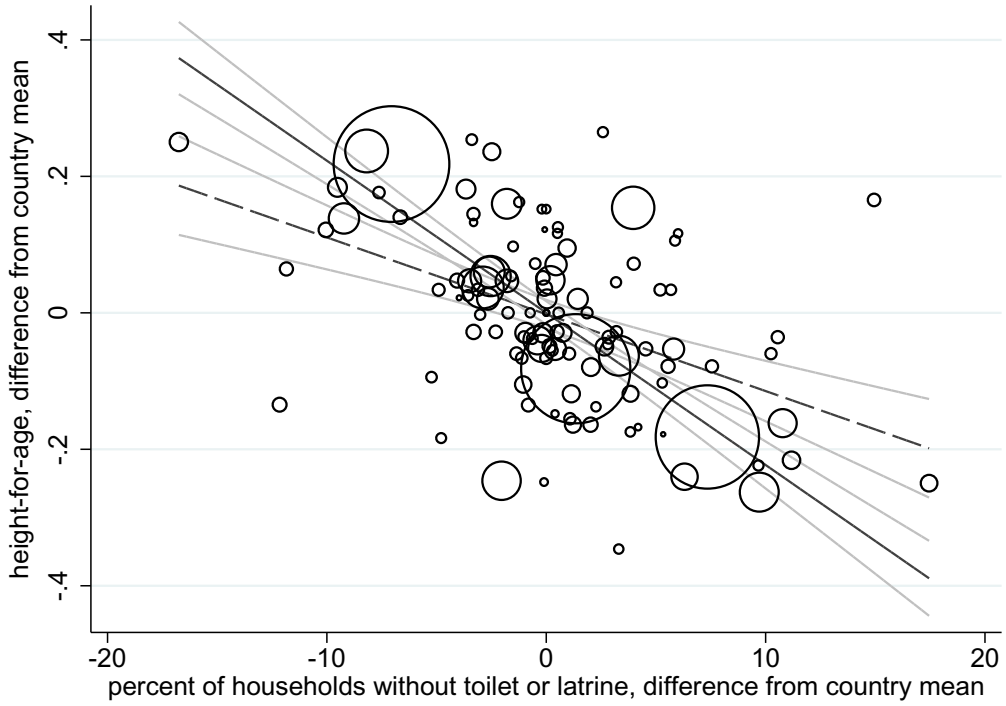


Figure 1.2: Wealth does not account for the sanitation-height association

(a) children born in the last 3 years



(b) children born in the last 5 years

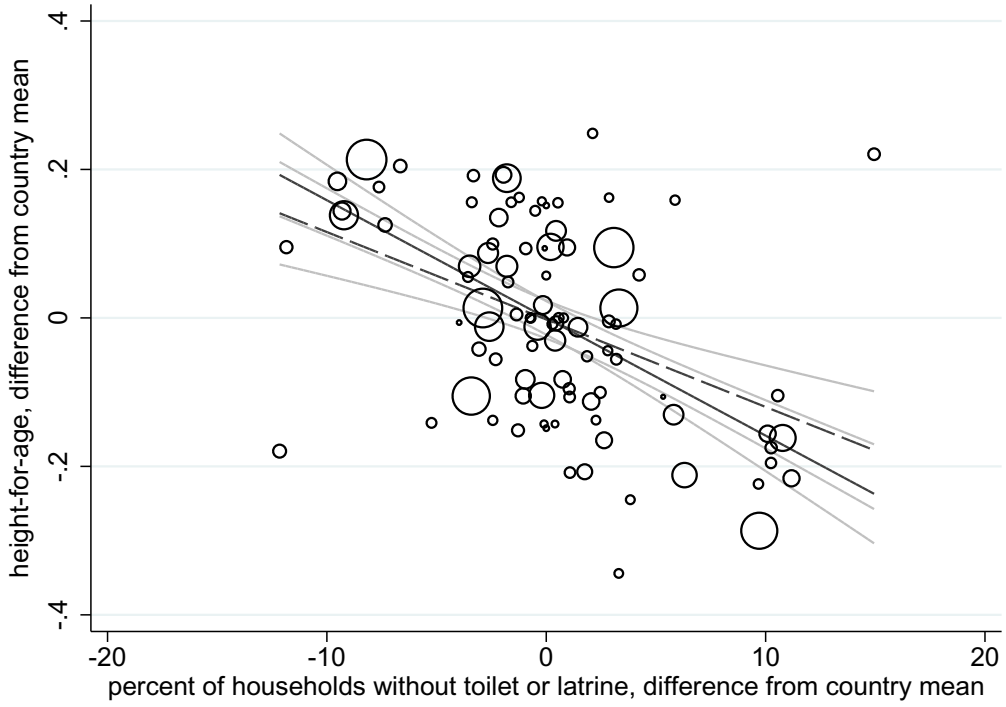
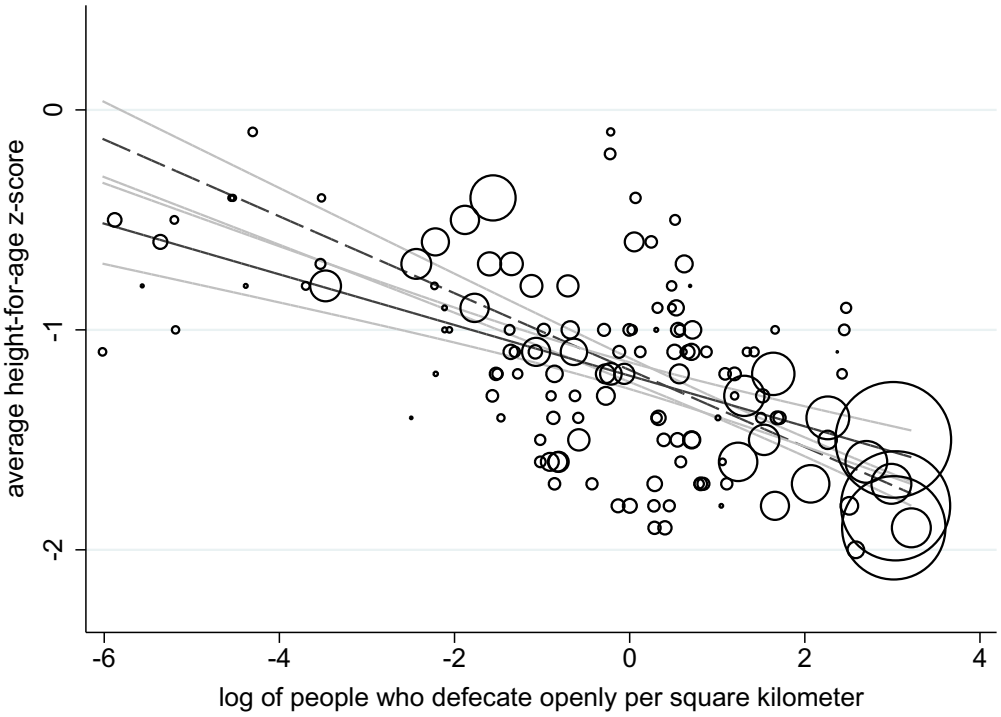


Figure 1.3: Difference from country mean sanitation explains height difference

(a) children born in the last 3 years



(b) children born in the last 5 years

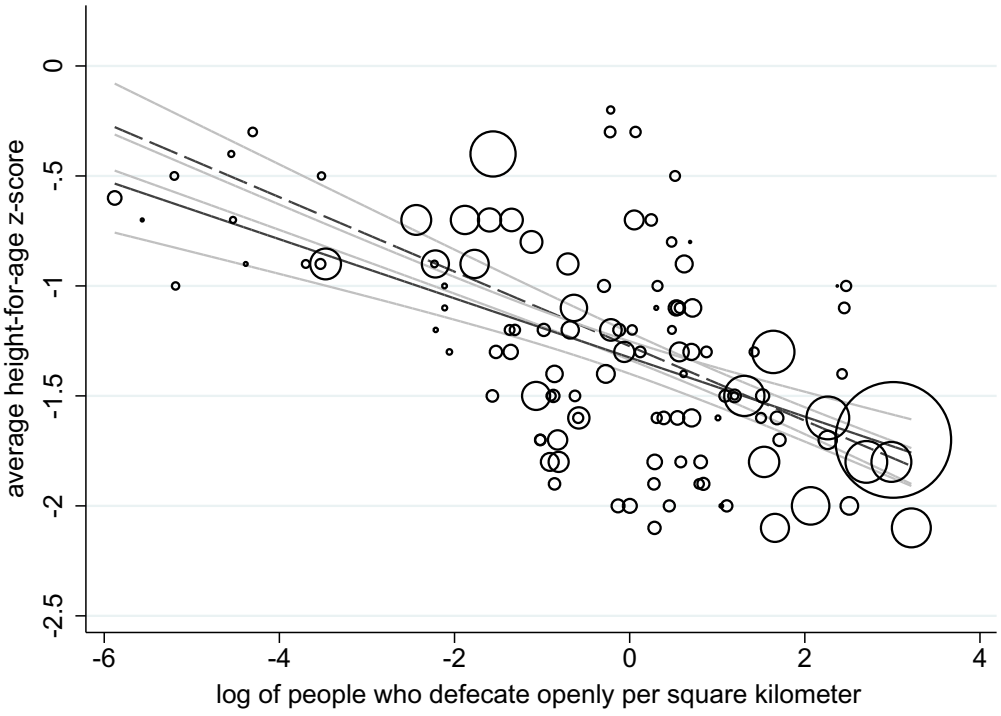


Figure 1.4: Open defecation interacts with population density to predict child height

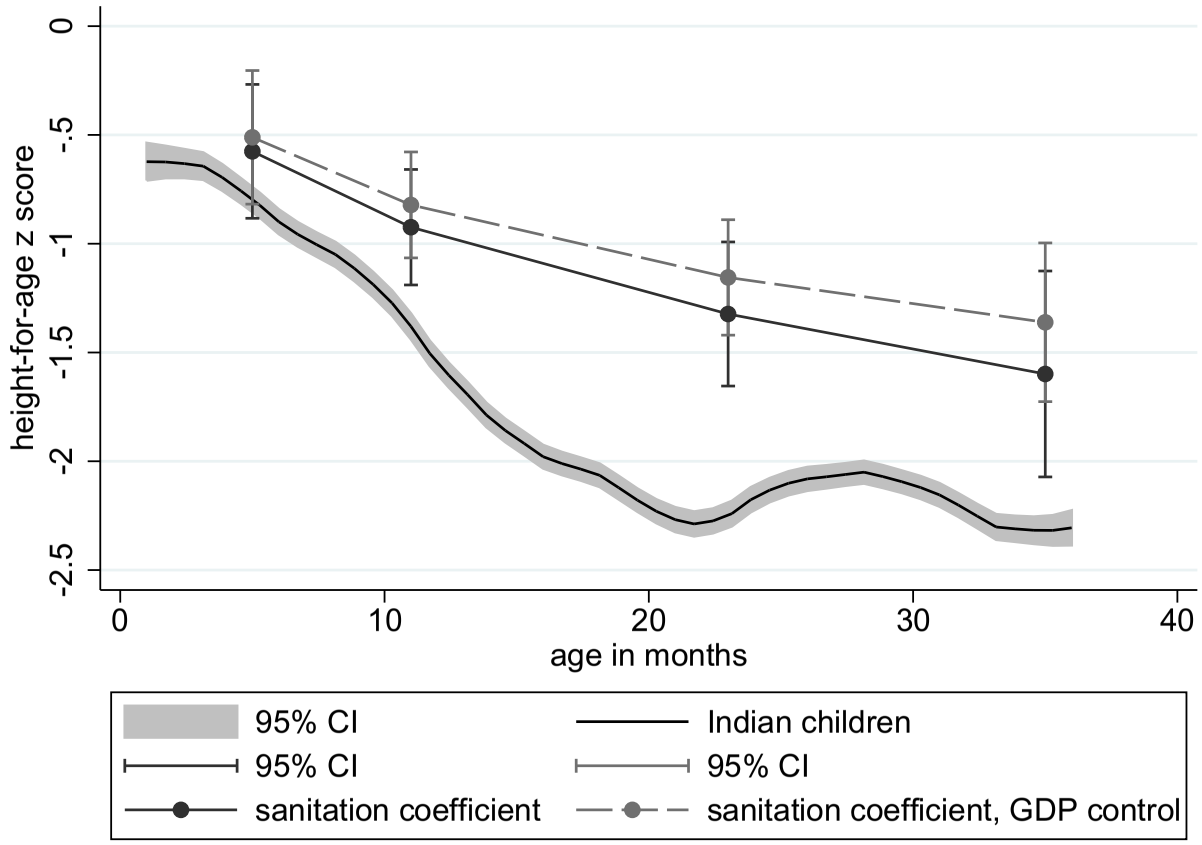
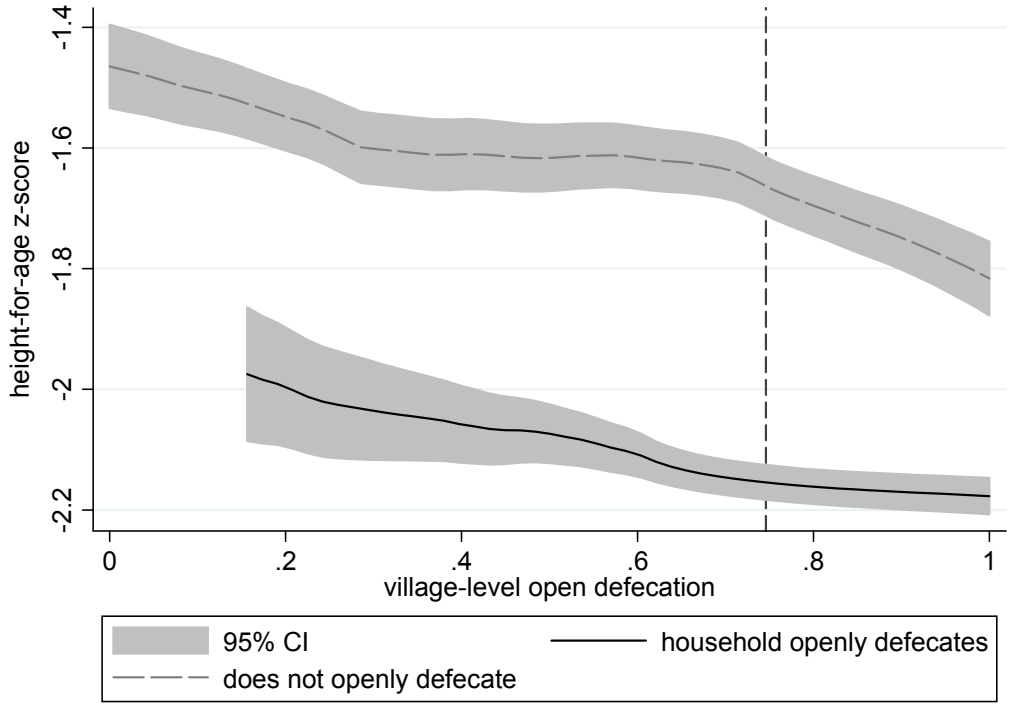


Figure 1.5: Open defecation is more steeply associated with child height at older ages

Confidence intervals are estimates of the coefficient from a regression of average country-level child height-for-age on open defecation, restricting the sample to four age categories. The curve plots the average height-for-age z-score of Indian children by age, for reference.

(a) India (NFHS-3)



(b) five African countries (8 pooled DHS surveys)

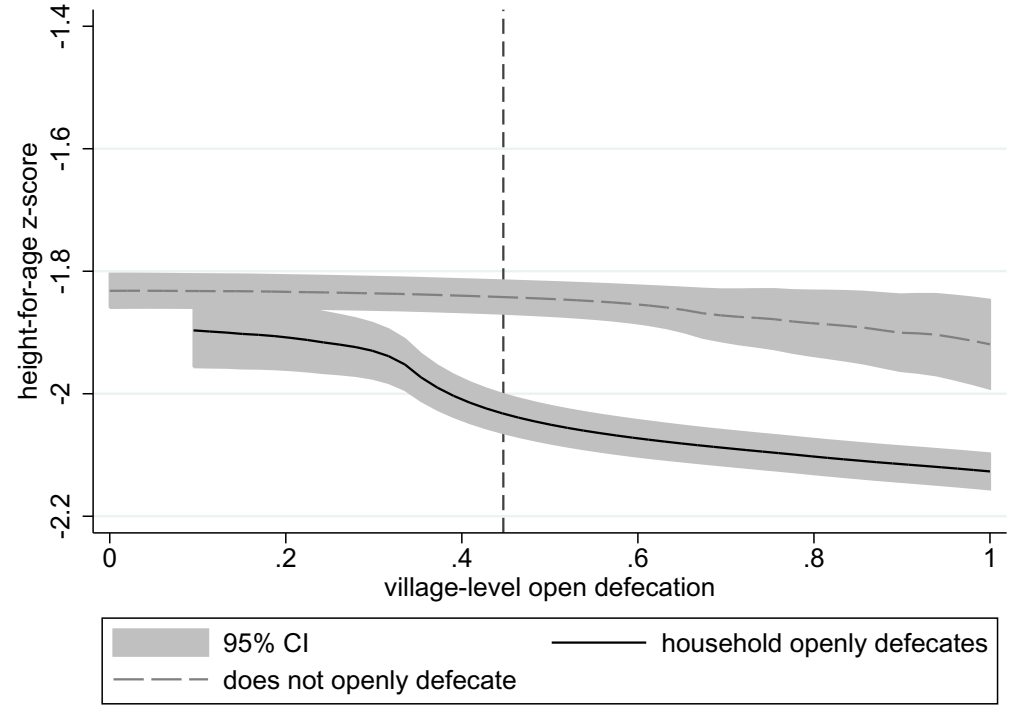
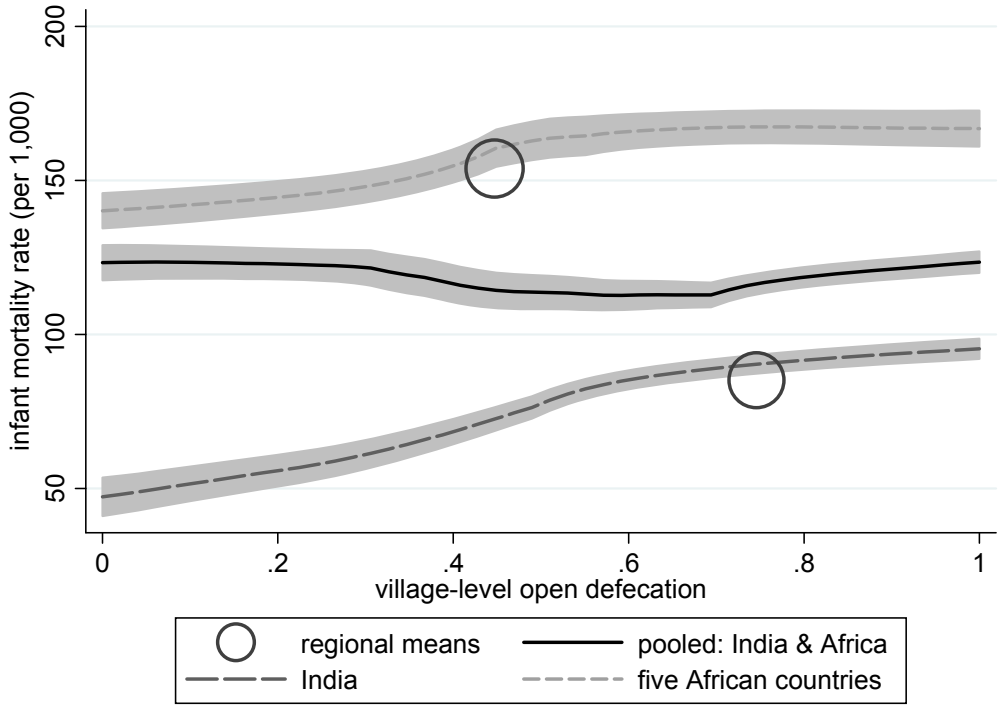


Figure 1.6: Negative externalities: Village-level open defecation predicts child height

Vertical dotted lines mark the overall mean open defecation fraction in these two rural samples.

(a) infant mortality



(b) asset ownership

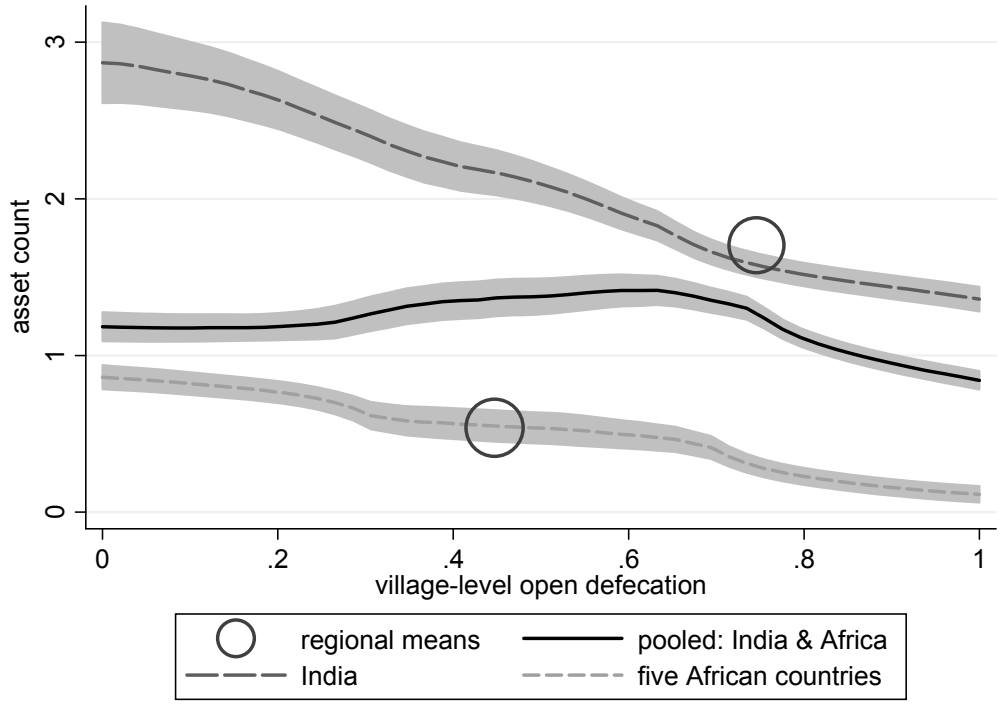


Figure 1.7: Simpson's paradox: Open defecation and well-being across and within regions

Table 1.1: Open defecation predicts child height across DHS surveys

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Average height-for-age z-score of children born in last 3 years							
open defecation	-1.239***	-1.326***	-1.002***	-0.962*	-1.028†	-1.111*	-0.663***
	(0.226)	(0.158)	(0.156)	(0.434)	(0.583)	(0.505)	(0.181)
open defecation							-1.499*
× density							(0.631)
ln(GDP)			0.202**	0.512***	0.472**	0.757†	0.280***
			(0.0733)	(0.146)	(0.174)	(0.423)	(0.0525)
women's height				0.0130	-0.0564	-0.0369	0.0425**
				(0.0476)	(0.0904)	(0.106)	(0.0143)
population density							0.0418
							(0.190)
year FEs		✓	✓	✓	✓	✓	✓
country FEs				✓	✓	✓	
region time trends					✓		
controls						✓	
n (DHS surveys)	140	140	140	130	130	102	130
R^2	0.542	0.679	0.744	0.988	0.990	0.991	0.862

Table 1.1, continued

Panel B: Average height-for-age z-score of children born in last 5 years

open defecation	-1.211***	-1.443***	-0.910***	-1.335**	-1.175	-1.354†	-0.689***
	(0.290)	(0.203)	(0.208)	(0.474)	(0.806)	(0.704)	(0.154)
open defecation							-1.446*
× density							(0.549)
ln(GDP)			0.276**	0.270	0.208	0.178	0.341***
			(0.0859)	(0.208)	(0.308)	(0.226)	(0.0454)
women's height				0.0674	-0.0180	0.0352	0.0544***
				(0.0774)	(0.145)	(0.107)	(0.0147)
population density							-0.0181
							(0.137)
year FEs		✓	✓	✓	✓	✓	✓
country FEs				✓	✓	✓	
region time trends					✓		
controls						✓	
n (DHS surveys)	117	117	117	108	108	104	108
R^2	0.369	0.555	0.689	0.990	0.991	0.991	0.893

Standard errors clustered by country in parentheses (65 countries in panel A, 59 in panel B). p-values: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1. Population density is demeaned to preserve the interpretation of open defecation. Controls are calorie deficit, female literacy, water within 15 minutes, knowledge of oral rehydration, polity score, and autocracy score; see the text for more complete variable definitions.

Table 1.2: Open defecation predicts child height, omitting each world region in turn

	(1)	(2)	(3)	(4)	(5)	(6)
omitted region:	S.S. Africa	S. Asia	M.E.&N.Af.	C. Asia	S.E. Asia	L. Amer.
Panel A: year fixed effects, no controls						
open defecation	-1.711***	-0.988***	-1.081***	-1.316***	-1.352***	-1.186***
	(0.240)	(0.140)	(0.216)	(0.165)	(0.161)	(0.186)
n (DHS surveys)	62	129	121	135	135	118
R^2	0.808	0.517	0.620	0.680	0.683	0.721
Panel B: year fixed effects, control for ln(GDP)						
open defecation	-0.733***	-0.387*	-0.866***	-1.001***	-1.028***	-1.065***
	(0.195)	(0.151)	(0.178)	(0.162)	(0.157)	(0.169)
n (DHS surveys)	62	129	121	135	135	118
R^2	0.925	0.703	0.693	0.741	0.747	0.736
Panel C: year fixed effects, control for average height of women						
open defecation	-0.679†	-0.994***	-0.864***	-1.064***	-1.096***	-0.911***
	(0.364)	(0.135)	(0.202)	(0.157)	(0.150)	(0.186)
n (DHS surveys)	57	121	112	125	125	110
R^2	0.880	0.522	0.639	0.698	0.701	0.768

Dependent variable is mean height-for-age of children under 3. Standard errors clustered by country in parentheses.
p-values: $\dagger$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1.

Table 1.3: Alternative "placebo" independent variables do not predict height, conditional on sanitation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
variable:	female literacy	nearby water	calorie deficit	electrification	immunization	polity score	breastfeeding
Panel A: Unconditional associations							
"placebo"	0.0143*	0.00478*	-0.00251*	0.00675*	1.070***	-0.0274*	-0.109†
	(0.00367)	(0.00203)	(0.00120)	(0.00254)	(0.178)	(0.0117)	(0.0605)
Panel B: Associations conditional on open defecation and other variables							
"placebo"	0.00348	0.00382	0.0000230	-0.00271	-0.159	-0.0119	-0.054
	(0.00327)	(0.00268)	(0.000964)	(0.00226)	(0.138)	(0.00787)	(0.036)
open defecat.	-0.880***	-1.077***	-1.002***	-0.918***	-1.003***	-0.843***	-1.123***
	(0.205)	(0.158)	(0.157)	(0.141)	(0.150)	(0.212)	(0.187)
ln(GDP)	✓	✓	✓	✓	✓	✓	✓
year FEs	✓	✓	✓	✓	✓	✓	✓
n	138	124	140	135	126	138	139
R^2	0.752	0.792	0.744	0.726	0.802	0.753	0.689

Dependent variable is mean height-for-age of children under 3. Standard errors clustered by country in parentheses. p-values: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction.

Table 1.4: Open defecation is more steeply associated with child height in urban than in rural areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
subsample:	total	urban	urban	urban	rural	rural	rural
Panel A: Height-for-age z-score of children under 3							
open defecation	-1.239***	-2.577***	-1.577**	-1.603***	-0.853***	-0.572***	-0.689***
	(0.226)	(0.688)	(0.504)	(0.447)	(0.173)	(0.157)	(0.143)
rural = urban					$\chi^2 = 9.50$	$\chi^2 = 6.07$	$\chi^2 = 6.69$
					$p = 0.002$	$p = 0.014$	$p = 0.009$
women's height			0.0512***	0.0506***		0.0407**	0.0409***
			(0.0130)	(0.00924)		(0.0144)	(0.0107)
rural = urban						$\chi^2 = 2.20$	$\chi^2 = 1.79$
						$p = 0.138$	$p = 0.181$
year FEs				✓			✓
n (DHS surveys)	140	140	130	130	140	130	130
R^2	0.542	0.403	0.518	0.638	0.467	0.506	0.624

Table 1.4, continued

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
subsample:	total	urban	urban	urban	rural	rural	rural
Panel B: Height-for-age z-score of children under 5							
open defecation	-1.211***	-2.219**	-1.620**	-1.875***	-0.755**	-0.527**	-0.743***
	(0.290)	(0.825)	(0.586)	(0.487)	(0.233)	(0.198)	(0.161)
rural = urban					$\chi^2 = 5.16$	$\chi^2 = 7.01$	$\chi^2 = 10.74$
					$p = 0.023$	$p = 0.008$	$p = 0.000$
women's height			0.0620***	0.0606***		0.0489**	0.0523**
			(0.0141)	(0.0138)		(0.0172)	(0.0152)
rural = urban						$\chi^2 = 4.90$	$\chi^2 = 2.02$
						$p = 0.027$	$p = 0.155$
year FEs				✓			✓
n (DHS surveys)	117	117	108	108	117	108	108
R^2	0.369	0.201	0.540	0.643	0.270	0.423	0.548

Standard errors clustered by country in parentheses (65 countries in panel A, 59 in panel B)
p-values: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1.
χ^2 tests in columns 5, 6, and 7 test for equality with the coefficients in columns 2, 3, and 4, respectively.

Table 1.5: Open defecation linearly explains much of the South Asia-Africa height gap

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
sample	full	restricted	restricted	restricted	restricted	restricted	restricted	restricted	restricted
South Asia	-0.349***	-0.360***	-0.253**	-0.061	-0.424***	-0.488***	-0.129	-0.364***	0.029
	(0.0469)	(0.0521)	(0.0734)	(0.162)	(0.050)	(0.109)	(0.181)	(0.084)	(0.184)
percent explained			30%	83%			73%		92%
open defecation			-0.456*						
			(0.173)						
open defecators				-1.758*			-2.184*		-1.998†
per square km				(0.865)			(0.961)		(1.094)
population density					0.185**				
					(0.062)				
ln(GDP)						0.205	0.227†	0.077	0.068
						(0.122)	(0.123)	(0.084)	(0.082)
year FEs								✓	✓
n (DHS surveys)	100	89	89	89	89	89	89	89	89
R^2	0.390	0.391	0.497	0.433	0.416	0.495	0.727	0.786	0.800

Dependent variable is mean height-for-age of children under 3. Standard errors clustered by country in parentheses. p-values: $\dagger$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1. South Asia is a dummy indicator. The "restricted" sample is the set of observations with information on both height and open defecation. The "full" sample includes all observations from South Asia or sub-Saharan Africa.

Table 1.6: Change over time in a panel of Indian districts, NFHS-1 1992-3 to NFHS-2 1998-9

	(1)	(2)	(3)	(4)
Panel A: Repeated cross section (OLS)				
district open defecation	-0.779**	-0.279		
	(0.155)	(0.193)		
village open defecation		-0.537**	-0.523**	-0.774**
		(0.128)	(0.104)	(0.170)
village open defecation^2				0.134
				(0.367)
survey round fixed effect	✓	✓	✓	✓
age-in-months × sex	✓	✓	✓	✓
control variables			✓	
Panel B: District fixed effects				
district open defecation	-0.525†	0.106		
	(0.307)	(0.321)		
village open defecation		-0.553**	-0.353**	-0.710**
		(0.125)	(0.117)	(0.179)
village open defecation^2				-0.548
				(0.337)
survey round fixed effect	✓	✓	✓	✓
age-in-months × sex	✓	✓	✓	✓
control variables			✓	
n (children under 3)	23,588	23,588	23,588	23,588

Dependent variable is height-for-age z-score of children under 3. Standard errors clustered by district (across survey rounds) in parentheses. p-values: $\dagger$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1. Controls are at the household or child level: electrification, water supply, household size, indictors for being Hindu or Muslim, a full set of birth order indicators interacted with the relationship of the child's mother to the head of the household, twinship indicators, and month-of-birth indicators. "Linear predicted height change" multiplies the statistically significant coefficient on open defecation by 0.063, the change in rural open defecation between the NFHS-1 and NFHS-2, to make a linear prediction, based only on sanitation, of the change in height, which was about 0.022. Only rural subsamples are used.

Table 1.7: Village open defecation predicts child height, India and African DHS data

	(1)	(2)	(3)	(4)	(5)
Panel A: India (2005 NFHS-3, rural sub-sample)					
village open defecation	-0.305**	-0.289**	-0.257**	-0.217*	-0.216*
	(0.0644)	(0.0636)	(0.0579)	(0.0886)	(0.0879)
village open defecation^2		-0.208	0.0746	-0.111	-0.174
		(0.210)	(0.191)	(0.224)	(0.220)
household open	-0.413**	-0.413**	-0.0999**	-0.419**	-0.181**
defecation	(0.0356)	(0.0356)	(0.0380)	(0.0353)	(0.0376)
controls			✓		✓
state fixed effects				✓	✓
n (children under 5)	26,832	26,832	26,832	26,832	26,832
Panel B: five African countries (8 DHS surveys, rural sub-samples)					
village open defecation	-0.294**	-0.361**	-0.179**	-0.0551	0.00776
	(0.0575)	(0.0598)	(0.0599)	(0.0712)	(0.0696)
village open defecation^2		-0.726**	-0.572**	-0.569**	-0.523**
		(0.193)	(0.187)	(0.196)	(0.191)
household open	-0.0783†	-0.0767†	0.0161	-0.0804*	-0.000619
defecation	(0.0404)	(0.0403)	(0.0400)	(0.0404)	(0.0402)
controls			✓		✓
DHS survey FEs				✓	✓
n (children under 5)	44,216	44,216	44,216	44,216	44,216

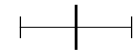
Dependent variable is height-for-age z-score of children under 5. Standard errors clustered by village (survey PSU). p-values: $\dagger$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Village open defecation is a fraction 0 to 1; household open defecation is an indicator 0 or 1. Controls are at the household or child level: 120 age-in-month by sex indicators; indicators for household dirt floor, access to piped water, electricity, TV, bicycle, motorcycle, and clean cooking fuel; and mother's literacy, knowledge of oral rehydration, age at first birth, count of children ever born, and relationship to the head of the household.

Table 1.8: Decomposition of rural India-Africa height difference: Fraction due to open defecation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
covariates:	none		birth demography		socio-economic		IMR (village-level)		mother's height	
sanitation:		✓		✓		✓		✓		✓
n (children)	71,048	71,048	71,048	71,048	71,048	71,048	71,048	71,048	70,596	70,596
Panel A: Pooled OLS regression, with an indicator for the Indian sub-sample:										
India	-0.142	-0.001	-0.191	-0.049	-0.301	-0.233	-0.243	-0.099	0.134	0.262
	(0.026)	(0.026)	(0.026)	(0.027)	(0.028)	(0.030)	(0.026)	(0.028)	(0.027)	(0.027)
village open		-0.480		-0.432		-0.177		-0.418		-0.420
defecation		(0.035)		(0.035)		(0.036)		(0.035)		(0.034)
R^2	0.0015	0.0090	0.0073	0.0142	0.0308	0.0318	0.0059	0.0124	0.0271	0.0337
explained:	0.994		0.754		0.225		0.593		not applicable	
India equal:	$\chi^2_1 = 136$		$\chi^2_1 = 130$		$\chi^2_1 = 24$		$\chi^2_1 = 120$		$\chi^2_1 = 136$	
Panel B: Oaxaca-Blinder decomposition adding linear village-level sanitation, unexplained difference:										
difference	-0.142	0.024	-0.203	-0.045	-0.316	-0.237	-0.258	-0.082	0.149	0.297
	(0.026)	(0.026)	(0.025)	(0.027)	(0.028)	(0.030)	(0.027)	(0.029)	(0.027)	(0.027)
explained:	1.169		0.781		0.250		0.682		not applicable	
Panel C: Non-parametric reweighting decomposition, counterfactual difference in means:										
difference	-0.142	0.061	-0.242	-0.022	-0.367	-0.186	-0.154	-0.005	0.215	0.396
change:	0.203		0.220		0.181		0.149		0.181	
explained:	1.43		0.908		0.493		0.967		not applicable	

"India" is an indicator for Indian observations, and "India equal" tests that the coefficient on the indicator is equal with and without the control for open defecation.

Table 1.9: What fraction of international height differences can open defecation explain?

				fraction of Indian height gap "explained"	
estimation strategy	source	slope	0 1 2 3 4 5	Africa (8 DHS)	U.S.
India's TSC, high	S (2012)†	1.592		3.542	0.440
India's TSC, low	S (2012)†	1.153		2.565	0.319
experiment, OLS	H&S (2012)†	0.786		1.748	0.217
experiment, TOT IV	H&S (2012)†	2.928		6.514	0.810
DHS means, FEs	Table 1	0.962		2.140	0.266
DHS means, controls	Table 1	1.111		2.471	0.307
DHS means, SA&SSA	Table 5*	0.721		1.604	0.199
India, district FEs	Table 6	0.533		1.185	0.148
India and Africa	Table 8	0.420		0.934	0.116

† These are probably overestimates of the effects of physical sanitation infrastructure because the programs they study also motivate use of existing latrines. DHS measures study latrine use. *The "Table 5" estimate is not, in fact, from table 5, but is found by estimating the regression in column 3 of table 5 without the South Asia indicator variable. Error bars in the fourth column plot 95 percent confidence intervals for the estimate of linear difference in height-for-age standard deviations associated with a 0 to 1 difference in the fraction openly defecating. The fourth column divides the point estimate by 0.45.

Chapter 2

Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from a Local Governance Incentive in India

2.1 Introduction

According to joint UNICEF and WHO (2012) estimates for 2010, 15 percent of people in the world defecate in the open without using 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, and to learn about effective strategies for motivating local government agents to implement such programs.

Well-identified evidence on the effectiveness of sanitation policy tools available in poor countries remains absent from the literature.[1] Past studies have focused on water supply, or have ignored the negative externalities that make open defecation a local public bad. Moreover, economists are increasingly understanding the importance of agency problems that limit state capacity (Banerjee, 1997; Niehaus et al., 2013). Could a rural sanitation program that is feasible to a poor country government improve human capital accumulation and health production?

This chapter estimates effects of India's Total Sanitation Campaign (TSC) on infant mortality and on children's height. This campaign represents a large effort to motivate local governments 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. In contrast, the TSC was implemented by the Indian government by local agents of central principals; many of the estimates presented will be representative of rural India. Thus, this chapter studies a full-scale program, and its

[1]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.

benefits and costs reflect real implementation, achieved using a novel *ex post* incentive mechanism designed for the constraints facing a low-capacity state.

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.

This chapter 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 infant mortality 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. The chapter documents externalities, showing that effects spill over onto children in households that do not use latrines. Second, this chapter 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 novel, successful governance mechanism, consistent with economic reasoning: the findings suggest that *ex post* incentives may motivate government agent performance in weak states and that incentivized local leaders can promote technology adoption. Finally, this chapter 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. Poor sanitation may represent an important constraint to global human capital, especially in South Asia.

Estimating causal effects of infrastructure is always challenging, given possibly endogenous construction (Dinkelman, 2011); this chapter combines converging evidence of causality from two approaches to identification. The first exploits heterogeneity in the timing of program implementation across districts. In order to combine 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 estimate 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. Finally, a further estimate applies a difference-in-differences 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 second strategy exploits a monetary prize offered by the Indian government to village officials for successfully implementing the program. The importance of this strategy derives from what it reveals about the mechanisms of the program, and from the more general lesson that *ex post* incentives can improve state capacity. The design of the prize rules created an incentive that was 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 and child stunting 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.

2.1.1 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 chapter is the first study of which I am aware to present well-identified estimates of the effects of public investments in rural household latrines in a poor country.[2] This chapter differs from related studies of water or sanitation and health 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, and can last many years (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

[2]Earlier papers have studied rural sanitation (Esrey et al., 1991), but much of this literate predates recent econometric emphasis on "design-driven studies" 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).

India, so-called "dry latrines" are common: corners or concrete slabs in or near homes where people defecate on the ground or 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 chapter 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.[3] 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

[3]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.

health effects beyond the households that practice it. Other evaluations of rural sanitation have focused 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, see (Bennett, 2012). The independent variable in this chapter 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.

2.1.2 India's Total Sanitation Campaign

The Total Sanitation Campaign was one of seven "national flagship programs" of the Government of India.[4] The TSC encouraged 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. In 2012 the TSC was replaced with a new approach to sanitation policy, the Nimral Bharat Abhiyan.

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 important features: it offered an *ex post* incentive to local government agents for achieving the desired outcome, and it made use of existing village social structure.

[4]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.

In October of 2003, the Indian government announced the Clean Village Prize (NGP in Hindi initials, for Nirmal Gram Puraskar), an incentive for villages (throughout this chapter, meaning Gram Panchayats) that achieve "open defecation free" status. When a village's chairman decides the village is eligible, he or she 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.[5] 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 chapter'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 important source of variation in program intensity in section 2.3.

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 (here, requiring latrine construction *and* use) 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. Yet, 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.

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

Unlike these incentives, however, the TSC provided an incentive to local governments, not households.

Although the NGP incentive money is nominally intended to be spent on village development, there is much evidence of local elite capture of 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 informally motivate the rest of the community.[6] 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 social "shaming" treatment, latrine ownership increased from 6% to 32%, but over the same period there was no increase in ownership in control villages. This can be seen as an independently estimated "first stage" that increases the plausibility of this chapter's approach.

[6]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).

2.1.3 Outline and data sources

This chapter combines several sources of individual and district survey data and district census data with TSC rules and administrative records to estimate effects of the TSC using two complementary identification strategies. Table 2.1 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 2.2, identifies an effect of the TSC on mortality and child height using heterogeneity in program timing across districts. Section 2.2.2 investigates the 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 this cross-sectional birth history, allowing panel data methods. Section 2.2.4 approximately replicates the within-district result with a long 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 will not be available for several years. 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 2.2.4 presents a further replication, using measures of TSC intensity independently collected by UNICEF in 47 districts of Uttar Pradesh, India's largest state, and finds similar results. Section 2.2.5 uses the India Human Development Survey to estimate an effect of early life sani-

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

tation on height-for-age, and important indicator of child human capital, using the same individual-level identification strategy as section 2.2.2.

Section 2.3 introduces the second identification strategy. In this section I use a discontinuity in the NGP incentive for local governments in order to construct a measure of predicted variation in district-level TSC intensity based on village-level 2001 census data. These 2001 census populations were measured prior to 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. Similarly, in survey data collected in 2010-11, stunting rates were lower in districts where the discontinuity created a greater incentive for local governments to improve sanitation. Finally, section 2.4 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.

2.2 Evidence from program timing: Fixed effects and changes

Did children exposed to better sanitation early in life experience better health? Because very early life is a critical period for determination of child height, and because infant mortality is defined to occur only in the first year of life, the disease environment most critically shapes these health outcomes during this short window of time. This section follows a growing economics literature on early life health in that it exploits variation in the timing of TSC implementation across districts to identify an effect on mortality and human capital accumulation.

2.2.1 Administrative records on TSC implementation

As its key independent variable, this section of the chapter 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 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/. Although section 2.2.4 will demonstrate that the key results are robust to alternative measures of program intensity that are independent of the TSC, most of this section estimates results of latrine coverage as reported by the TSC (the discontinuity-based estimates in section 2.3 make no use of government administrative records). These records likely present 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 construction of school latrines. It is an overestimate if bureaucrats inflate construction counts (*cf.* Imbert and Papp, 2011). I assume only that the true intensity of the TSC is, on average, 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 2.4).

A "process evaluation," rather than an "impact evaluation," would conclude that the TSC was far from completely implemented throughout India. Indeed, this heterogeneous rollout is part of what permits identification of causal effects. 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 2.3. However, implementers at

the village, district, NGO, and federal level have all insisted to me that a 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 section 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 attractiveness of the incentive to many local government officials.

2.2.2 Individual-level infant mortality: Empirical strategy

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.

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.[9] Each infant is matched to the number of TSC latrines per rural person

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

[9]Note that infants whose mothers die are likely to be excluded from the data, but it is very unlikely that there would be a biasing effect of the program here, as adults do not often die of ordinary diarrhea. In any event, there is little evidence of the program changing the characteristics of real or recalled fertility:

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

$$IMR_{idt} = \beta\, TSC_{dt} + \underbrace{D_i\gamma + H_i\zeta + M_i\mu + P_i\phi}_{\text{individual-level controls: } X_i\theta} + \alpha_t^{years} + \delta_d + \varepsilon_{idt}, \tag{2.1}$$

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.

the program does not predict the sex of the infant ($t = 0.10$), nor the mother's age ($t = 0.56$), nor whether the infant's household has a ration card for the poor ($t = 0.23$).

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).

2.2.3 Individual-level infant mortality: Results

Results are presented in Table 2.2. 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[10] scaled effects: latrine construction equivalent to one pre-program standard deviation is associated with a decline in IMR

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

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.[11] These effects remain quite similar with the full set of controls, as well as with state-year fixed effects or village fixed effects.

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. Additionally, although not reported in the table due to possible over-controlling, if district-specific linear time trends are added to equation 2.1'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. This makes clear that the result is not merely due to declining infant mortality over time or other district trends.

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.

Program intensity is measured per rural capita because what ultimately matters is whether every people'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

[11] 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.

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.

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 can never be directly tested, I replicate the estimation of equation 2.1 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. Note that this is also evidence against any mechanical effect of the mere passage of time, which would have also been present here.

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 2.2 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 2.3 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.

The individual-level DLHS data permit interactions indicating for whom TSC latrines were more effective. 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 2.1 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. A possible concern here is an omitted *interaction* variable between TSC intensity and something correlated with breastfeeding. However, when further interactions are added, one between TSC intensity and the DLHS wealth index and the other between TSC intensity and mother's literacy, the interaction does not change (from -74.84 to -77.47, $t = -2.37$) and the additions do not improve the fit. Column 8 shows a similar result for the IMR, instead of the PNNMR.

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.

To test this hypothesis, I added an interaction between TSC intensity and district rural population density to the basic specification of equation 2.1. 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, $s.e. = 0.0073$). Therefore, 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.[12]

[12]In contrast, the program is *not* more effective where the water source is closer (interaction $t = -0.76$). This is not consistent with poorly-identified evidence from the epidemiology literature suggesting that water and latrines are complements (Esrey et al., 1992), but is no surprise given that these are pit latrines, not flush toilets, and require very little water (Franceys et al., 1992), essentially only a cup's worth for

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 2.2 and 2.3, 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.

Although not presented here, I have produced 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 2.2 and 2.3. 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; the estimate of correctly timed construction is essentially unchanged. The results are similar (the coefficients on the lead and lag are small and not statistically significant) if the lead and lag independent variables are not included with the properly timed independent variable in the same specification, but instead are entered separately, each in their own regression.

hygiene. The result suggests, encouragingly for sanitation policy, that it is plausible that the TSC could have therefore been helpful even where water is scarce; I thank Abhijit Banerjee for this observation.

Externalities of sanitation

One reason that open defecation is an economic concern and sanitation is a public issue is negative externalities: one household's open defecation harms children living in other households. This is detectable in the data in the lack of an interaction between a household's own open defecation status and the district level sanitation environment variable that has been used throughout this section. If latrines only had private benefits, then we would expect the benefits of the TSC only to accrue to households that used them; in contrast, if sanitation were a public good, then all children should benefit.

Indeed, a household's own sanitation behavior does not interact with the effect of the TSC. In this rural sample, 74% of the infants studied live in a household that practices open defecation. However, when an indicator for open defecation is included in equation 2.1 along with an interaction with TSC intensity (that is, when this interaction is added to column 1 of table 2.2) the interaction has a very small coefficient of 0.86 (compared with a main effect if -87) and is not statistically significant ($t = 0.02$). Thus, it is no surprise that if the sample is restricted to children living in households without latrines, the estimate of the effect of the TSC on infant mortality is statistically indistinguishable to the effect in the full sample: -104 ($t = -2.10; n = 123,017$). These results provide evidence for the spillover effects of open defecation and the economic rationale for sanitation as a public good.

2.2.4 District-level infant mortality: long difference-in-differences

Section 2.2.2 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. Then, these results are approximately replicated for one state with data on TSC intensity from another source.

Empirical strategy

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 2.4 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 estimates the same coefficient with more conservative inference (Bertrand et al., 2004). The main specification is

$$\Delta IMR_d = \beta_0 + \beta_1 \Delta TSC\ \ latrines_d + \Delta X_d \theta + W_d \omega + \varepsilon_d, \tag{2.2}$$

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. As support for this specification, note that program intensity is not correlated with 2001 pre-program infant mortality ($t = 0.58$).

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

District-level results

Estimates of equation 2.2 are presented in columns 1 through 4 of table 2.4, 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.

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

Falsification tests

Table 2.5 presents three further opportunities to falsify the causal identification of the TSC's effect. Panel A repeats the regressions from table 2.4 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.[14]

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 an online government database similar to 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

[14]Note that table 2.3 already demonstrated that the apparent impact of the TSC is concentrated on *post-neonatal* mortality. Although the Annual Health Survey does not divide IMR into components, JSY – the main other change in government health policy during this time – would be expected to change *neonatal* infant mortality, if it had any effect at all.

government's online monitoring system. There are no statistically significant "effects" of NREGA on infant mortality.

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.[15] 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 2.6 presents estimates of equation 2.2 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.

2.2.5 Height and stunting: Effects on human capital

This section has presented evidence that the TSC reduced infant mortality. The early life disease environment also has an effect on children who survive. Height is an important

[15] 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, none of whom reviewed these conclusions.

indicator because the same early-life conditions that help children grown to their physical potentials help them grow to their human capital potentials: 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 reported 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 2005 India Human Development Survey (IHDS), matched with the same district-level time series of latrine construction used in section 2.2.2.

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

$$z_{idt} = \beta TSC_{dt} + (A_i \times sex_i)\,\Gamma + H_i\theta + M_i\vartheta + S_i\psi + \alpha_i^{years} + \delta_d + \varepsilon_{idt}, \tag{2.3}$$

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.

Results

Table 2.7 reports estimation results.[16] 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.[17] Within this sample, 1,311 households have

[16]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.

[17]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

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.[18] 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.

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 a mean of 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 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 2.7 may present *underestimates* of the effect of the TSC on survivors' heights. Section 2.2.2 found that

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.

[18] With the caveats stated in section 2.2.2, 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.

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.

2.3 Evidence from a discontinuity in a local governance incentive

As section 2.1.2 described, chairmen of village governments 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. Narrowly, this section can be seen as a robustness verification of the fixed effects results, studying a different dimension of heterogeneity and data that is not based on administrative records. More importantly, this section provides evidence on the program mechanism, and on a strategy to motivate local government agents to deliver public services in states with low capacity. Qualitative evidence from field visits indicates that the NGP was indeed implemented approximately as described here.

2.3.1 A discontinuity in the Clean Village Prize

Table 2.8 reprints the NGP rules, reporting the incentive offered in each population interval. There are four points where the incentive discontinuously increases. Most villages have populations below 2,000, so the 1,000 discontinuity is by far the most relevant. This

discontinuous incentive was devised solely for this program, and these cut-points are not associated with any other distribution scheme. Villages are not in competition: in principle, they could all qualify for the prize. The cash 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.

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.[19] Note also that it does not matter if the 2001 census population count was, in fact, an incorrect count, because it is nevertheless exactly this number that is used to determine the prize amount.

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

$$incentive(population) > \frac{cost(population)}{probability(population)}, \tag{2.4}$$

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

[19]Unsurprisingly, the density of 2001 village population is smooth through the discontinuities: in the full census, 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 ($F = 0.18$, $p = 0.67$).

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. Therefore, the program would be expected to produce healthier children in places just above the cut points.

2.3.2 First stage: Local governments responded to the incentive

Did a larger incentive per capita indeed lead to a larger response from local governments? Figure 2.1 presents evidence that it did, plotting the fraction of villages who won the prize for being open defecation free against the village size. The figure is based on data from the universe of villages in Rajasthan, a state in north India. Rajasthani data was used simply because data from the 2001 census, administrative records on whether a village won the NGP by 2012, and the TSC's baseline survey all happened to be available and able to be matched.

Figure 2.1 focuses on the discontinuity at 1,000 (this is the 76th percentile of the Rajasthani village population distribution; the next jump at 2,000 is uninteresting because it is at the 99th percentile). The dots plot means of population bins 25 wide; lines are linear, quadratic, and local polynomial regressions separately on each side of the discontinuity, and the shaded area is the 95% confidence interval for the local polynomial regression. The discontinuity is visible: larger villages are less likely to win, but the probability jumps up where the prize doubles.

Table 2.9 reports this result as a regression discontinuity of the form

$$\begin{aligned} winprize_{vd} = \; & \beta over1000_{vd} + \\ & \textstyle\sum_{p=1}^{4} \left(\alpha_p^{\ell} population_{vd}^{p} + \alpha_p^{r} population_{vd}^{p} \times over1000_{vd} \right) + \\ & X_{vd}\theta + \delta_d + \varepsilon_{vd}, \end{aligned} \tag{2.5}$$

where $winprize_{vd}$ is an indicator for having won the Clean Village Prize by the time the data were compiled in the summer of 2012; $over1000_{vd}$ is an indicator for a population greater than 1,000; β is the coefficient of interest; δ_d are district fixed effects; and X_{vd} is a set of control variables available with these data: baseline (before TSC) latrine coverage, fraction Scheduled Caste, fraction Scheduled Tribe (all three as quadratic polynomials), an indicator that the village is all Scheduled Caste, and the fraction of the village with Below Poverty Line cards. Adding controls and changing the degree of the running polynomial changes the precision of the estimate, but in all cases there is a similar, statistically significant discontinuity.[20] Because the jump in the prize amount at 1,000 is 50 rupees per capita, this suggests a local effect of a 2 to 3 percentage point increase in winning the prize that results from a 50 rupee per capita increase in the prize.

Of course, villages that did not win the prize may nevertheless have improved sanitation in the attempt, and even winning the prize does not necessarily be a guarantee that a village is open defecation free. Can we similarly detect an effect of the discontinuity on sanitation coverage? Unfortunately, the TSC's village level administrative data on latrine construction are of low quality where they exist, and in some cases are altogether absent. For example, the central government overwrites its records on past year village-level data when new data is received, which happens more often for some states than others. Results not reported here but available on request examine these village-level administrative data in what more detail is possible, and document some further evidence from data beyond the Rajasthani data discussed above that the NGP did motivate village chairmen more where effective incentives were greater.

[20]Standard verifications of RD validity are passed. For example, using placebo displaced discontinuities finds no effect: at 1,100 t-statistics of 1.33 and -0.58 for linear and cubic running variables, for example, and at 900, 1.55 and 0.51, respectively. A "donut RD" in the sense of Barreca et al. (2011) which omits villages from 975 to 1,025 estimates a similar discontinuity of 0.034, with a standard error of 0.011 and a t-statistic of 2.99.

2.3.3 Infant mortality

If local governments of villages with populations to the right side of the discontinuities were more motivated to improve sanitation, and if sanitation has an effect on child health, then one would expect improved health in these villages. Unfortunately, the NGP incentive operates at the village level, but IMR data is reported for districts, which are larger.[21] Therefore, this section matches district-level health outcomes with district-level measures of the strength of the prize incentive. The main results focus on a reduced-form approach, asking if health outcomes are associated with the incentive; section 2.3.3 documents that results are very similar to the fixed-effects estimates of the effect of latrine coverage if the effect on IMR is estimated using the incentive as an instrument.

Empirical strategy

I create this section's independent variables by constructing district-level variables to reflect the intensity of the incentive in each district, using 2001 census data on the population in each village.[22] 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. Thus the variable is the sum of the prize that each village could win, divided by the total population of the district. Note that even for districts of identical overall population, village count, and population per village, this could be very different if people are differently distributed to villages above or below the cut-points. There is a clear "first stage" effect of this variable on TSC latrine construction in the district administrative records. This variable is positively associated with district TSC

[21]The most recent health microdata for India dates to the 2008 DLHS, the 2005 IHDS, and the 2005 DHS.

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

construction (a 100 rupee increase in 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 ($t = -1.3$).

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.[23] 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.

[23]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.

Estimation

Table 2.10 presents estimates of district-level cross section regressions:

$$IMR_d^{2010-11} = \beta incentive_d + X_d\theta + \varepsilon_d, \tag{2.6}$$

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 2.2.4. 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 quantitatively 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 row, including the controls for that column: 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 statistically significantly associated with IMR with the predicted sign.[24] 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 down rows 2 through 4, as the bandwidth narrows on the discontinuity, and then from row 4 to row 5, as the incentive measure "flips" from an indicator of high average incentives to an indicator of low average incentives. When the

[24] 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 column of the table); with all three controls (compare with the last column) 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 rows 4 and 5 are very similar in absolute value. The reported p-values verify that the estimates in rows 4 and 5 are different, which would not be the case under the null hypothesis of no effect of the incentive or discontinuty. 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.

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 chapter'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 2.3.3 could then be considered the reduced form of this analysis which, again, would make no use of the TSC administrative records.

What can such an instrumentation strategy say? 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 linearly 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 2.2, which reports a similar estimate of 85 using the DLHS-3 and TSC administrative records. In contrast, there is no association if *urban* infant mortality is instead used as the dependent variable. The interpretation of this IV estimate, as always, is complicated by factors such as heterogeneity of treatment effects, but as a robustness check it is consistent with the rest of the evidence.

2.3.4 Height and human capital

Changes in the disease environment that impact infant mortality would also be expected to influence the heights of children exposed to them in early life. Did the discontinuity in the NGP incentive have an effect on child height? The available data constrain any ability to answer this question. The only large-scale child height data collected in India since the NGP has been implemented are in the HUNGaMA survey (Naandi Foundation, 2011). This survey was conducted as an advocacy tool by an Indian NGO from October 2010 and February 2011 in 3,360 villages across 9 states.[25] I use published district averages from the publicly available survey report. The data do not cover all districts, or even a random sample, but rather what the report identifies as "100 Focus Districts" at the bottom of a 2009 UNICEF child development district index, plus a few more selected by the organization, for a sample of 108 districts.

With such a small sample of districts, only a simple analysis is possible: estimating regression equation 2.6 of child stunting on the prize incentive per capita. The data do not report the average height-for-age of children surveyed in the district, but only the percent stunted (below -2 standard deviations) or severely stunted (below -3 standard deviations).

[25]However, most of the districts covered are from the poor northern states of Uttar Pradesh (37% of the sample), Bihar (21%), Jharkhand, (13%), and Madhya Pradesh (11%). I omit four surveyed districts – two in Kerala and two in Tamil Nadu – both out of necessity because the rest of the district-level data used throughout this chapter is from north India, and due to the fact that these would be unusual as the only two districts included in states (with many other districts) that are very different from the rest of the sample. The report specifically notes that these were included to be *unlike* the other districts (p. 8).

Such dichotomization is well-known to reduce statistical power (Royston et al., 2006), so any effect of the incentive on child height may be stronger than what is documented here.

Figure 2.2 depicts the main result of this section: stunting rates are lower after the program in districts where the incentive per capita is greater, at both reported levels of stunting. Table 2.11 verifies the statistical significance and robustness of this result. Controls verify that the correlation is not an artifact of baseline sanitation coverage or infant mortality, nor is it driven by female literacy, which emerges as a clear predictor of the HUNGaMA survey stunting data.[26]

2.4 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 2.2.2 and 2.2.4, then the TSC has eliminated just under one-fifth of those 20.75 baseline rural infant deaths that were due to fecal contamination.[27] There is a growing set of well-identified econometric estimates of impacts on infant mortality. Like this chapter, 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. This effect is of a plausible magnitude, given the literature. It is somewhat smaller than the effect estimated

[26]Because the data are a selected subset of Indian districts, a Heckman selection model may be appropriate. However, results are similar, and selection does not appear to be an issue. For example, implementing column 1 as a two stage Heckman selection model with the incentive per capita also used to predict inclusion does not change the result (from -0.222 in the table to -0.238 in the selection model, $z = 1.97$) and the incentive does not predict inclusion in the sample ($z = 0.07$).

[27]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.

by Galiani et al. (2005) of 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.[28] 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 chapter's are inclusive of all heterogeneity of administration and losses to corruption, under actual implementation at scale.

[28]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."

Latrine construction figures are almost certainly inflated, due to bureaucrats' incentives 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. A model available on request documents 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.

2.5 Conclusion

This chapter 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. Sanitation has positive externalities: an effect is seen even on children in households that do not adopt latrine use.

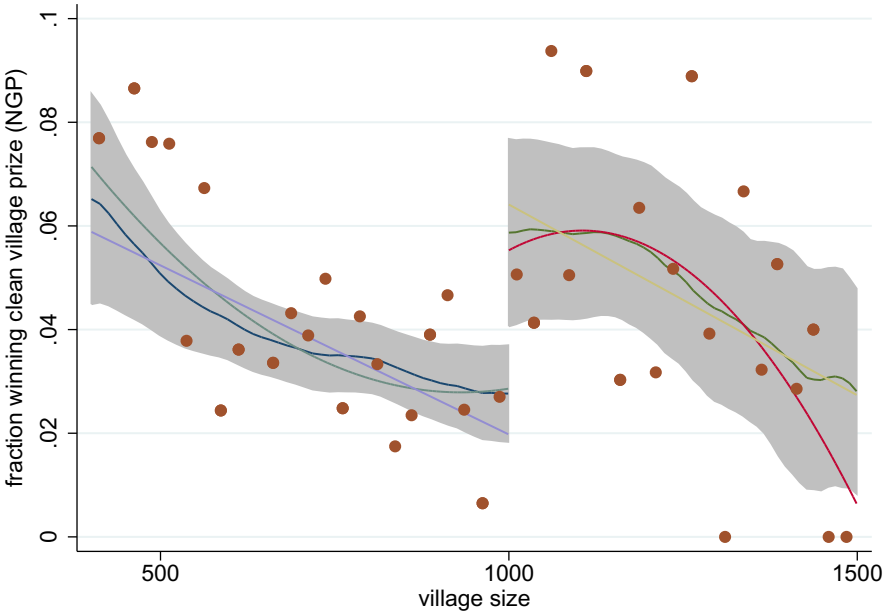
Beyond the TSC, this chapter shows that rural sanitation has an important contribution to make to children's health and that open defecation is a health hazard and public bad even in rural areas. These findings underscore the importance of chronic enteric infection to global poor children's health, and demonstrate the value of studying effects of fecal pathogens through reduced form health outcomes, in addition to or 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.

This chapter's results imply that as late as a decade ago, India had much room to improve its sanitation infrastructure, and likely does still. In India's 2005-06 DHS survey, 55.3 percent of households reported having no toilet or latrine.[29] As chapter one discussed, 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.[30] Yet, according to a linear extrapolation of the estimates in section 2.2.5, bringing India to Sierra Leone's 2008 23 percent open defecation rate 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.

[29] 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.

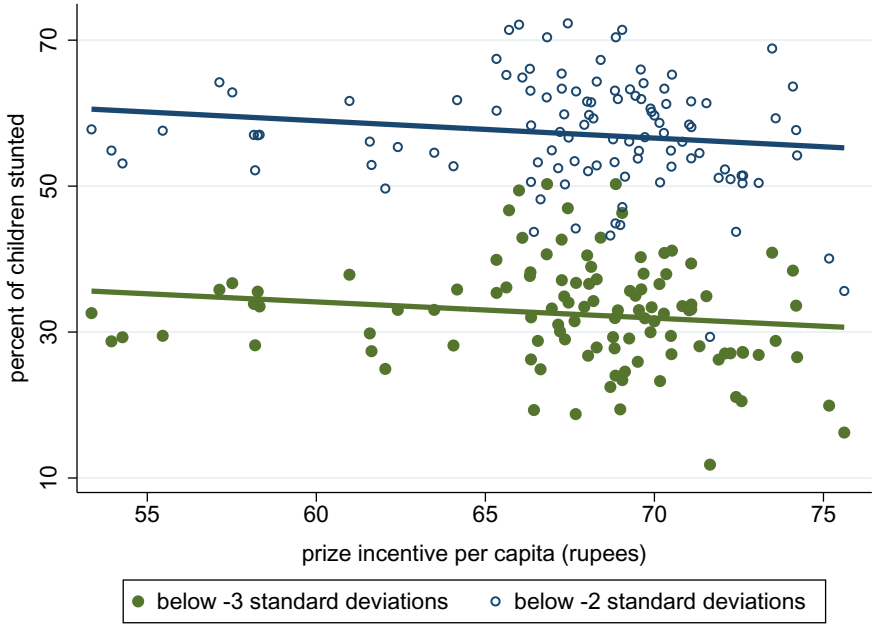
[30] For example, sanitation would not readily explain slower cohort-to-cohort growth in women's height than men's (Deaton, 2008).

Figure 2.1: Effect of discontinuity in incentive amount on sanitation prize, Rajasthan



Note: The figure is constructed with data on all villages in Rajasthan.

Figure 2.2: Effect of incentive discontinuity on district-level child stunting



Note: Stunting data from 2010-11 HUNGaMA survey.

Table 2.1: 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
children stunted (% 2010-11 HUNGaMA)	57.14	0.73
children severely stunted (% 2010-11 HUNGaMA, 108 districts)	32.43	0.70
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 2.2: 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 2.3: Mechanisms of the effect of the TSC on mortality, DLHS-3

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
mortality rate:	post-neonatal mortality rate (months 2-12)					neonatal (month 1)		IMR
TSC latrines	-50.01*	-48.12*	-51.51†	-47.17*	-48.21*	-35.69	-39.10	-76.30*
per capita	(24.04)	(24.33)	(29.60)	(23.55)	(23.63)	(28.74)	(27.25)	(34.43)
latrines × fraction after				-74.84*	-83.69*			-132.76*
non-breast milk food				(32.48)	(32.99)			(60.85)
fraction of first year after				-0.587	-2.734*			-5.486*
non-breast milk food				(1.310)	(1.358)			(2.446)
district fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
year fixed effects	✓	✓		✓	✓	✓	✓	✓
state × year fixed effects			✓					
controls		✓	✓		✓		✓	
n (infants)	164,795	164,762	164,762	164,795	164,762	164,795	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 2.4: 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				2.5×10^{-6}
				(2.1×10^{-6})
village count, 2001 census				-0.000900
				(0.00233)
popul. × village count				1.2×10^{-9}
				(1.1×10^{-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 2.5: 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 2.4.

Table 2.6: 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 2.7: 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: $\dagger$ $p < 0.05$. Clustered standard errors in parentheses. "Effect of program mean" multiplies the coefficient estimate by the mean of program intensity in these districts in 2005.

Table 2.8: 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 2.9: Clean Village Prize regression discontinuity, all Rajasthani villages

	(1)	(2)	(3)	(4)	(5)	(6)
	ever won Clean Village Prize by 2012, linear probability model					
over 1,000	0.0299**	0.0291*	0.0233*	0.0253†	0.0271†	0.0266†
	(0.0115)	(0.0114)	(0.0117)	(0.0131)	(0.0150)	(0.0160)
polynomial	quadratic	quadratic	cubic	cubic	quartic	quartic
controls		✓		✓		✓
district FEs		✓		✓		✓
n (Rajasthani GPs)	6,464	6,464	6,464	6,464	6,464	6,464

Standard errors clustered by 209 blocks in parentheses. Two-sided p-values: $* \; p < 0.05$ $** \; p < 0.01$, $* * * \; p < 0.001$. One-sided p-values: $\dagger \; p < 0.05$. FE = fixed effects. GP = Gram Panchayat. Controls are baseline (before TSC) latrine coverage, fraction Scheduled Caste, fraction Scheduled Tribe (all three as quadratic polynomials), an indicator that the village is all Scheduled Caste, and the fraction of the village with Below Poverty Line cards.

Table 2.10: Effect of NGP incentive on endline (2010/11) rural IMR, Annual Health Survey

	(1)	(2)	(3)	(4)
	Endline (2010-11) infant mortality			
prize incentive per capita	-0.209	-0.285	-0.277	-0.461
	(0.129)	(0.094)	(0.094)	(0.199)
fraction within 50% above cut	-28.72	-29.84	-30.05	-32.22
	(14.44)	(10.54)	(10.29)	(15.40)
fraction within 40% above cut	-33.95	-31.61	-32.22	-27.92
	(17.41)	(12.95)	(12.57)	(16.66)
fraction within 30% above cut	-42.22	-37.63	-39.73	-30.36
	(22.75)	(16.84)	(16.47)	(18.77)
fraction within 30% below cut	100.2	40.20	42.53	38.72
	(28.52)	(17.36)	(21.18)	(20.44)
30% above = 30% below	$p = 0.000$	$p = 0.001$	$p = 0.002$	$p = 0.013$
IMR, 2001		✓	✓	✓
sanitation, 2001			✓	✓
population per village				✓
n (districts)	280	280	280	280

Each estimate is from a separate regression of 2010/11 IMR on that row's independent variable and the checkmarked controls in its column. Heteroskedasticity robust standard errors in parentheses. The p-values after the coefficients reflect a test of the null hypothesis that the coefficients on 30 percent above and below are equal.

Table 2.11: Clean Village Prize incentive and child stunting, HUNGaMA districts 2010-11

	(1)	(2)	(3)	(4)	(5)	(6)
	percent severely stunted				percent stunted	
incentive	-0.222†	-0.412**	-0.344*	-0.340*	-0.238	-0.307*
per capita	(0.127)	(0.126)	(0.134)	(0.135)	(0.145)	(0.141)
IMR, 2001		0.206***	0.166**	0.171**		0.156**
		(0.0484)	(0.0507)	(0.0519)		(0.0507)
female			-0.222**	-0.248**		-0.266*
literacy, 2011			(0.0804)	(0.0859)		(0.108)
latrine				0.0595		0.0406
coverage, 2001				(0.0592)		(0.0581)
constant	47.49***	43.12***	53.04***	52.52***	73.24***	77.62***
	(8.452)	(7.914)	(9.635)	(9.387)	(9.593)	(10.89)
n (districts)	108	108	106	106	108	106

Heteroskedasticity robust standard errors in parentheses. Two-sided p-values: $*\ p < 0.05$ $**\ p < 0.01$, $***\ p < 0.001$. One-sided p-values: $\dagger\ p < 0.05$. Stunted: height-for-age below -2 standard deviations; severely stunted: below -3 standard deviations.

Chapter 3

Village sanitation externalities and children's human capital: Evidence from a randomized experiment by the Maharashtra government

with Jeffrey Hammer

3.1 Introduction

Indian children suffer some of the highest rates of average stunting in the world, with lifelong implications for health and human capital. Simultaneously, India leads the world in open defecation; over half of the population defecates openly without a toilet or latrine. Fortunately, prior non-experimental research indicates that improvements in rural sanitation that are feasible to the Indian government could importantly improve early life health.

Adding to this evidence base, this chapter contributes the first econometric analysis of a village-level rural sanitation program experimentally implemented in a randomized, controlled trial. We study a community sanitation program that was conducted by the government of Maharashtra, India in early 2004. The program sent a representative to promote or "trigger" community investments in latrine use, in the context of the Indian government's Total Sanitation Campaign.

Although the government of Maharashtra originally planned to implement the program in three districts, it ultimately implemented the experiment only in Ahmednagar district, randomizing within this district. We find that, where the experiment was implemented, the program was associated with an increase in average child height comparable to non-experimental estimates in the literature.

This chapter makes five contributions to the economics literature. First, we present a rigorous econometric analysis of a community-level sanitation experiment. Second, we offer the first causally well-identified evidence of sanitation externalities: effects were found even on children in households that never adopted latrine use. Third, we reflect on the implications of the fact that the government originally planned to implement the experiment in a larger set of villages than it did. Relatively few experiments highlight the selection of the group to be experimented upon, despite the fact that this selection could importantly shape resulting parameter estimates. Fourth, we note that this change in plans underscores the implementation constraints facing the Indian government and the many remaining gaps in rural sanitation coverage. Importantly we study implementation of an experiment by the government, not an NGO-academic partnership. Fifth, we demonstrate how use of non-parametric statistical inference can partially respond to Deaton's (2012) recent observation that outliers may inappropriately determine conclusions in small-sample field experiments in development economics.

This chapter also makes three contributions to policy debates, especially in India. The first regards the allocation of public funds in India: few prior studies have shown effects

of public policies on health status. That is, although some studies have shown impacts of interventions by NGOs or medical researchers, none have been implemented by the Indian government (Das and Hammer, forthcoming). Learning about the effects of "scaled up" programs implemented by the Indian government requires studying government implementation. The second notes that because sanitation has important externalities, it has a strong theoretical claim to public resources, empirically validated by this experiment, in a way that purely curative care sometimes may not. Finally, this chapter studies the height of children, which is widely agreed to measure the long-term net nutritional status of children. Indian children are exceptionally short by international standards, a major policy concern which has attracted the recent attention of many economists (*e.g.* Deaton, 2007; Tarozzi, 2008; Jayachandran and Pande, 2012; Panagariya, 2012). Stunting is often referred to as "malnutrition," which suggests to many that providing food is the appropriate policy response. Yet, net nutrition is a matter of food intake, of food absorption and use by the body, and of losses of nutrition due to disease. Diarrheal and other chronic intestinal disease can limit children's ability to absorb and use improved nutrition, and may be responsible for an important part of stunting among Indian children, as discussed in chapter one.

3.1.1 Open defecation is widespread in India

According to joint UNICEF and WHO (2012) estimates for 2010, 15 percent of people in the world, and 19 percent of people in developing countries, openly defecate without using any toilet or latrine. Of these 1.1 billion people, nearly 60 percent live in India, which means they make up more than half of the population of India. These large numbers correspond with the estimates in the Indian government's 2011 census, which found that 53.1 percent of all Indian households – and 69.3 percent of rural households – "usually" do not use any kind of toilet or latrine. In the 2005-6 National Family Health Survey, India's version of

the DHS, 55.3 percent of all Indian households reported defecating openly, a number which rose to 74 percent among rural households.

These statistics give several reasons to be especially concerned about open defecation in India. First, open defecation is much more common in India than it is in many countries in Africa where, on average, poorer people live.[1] Second, despite accelerated GDP growth in India, open defecation has not rapidly declined in India over the past two decades, not even during the rapid growth period since the early 1990s. In the DHS, where 55.3 percent of Indian households defecated openly in 2005-06, 63.7 did in the earlier 1998 survey round, and 69.7 did in 1992. In 2010, 86 percent of the poorest quintile of South Asians usually defecated openly.

3.1.2 Non-experimental evidence for effects of sanitation on health

We report the first, to our knowledge, econometric analysis of a randomized controlled experiment about the effects of a village-level[2] community sanitation program on child health.[3] In a review of evidence on rural water and sanitation interventions, Zwane and Kremer (2007) conclude that "many of the studies that find health effects for water and sanitation infrastructure improvements short of piped water and sewerage suffer from critical methodological problems" (10). Importantly, however, two existing literatures indicate that a large effect of sanitation is plausible.

First, medical and epidemiological literatures have documented the mechanisms linking open defecation to poor health and early life human capital accumulation. Checkley

[1] Chapter one noted that population density is also much greater in India than in Africa, providing more opportunities for contact with other people's fecal pathogens.

[2] Some prior evaluations of rural sanitation have focused 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). Such an approach would ignore negative externalities of open defecation.

[3] Lisa Cameron, Paul Gertler and Manisha Shah have presented preliminary results from an excellent sanitation experiment in progress in Indonesia that are quite complementary to our findings.

et al. (2008) use detailed longitudinal data to study an association between childhood diarrhea and subsequent height. Perhaps more importantly, Humphrey (2009) and Korpe and Petri (2012) note that chronic but subclinical "environmental enteropathy" – a disorder caused by repeated fecal contamination which increases the small intestine's permeability to pathogens while reducing nutrient absorption – could cause malnutrition, stunting, and cognitive deficits, even without necessarily manifesting as diarrhea (*see also* Petri et al., 2008; Mondal et al., 2011).

Second, chapter two reported a well-identified retrospective econometric study that found an effect of a government sanitation program in rural India.[4] Averaging over implementation heterogeneity throughout rural India, chapter two showed that the TSC reduced infant mortality and increased children's height. In a follow-up study, Spears and Lamba (2012) find that early life exposure to improved rural sanitation due to the TSC additionally caused an increase in cognitive achievement at age six, using a similar approach to identification.

We study a village-level program inspired by the Community-Led Total Sanitation movement (Bongartz and Chambers, 2009), implemented in the context of the government's TSC. Alok (2010), in his memoirs as an administrative officer responsible for the TSC, describes Maharashtra as an early and rapid adopter of the TSC. Among Indian states, Maharashta has the most villages which have won the Clean Village Prize for eliminating open defecation. Our data come from a study done early in the implementation of the TSC in that state.

3.1.3 Overview

The next section outlines our empirical strategy: analysis of a randomized, controlled experiment. Although the original decision of the Maharashtra government was to conduct

[4]Additionally, other studies find effects of large scale piped water investments on health, especially in the history of now-rich countries (Cutler and Miller, 2005; Watson, 2006).

an experiment in three districts, the experiment that occurred was confined to one district. Nevertheless, in all three districts there were three rounds of survey data collected, one before the experimental intervention and two after.[5] Section 3.3 presents and analyzes the results of the experiment. Section 3.4 discusses policy implications, considering treatment heterogeneity and the consequences of the government's original decision to experiment. Section 3.5 concludes.

3.2 Method: A randomized field experiment

The timeline of this experiment contained four events: the experimental intervention in early 2004 and three survey rounds.

- February 2004: baseline survey data collection
- shortly thereafter: community level sanitation "triggering" intervention
- August 2004: midline survey data collection
- August 2005: endline survey data collection

Therefore about 18 months elapsed between the experimental intervention and the final observations of outcomes.

3.2.1 The program: A community sanitation motivation intervention

The experimental program studied here was conducted in the context of the initial introduction of India's Total Sanitation Campaign by the Maharashtra government. In Maharashtra, as in other states in India, it is the responsibility of district government staff

[5]This unusual circumstance of data collected beyond the experimental panel was due to funding rigidities of large bureaucracies, in this case both the World Bank and the Government of Maharashtra.

to implement the TSC, and different districts pursued the program goals with different levels of intensity at different times. The TSC was a large government effort throughout India, which it is not the purpose of this chapter to evaluate (*see* chapter two). Instead, this chapter evaluates a modest experimental addition to the TSC in one district. Thus, whenever this chapter refers to "the program" studied, we mean only this special, randomized sanitation promotion intervention, and certainly not to the entire TSC. However, it is important to note that, because this experiment happened in the early days of broader TSC implementation, it occurred in a local context of minimal to no other sanitation policy activity, beyond in principle providing funds for latrine construction to village leaders who elected to draw upon them.[6]

The experimental program studied here is community-level sanitation motivation by a representative of the district government. Inspired by the procedures of the Community-Led Total Sanitation movement, the program sent a sanitation promoter to visit the village and convene a meeting where information, persuasion, demonstration, and social forces were employed in an attempt to "trigger" a community-wide switch to latrine use. For more details on the exact procedures of a sanitation "triggering," please see Bongartz and Chambers (2009). It is important to emphasize, however, that the program studied was not a traditional CLTS implementation because it also included government subsidies for latrine construction.

Is it plausible that such a motivational visit could have positive effects? "First-stage" evidence from other studies indicates that such an event can successfully change behavior. For example, in context of India's TSC, Pattanayak et al. (2009) find in a randomized, controlled trial in two blocks in a district of Orissa that in villages receiving a social "shaming" treatment, latrine ownership (and reported use) increased from 6% to 32%, but over the same period there was no increase in ownership in control villages.

[6]Presumably, in the absence of these available funds for latrine construction, the experimental program would have been much less successful; in this context, they were plausibly necessary but not sufficient for program success.

3.2.2 What was supposed to happen?: Three districts in Maharashtra

Districts are the administrative unit of the Indian government that make up states; districts, in turn, comprise blocks, which contain Gram Panchayats, which we will often call "villages." When the government of Maharashtra initially decided to conduct this experiment, it selected three districts: Ahmednagar, Nanded, and Nandurbar. Randomization would occur separately stratified within each district to assign 60 villages to treatment and control groups, with 30 villages each in each district.

Table 3.1 compares the three districts with average properties of rural Maharashtra and all of rural India. In general, Ahmednagar and Nanded look similar to one another, while Nandurbar appears poorer and has a larger Scheduled Tribe population. Ahmednagar has better sanitation coverage than Nanded and Nandurbar, but these figures are difficult to interpret because they are from after the program studied was implemented. These districts were chosen because Ahmednagar and Nanded district officials requested early implementation of the TSC at a state level workshop in 2002; a state official selected Nandurbar so that a particularly poor district would be included.

3.2.3 What did happen: An experiment in one district

Although the government of Maharashtra originally planned to implement an experiment in three districts, in fact, it ultimately only implemented the experiment in Ahmednagar. In this district, the program was indeed implemented in 30 villages randomly selected out of 60 eligible for the treatment or control groups. As table 3.1 shows, Ahmednagar has better average sanitation coverage than the other two possible districts, and is otherwise similar to Nanded and less poor than Nandurbar.

Due to some confusion (later resolved) about whether the experiment would be implemented in all three planned districts, the World Bank had already contracted with a survey

organization to collect data in all three districts. Therefore, the data collection continued in all three districts, as originally planned.

This change of government plans – and seemingly unnecessary data collection – presents an unusual econometric opportunity in the analysis of a field experiment. One important conclusion is already clear, even before any statistical analysis: there are important limits to the ability of the Indian state to translate decisions from high-ranking officials into activities, programs, and services in villages.[7]

3.2.4 Empirical strategy

The empirical strategy of this chapter is built upon the random assignment of villages to treatment or control groups. However, a randomized, controlled experimental intervention only occurred in Ahmednagar district. It would be difficult to know how often planned field experiments are canceled,[8] but our case is unusual among these because data were still collected about the originally intended sample.

How does this change our econometric strategy? Importantly, randomization happened *within* districts. Sixty villages in each district were identified as eligible for randomization, and of these 30 each were randomly assigned to treatment and control groups using pseudo-random number generator functions in Microsoft Excel, in a different "worksheet" spreadsheet page for each district. This means that an experiment occurred in Ahmednagar independently of whatever happened in the other two districts.

[7]For this reason, Pritchett (2009) has described India as a "flailing state": the head cannot send signals to the hands and feet. While important for sanitation policy, which has yet to reach many of the rural areas where open defecation remains nearly universal, constraints on effective implementation exist for all health related policies. For example, problems of absenteeism and low-quality of publicly provided health services are widespread, as well.

[8]Soon after this project a similar problem of lack of or faulty implementation that led to a similar disconnect of intervention and evaluation affected a World Bank project in Karnataka, India. Contrasting this pattern of difficulty evaluating *government* programs with the success of experimental partnerships between researchers and NGOs illustrates Ravallion's (2012) observation that "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).

Therefore, we can produce internally valid estimates of causal effects in Ahmednagar, in exactly the same way as if the experiment had only ever been intended to occur there. Of course, the estimated effect may not have external validity for or be "representative" of the average effect across all of Ahmednagar, Nanded, and Nandurbar. However, it arguably would not have been anyway: the 60 villages in each district were not randomly sampled from the set of all possible villages. Even if they had been, these three districts were not chosen randomly from the set of all districts in Maharashtra or India. Indeed, two were volunteered by interested district officials, and one was chosen for its remoteness and difficulty to work in by political advocates of community sanitation projects, exactly in order to demonstrate that such sanitation programs can be widely successful.[9]

So, the effect of the planned experiment that did not occur in Nanded and Nandurbar is to change the set of villages eligible to be randomized into the treatment or control groups. Deaton (2012) observes that econometric theory appears to have no ready name for this set, and suggests it be called the "experimental panel." Because we can still produce internally valid results for Ahmednagar, our analysis will focus on this district. However, we will replicate our time-by-treatment group difference-in-difference results in this district with triple difference estimates using Nanded and Nandurbar as a counterfactual. Additionally, we will verify that there was no "treatment" effect of randomization in these untreated districts.

Effects on child height

Physical height has emerged as an important variable for economists studying development, labor, or health (Steckel, 2009). Height is a persistent summary measure of early life health; early-life height predicts adult height (Schmidt et al., 1995), as well as human capital and economic productivity (Case and Paxson, 2008; Vogl, 2011). Puzzlingly, South Asian people

[9]It should be noted that such choices are usually motivated by the opposite concern: to place programs in the most favorable circumstances. It showed quite a bit of courage to tackle the hardest cases first.

are much shorter than their income would predict (Deaton, 2007; Ramalingaswami et al., 1996): they are, for example, shorter on average than people in Africa who are poorer, on average. Moreover, children's height is much more steeply correlated with cognitive achievement in India than in the U.S., suggesting more depth and greater variance of early life disease (Spears, 2012). Medical and epidemiological evidence, as well as chapter one's econometric decomposition, suggests that widespread open defecation could be an important part of the explanation for Indian stunting (Humphrey, 2009).

Height of children under 5 is, therefore, the central dependent variable in our analysis.[10] In particular, we study the height of all children under five in a randomly selected 75 percent of households in the villages surveyed. This age group is the focus of WHO growth reference charts, and is a commonly selected population in height studies. In our main results, we transform height into z-scores using the 2006 WHO reference population,[11] but we will show that our results are robust to using log of height in centimeters – unrelated to an external standard – as the dependent variable instead.

Our preferred specification is a difference-in-differences at the individual child level, using only data from Ahmednagar district:

$$\begin{aligned} z_{ivt} = \quad & \beta_1 treatment_v + \beta_2 treatment_v \times midline_t + \beta_3 treatment_v \times endline_t + \\ & A_{ivt}\Gamma + \alpha_v + \gamma_t + \varepsilon_{ivt}, \end{aligned} \tag{3.1}$$

where i indexes individual children, v indexes villages, and t indexes the three survey rounds: baseline, midline, and endline. The dependent variable z is the child's height-for-

[10] Although many epidemiological studies use survey-reported diarrhea as a dependent variable, we make no attempt to study this noisy and unreliable outcome measure (Schmidt et al., 2011). For example, Zwane et al. (2011) show that households randomly selected to be surveyed more frequently report less child diarrhea. More broadly about survey reported morbidity, 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. Finally, Humphrey's (2009) evidence of height shortfalls due to chronic enteropathy indicate that diarrhea is not an indicator of or necessary condition for losses in human capital.

[11] The international reference population used to create these standards includes a population of Indian children raised in healthy environments in south Delhi. Such Indian children grow, on average, to the WHO reference mean heights (Bhandari et al., 2002).

age z score, $treatment_v$ is an indicator for living in a village assigned to the treatment group (notice it is only indexed by village), and $midline_t$ and $endline_t$ indicators for survey round. Survey round fixed effects γ_t will always be included, and to this a set of 120 age-in-months-times-sex indicators A_{ivt} and village fixed effects α_v will be added in stages to demonstrate that they do not change the result. We replicate the result using a similar specification

$$z_{ivt} = \beta_1 treatment_v + \beta_2 treatment_v \times after_t + A_{ivt}\Gamma + \alpha_v + \gamma_t + \varepsilon_{ivt}, \tag{3.2}$$

where the $midline_t$ and $endline_t$ indicators have been collapsed into the single variable $after_t$, which is 1 for observations in the midline or endline survey round and 0 for observations in the baseline survey round.

As a further robustness check – exploiting all of our data and the fact that no experimental intervention occurred in Nanded or Nandurbar – we use a triple difference, comparing the effect of random assignment to the treatment group in Ahmednagar to the effect of assignment in the unexposed districts. Thus, we estimate

$$\begin{aligned} z_{ivt} \;=\; & \beta_1 treatment_v + \beta_2 treatment_v \times midline_t + \beta_3 treatment_v \times endline_t + \\ & \beta_4 treatment_v \times Ahmednagar_v + \beta_5 treatment_v \times midline_t \times Ahmednagar_v \\ & + \beta_6 treatment_v \times endline_t \times Ahmednagar_v + A_{ivt}\Gamma + \alpha_v + \gamma_t \times \delta_v + \varepsilon_{ivt}, \end{aligned} \tag{3.3}$$

and

$$\begin{aligned} z_{ivt} \;=\; & \beta_1 treatment_v + \beta_2 treatment_v \times after_t + \\ & \beta_4 treatment_v \times Ahmednagar_v + \beta_5 treatment_v \times after_t \times Ahmednagar_v + \\ & A_{ivt}\Gamma + \alpha_v + \gamma_t \times \delta_v + \varepsilon_{ivt}, \end{aligned} \tag{3.4}$$

where β_5 and β_6 (where present) are the estimates of the experimental effects. In these equations, $\gamma_t \times \delta_v$ are eight indicators for the district-by-survey round interaction cells, always included.

Because the experimental treatment was assigned at the village level, in all regression estimates we calculate standard errors clustered by village. Even in the Ahmednagar-only samples, there are 60 villages, which exceeds Cameron et al.'s (2008) threshold of 50 clusters for reliable standard errors. However, our sample is small if *villages* are thought of as the independent unit of observation. An outlier village could have a outsized effect on our results (Deaton, 2012).[12] To verify our results, we collapse our data into 60 observations (for example, the change in mean height in each village) and perform non-parametric statistical significance tests – in particular, tests based on rank of observations, and not absolute magnitude. Of course, this only protects from outliers binary inference about a null hypothesis of no effect, and not the distribution of average treatment effect estimates. Additionally, we replicate our main result omitting each village in turn.

Effects on the Clean Village Prize

As part of its Total Sanitation Campaign, the Indian government awards villages a Nirmal Gram Puraskar — Hindi for Clean Village Prize — in recognition for becoming open defecation free. Villages certified by government auditors to be open defecation free receive a trophy and a cash prize, presented to the village chairman at a prestigious ceremony in the state or national capital. Although only about 4 percent of all Indian villages have won the prize, this number is much larger in Maharashtra, where over 9,000 prizes have been won, more than any other state and, indeed, about one-third of the total number of prizes awarded. For more information on the Clean Village Prize, please see chapter two.

[12]For example, we could estimate a large average effect merely because one village where the potential program effect was large happened to be in the treatment group rather than the control group.

We treat receipt of the Clean Village Prize as measure of village sanitation coverage that is independent of our data collection, if coarse and noisy. We obtained administrative records from the Indian central government on which villages in Ahmednagar had ever won the Clean Village Prize. Prizes were first awarded to any of the villages we study in 2006. Through the summer of 2012, 12 of the 60 villages studied have won the prize. We will use regression, Mann-Whitney, and Fisher exact statistical significance tests to investigate whether villages assigned to the treatment group were more likely to go on to win this sanitation prize.

3.3 Results

What did the community sanitation intervention achieve? As outlined in section 3.2.4, we concentrate in this section on Ahmednagar, the only district that, in fact, received the experiment. We document evidence for an effect in four stages. First, the experiment balanced observed baseline properties. Second, the experiment improved sanitation coverage, but not completely. Third, in an independent confirmation of this effect on sanitation, villages assigned to the treatment group were more likely to subsequently win a government prize for being open defecation free. Finally, we show a statistically robust effect on children's height.

3.3.1 Balance of observed baseline properties

Did the random assignment of villages to treatment and control groups achieve balance on observed baseline characteristics? Table 3.2 shows that the data offer no evidence of imbalance, both in the district Ahmednagar where the program was implemented, and in the other two districts where it was not. Across a range of variables, in no case is there a statistically significant difference between the treatment and control groups in variables observed in February 2004, before the program. The greatest t-statistic in absolute value

is 1.52 for having a separate kitchen, in Nanded and Nandurbar districts. Households in the treatment and control groups are similar in the first and second principal components of a vector of assets asked about in the baseline survey. The summary statistics in the table reflect the poverty and poor health in the studied districts. As an illustration of their poverty, we note that only about three-fourths of households owned a clock or watch.

3.3.2 An effect on latrine ownership

Is there evidence in the data, beyond the government's claim to the World Bank researchers, that the program happened at all? As a first piece of evidence, in Ahmednagar, in the mid-line survey shortly after the intervention, respondents were 7.2 percentage points more likely to report a recent visit by a sanitation promoter in the treatment group than in the control group, compared with a 1.5 percentage point difference in Nanded and Nandurbar. This difference is only barely statistically significant, but it is not surprising that villagers would remember this particular government visit with much noise. The exact two-sided p-value differs across household-level regressions with standard errors clustered by village, regressions collapsed to 60 village-mean observations, and in a non-parametric Mann-Whitney-Wilcoxon test on the collapsed means (0.116, 0.099, and 0.066).

More importantly, villages in the treatment group built more latrines. In the final survey round in August of 2005, treatment village household latrine coverage in Ahmednagar had increased by 8.2 percentage points more than for control households,[13] compared with a difference of -0.9 percentage points in Nanded and Nandurbar. The Ahmednagar difference has a two-sided p-value of 0.073 and 0.072 in uncollapsed (clustered) and collapsed regressions, respectively, although curiously the corresponding figure is 0.22 in a collapsed Mann-Whitney-Wilcoxon test.

[13] If each of the 60 villages in Ahmednagar is omitted in turn, this point estimate ranges from 5.5 percentage points to 9.1 percentage points.

This effect of the program on latrine coverage is interesting in two further ways. First is that it is modest: an 8.2 percentage point increase left many people openly defecating. Indeed, we do not know exactly what the change in latrine *use* was. However, it is comparable to the average change in latrine coverage due the the TSC throughout India in the Indian government's inflated official statistics. The fact that the program had an effect on height without resulting in universal latrine coverage suggests one of two possibilities. One is that here it was not the case, as is sometimes claimed in the policy literature, that only eliminating open defecation can have an important effect on health. An effect on health of improvements short of eliminating open defecation could be plausible in these rural, low-population density villages, where fecal pathogens are unlikely to influence the health of children living some distance away. Another possibility is a consequence of the fact that latrine ownership and use often diverge. Just as latrine ownership does not guarantee a complete eschewal of open defecation,[14] nor does individual use require personal ownership of a latrine. Public and school latrines are also part of the program, as is a general, if not universal, increase in use of latrines that had already been built.

A second property of the effect on latrine coverage is that the distributions of village sanitation coverage in the treatment and control groups are different throughout. Figure 3.1 plots the CDFs of village latrine coverage for the treatment and control groups in Ahmednagar in the endline data.[15] Thus, each line reflects 30 data points, each point an average of an indicator of household latrine ownership for a village. The treatment group distribution is always to the right of the control group distribution, and there is clear separation at both the top and the bottom. The figure highlights that only a few treatment group villages achieved more than 50% coverage.

[14] In the baseline data, about a quarter of latrine-owning households still practiced open defecation.

[15] To be clear, it is not the case that subtracting the curves gives a distribution of treatment effects.

3.3.3 An effect on the Clean Village Prize

What evidence do we have that the program was successful from data sources outside of this experiment? The central government of India awards a Clean Village Prize, intended for villages where nobody defecates openly, but instead disposes of feces safely. Although certainly the government prize awarding process is not perfectly accurate, we believe it is at least positively correlated with sanitation.

We merged our data – by hand but blindly to treatment and control status – with central government administrative records of which villages won the prize. The program studied occurred in 2004, our endline data were collected in 2005, and the first village in our sample ever to do so won the prize in 2006. We received data on prize winners in July 2012, indicating which villages in Ahmednagar had ever won the prize by that time.

In the treatment group, 9 of 30 villages have won the Clean Village Prize; in the control group, 3 of 30 villages have won the prize. This 20 percentage point difference is statistically significant with robust regression t, Mann-Whitney, and Fisher exact tests with p-values of 0.054, 0.055, and 0.052, respectively. Because these prizes were awarded several years after our experiment ended, because they involve several investigations by various agents, and because during the time period studied prizes were ultimately approved by the central government in Delhi rather than the state government, we consider it to be very unlikely that the prize outcomes were manipulated to create the appearance of an effect of this experiment. Therefore, we interpret this finding, that treatment group villages were more likely to go on to win the prize than control group villages, as independent confirmation that the experiment happened and caused an improvement in sanitation.

3.3.4 An effect on children's height

Did the program increase children's heights? Table 3.3 presents regression evidence from Ahmednagar that it did. The table reports results from 16 specifications in order to demon-

strate the robustness of the finding. Results are collected into four panels, corresponding with regression equations 1 through 4, respectively:

- **Panel A:** Double difference (Ahmednagar only, treatment × time), midline and endline separated, that is, treatment and control villages were compared only using Ahmednagar data, comparing the differences over time between the two groups,
- **Panel B:** Double difference (Ahmednagar only, treatment × time), midline and endline collapsed into "after,"
- **Panel C:** Triple difference (Nanded & Nandurbar included, treatment × time × Ahmednagar), midline and endline separated, that is, including all villages in the sample (not just in Ahmednagar), with the difference between treatment and control time trends also being compared across districts, and
- **Panel D:** Triple difference (Nanded & Nandurbar included, treatment × time × Ahmednagar), midline and endline collapsed into "after."

Within each panel, four specifications are included:

- **Column 1:** The basic double or triple interaction, and nothing else.
- **Column 2:** To column 1, we add 120 dummies for age in months 1-60, separately for boys and girls. This accounts for the unfolding of stunting over time, for any mean differences between our population and the WHO reference population, and for any differences in age structure across experimental groups. Adding these controls slightly increases the experimental point estimate in two cases and decreases it in two cases, but in no case makes an important difference.
- **Column 3:** To column 2, we add village fixed effects (constant across the three survey rounds). Because the treatment was randomly assigned to villages, we would not expect these to have an effect, and they do not, other than to slightly reduce standard errors.

- **Column 4:** As recommended in the WHO height-for-age reference table guidelines, we omit observations more than 6 standard deviations from the mean. Column 4 replicates column 2 using truncated regression[16] to demonstrate that this truncation has little effect. Section 3.3.4 further documents that mismeasured ages and dispersion in height-for-age z-scores are not responsible for our results.

In all cases an effect of the program is seen, typically in the range of 0.3 to 0.4 height-for-age standard deviations, or about 1.3 centimeters in a four-year-old. McKenzie (2012) recommends longer time series in experimental studies than simple before-and-after. Although we only have two post-intervention survey rounds, it is notably consistent with our interpretation of the results as representing an effect of the program that the point estimate for the endline is greater than the point estimate for the midline in every case, perhaps as the effects of reduced enteric infection have had an opportunity to accumulate. So, in panel A, the effect ranges from 0.236 to 0.278 at midline, and from 0.379 to 0.448 at the end. Without making the distinction of endline to midline – that is, ignoring the length of exposure to the program in panel B – the effect is unsurprisingly right in the middle: 0.324 to 0.357.

Non-parametric statistical significance tests, collapsed to the village level, confirm these findings. It is important to verify collapsed results – where villages are observations – because the treatment was randomly assigned and implemented at the village level. Twice, once for the midline and once for the endline, we create a dataset of 60 observations: for each village we compute first the mean height-for-age z score in each round, and then the change since the first round. We perform a Mann-Whitney-Wilcoxon rank sum test. The null hypothesis that the distributions from which the changes in mean height were drawn were the same in treatment and control villages is almost rejected in the baseline-to-midline case with a two-sided p-value of 0.103 and is rejected in the baseline-to-endline case with

[16] A'Hearn (2002) recommends this procedure for studying truncated height samples. Strong normality assumptions are required, although this might not be such a misleading assumption for normalized height.

a p-value of 0.065. Repeating this procedure a third time with the midline and endline collapsed into a single "after" period produces a p-value of 0.048.

A further alternative specification is to omit any use of z-scores by using height in centimeters as the dependent variable, in logs to account for different effect sizes at different ages. The effect of the program in the endline period is to increase height by 1.8 percent ($t = 2.20$) in the double difference[17] (comparable in functional form to column 2 of panel A) and by 2.7 percent ($t = 2.45$) in the triple difference (comparable to column 2 of panel C).

A final test responds directly to Deaton's (2012) concern that the overall result could be driven by one village with a large potential treatment effect or other special properties. We replicate the estimation of the "after" treatment effect in Ahmednagar 60 times, omitting each village in turn. The point estimate ranges from a minimum of 0.28 to a maximum of 0.37 and the t-statistic ranges from 1.94 to 2.66, with a mean of 2.20. Thus our result does not merely reflect any one outlier village.

Improvement in height, but not to the WHO standard

How large is the estimated effect on children's height? One way to understand the effect is to compare it with chapter two's estimates of the effect of the government's Total Sanitation Campaign throughout India. Averaging over incomplete and heterogeneous implementation, that chapter reported that, on average, the program increased height-for-age z-scores by about 0.2 standard deviations. Our experimental estimates are about 1.5 to 2 times as large, consistent with the fact that they are derived from a focused, relatively high-quality experimental implementation.

Another way to understand the effect is to compare it with the gap between the average Indian child and the WHO reference population mean. On average, Indian children older

[17] As one step further, if age in months is restricted to under 18 (to only include children exposed or not exposed to the program) this estimate is a nearly identical 1.9 percent, but with a sample about one-third the size, the two-sided p-value rises to 0.17 ($t = 1.38$).

than 24 months are about 2 standard deviations below the WHO reference mean, and the children in our study are even shorter. Figure 3.2 plots the average endline heights at each age in the treatment and control group in Ahmednagar (as kernel-weighted local polynomial regressions), alongside the mean height of the WHO reference population.[18] The waviness in the graph is due to age heaping of children at round ages. The figure shows that treatment group children are taller than control group children, although not by nearly enough to reach the WHO reference mean.

This graph suggests a further, non-parametric statistical significance test of the main effect of the program on height in Ahmednagar. We collapse the data into 120 observations: a mean for each age-in-months for the treatment and control groups. A matched pairs sign test rejects that the median of the differences is zero with a p-value 0.078. If, instead, 240 observations are created, allowing separate means for boys and girls, the p-value is 0.039. Because this test compares children within age-in-months by sex categories, it also is unaffected by any concern that the WHO reference population may not be appropriate, due to, say, age or gender bias or any international genetic differences (Panagariya, 2012). That is, the mean deviation of height from the reference population will be the same for treatment and control groups within each age-by-sex category, so we are simply comparing the difference in heights.

Negative externalities: Effects in households without latrines

Existing evidence such as the interaction of open defecation with population density in its effect on children's health in chapters one and two suggests negative externalities, effects of one household's open defecation on another's children. But perhaps these prior studies cannot definitively rule out that the effect is purely due to a household's use of its own latrine, or that associations between child height and community sanitation averages reflect

[18]Note that although this resembles a growth curve, it is from a *synthetic* cohort – that is to say, a cross-section – and does not plot the longitudinal growth of any child.

omitted variables. Our experiment randomized the intervention at the community level (Miguel and Kremer, 2004). Therefore, an effect on the heights of children whose households did not use latrines, even at the endline after the program, would be evidence of spillovers of sanitation onto other local households.

Indeed, even after the program most children lived in households without latrines. Restricting the sample to this subset[19] (74.6 percent of the Ahmednagar sample) and estimating the simple difference-in-differences in panel B of table 3 finds that the program caused even children in this group to be 0.42 standard deviations taller (standard error = 0.19, $n = 2,562$). Although this point estimate appears slightly larger than the estimate for the full sample, in fact it is not statistically detectably different. When the full sample is used with a fully-interacted triple difference, the effect of the program on households with a latrine at endline is no different than the effect on households without a latrine at endline: the estimate of the triple difference (treatment $\times$ after $\times$ latrine at endline) is 0.001 with a standard error of 0.20 and a t-statistic of 0.01.

Therefore, this community-level experiment offers evidence of spillover effects of open defecation. The program made children taller even in households who did not themselves own latrines, and there is no evidence that the effect of the program differed across children in households which did or did not.

Differences throughout the height distribution

Were the final differences between the treatment and control groups concentrated on taller or shorter children? Implementing randomization only ensures a consistent estimate of the average treatment effect, not of the full distribution of treatment effects, but recognizing this, it still could be informative to compare the height distributions in the treatment and control groups.

[19]The subset is children who live in households who did not have a latrine at endline; this excludes children who live in households who did not have a latrine at baseline but who acquired one by endline.

Panel A of figure 3.3 plots height CDFs in the treatment and control groups in Ahmednagar in the baseline data, from before the program. The lines are very close to each other, as we would expect, with the slight separation at the bottom suggesting that the shorter children in the control group were not as short as the shorter children in the treatment group before the program.

Panel B presents the same CDFs from the endline data, after the program. Almost throughout the range, the treatment group distribution has moved to the right of the control group distribution (especially recalling their baseline separation among shorter children). If so, this is consistent with open defecation being a public bad with consequences for many people and with widespread stunting and enteropathy in the population, such that even relatively tall children have room to grow relative to their genetic potentials. A Kolmogorov-Smirnov test for equality of distribution rejects that the treatment and control distributions of height are the same after the program ($p = 0.03$) but does not using the data from before ($p = 0.23$).

Finally, figure 3.4 plots quantile regression estimates of the difference-in-differences "effect" of the program at deciles 0.1 through 0.9. Thus, these are quantile regression equivalents of equation 3.2, corresponding with panel B of table 3. As the CDFs suggested, the quantile coefficients are similar to the OLS estimate of the program effect throughout the height distribution. There is a slight trend of greater quantile coefficients among shorter children, and indeed a linear regression of the nine coefficient estimates on the quantiles 0.1 through 0.9 finds a negative slope (p-value $= 0.065$). Therefore, if anything there is a slightly greater quantile difference-in-differences towards the bottom of the height distribution.

Mismeasured ages and truncation are not important here

Height-for-age z-scores, especially for young children, require accurately measured age in months. Mismeasuring ages (as well as mismeasuring heights) will add noise to the z-scores.

Thus, in our data, the standard deviation of scaled height-for-age is 2.1, more than twice as much as would be expected from a standardized normal distribution.

One consequence of this noise is to reduce power by increasing standard errors. Another consequence is to require a truncated data set, by following the WHO recommendation of omitting heights more than 6 standard deviations from the mean – which, in our case, means using only z-scores from -8 to 4.

Could this truncation be responsible for our results? Would other endpoints produce different answers? In order to answer this question, table 3.4 reports results from 49 alternative combinations of endpoints, in 0.5 standard deviation increments of a lower bound of -9 to -6 and an upper bound of 3 to 6. In particular, the table replicates column 2 of panel B, using the double-difference with a collapsed "after" so that there is only one treatment effect estimate to report for different combinations.

Changing the cut-points has little effect on the result. All coefficient estimates are positive, and 94 percent (all but the extreme lower-right corner) are between 0.2 and 0.4 height-for-age standard deviations, comparable to the range in table 3.3. The bottom panel reports corresponding t-statistics; most exceed 2, especially near the WHO-recommended cutpoints that we use. Unsurprisingly, estimates become less precise as the bounds are widened and noisier observations are included. The mean across all combinations of cut-points is an effect of 0.3 standard deviations and a t-statistic of 1.96.

An alternative approach, introduced in section 3.3.4, is to use log of height as the dependent variable, omitting z-scores altogether. If this is done with the widest set of cutpoints used in the table, -9 to 6, we find that the program increased height in the endline by 3.3 percent with a t-statistic of 2.42.

3.4 Policy implications

This section considers two ways to draw and interpret policy conclusions from this experiment. The first considers the average effect, or lack of an effect, of the original government official's plan to conduct an experiment in three districts. The second explores heterogeneity in the effect of the program across geographic places.

3.4.1 Researcher-implementer partnerships and the "effect" of an official's decision

We have been reporting the effect of the program *in Ahmednagar*, the only district studied that actually received an experiment. However, a high-ranking government official, in approving this project, had intended for all three districts to experience the experimental treatment. Looking at the effects of initial random assignment of villages to treatment and control groups, we can estimate a confidence interval for the effect of this decision. This is equivalent to replicating the regressions in panel A of table 3.3 using all three districts together, rather than only Ahmednagar.

Unsurprisingly, there is no evidence of an effect. For both period 2 and period 3, the wide 95% confidence intervals include zero: -0.21 to 0.35 and -0.30 to 0.25 respectively. Also unsurprisingly, there is no evidence of an effect in either Nanded or Nandurbar, studied separately. The greatest t statistic in absolute value over the two periods and two districts is 1.08. So, the "intent to treat" – which ordinarily is the only effect of a policy that is under the control of the deciding policy-maker – is not significantly different from zero in the pooled sample of all three districts.[20] If we did not know that the program simply was

[20]This is often contrasted with a "treatment on the treated" estimate resulting from instrumenting for actual implementation using originally intended implementation. If we compute this two-stage least squares estimate with the pooled data from all three districts, we find an "after" program effect of 0.253 height-for-age standard deviations ($p = 0.064$), or 0.272 ($p = 0.051$) with the age controls, and a first stage t-statistic of 36 on random assignment to the treatment predicting the treatment, addressing weak instrument concerns. However, these estimates are subject to all of the ordinary limits of instrumental variables estimates, including the opacity of local average treatment effects.

not implemented by two district governments, this would be the end of our analysis. Many experiments avoid this risk by having the intervention done such that it is fully implemented by and under the control of the researchers, at a difficult to quantify cost to external validity (Ravallion, 2012). Arguably, however, the very risk of non-implementation that deters would-be researchers is of first-order importance to policy-makers. Keeping better records with better monitoring of how and why programs are ultimately implemented may reveal important, systematic, or predictable problems of implementation (Pritchett et al., 2012).

3.4.2 Effect heterogeneity and experiment generalizability

Although rarely reported in the literature, it seems probable that many experiments are planned, perhaps approved by high-ranking officials, and never implemented, like ours in Nanded and Nandurbar. If so, the only thing unusual about our data would be that we nevertheless surveyed households in those districts, that is, we set up and carried out an evaluation despite not doing the project. Of course, we will never know what the effect of the intervention would have been in these districts if it had occurred. Plausibly the effect would have been smaller. In general, as others have noted before, potential effect size may or may not be correlated with the political, bureaucratic, or financial factors that determine whether an experiment is possible or likely in a particular place.

Two further procedures can help us asses the generalizability of our results (but only to a very limited degree). One approach, recommended by Allcott and Mullainathan (2012), is to interact the treatment effect estimate with indicators for geographic regions of intermediate size (smaller than the total experiment area, but larger than the unit of observation) and to report an F-test of the null hypothesis that the experimental effect does not vary across these sub-regions. In our case, this would be blocks, which divide districts and contain villages. In Ahmednagar district, there are 11 blocks that contain at least one treatment and control village, allowing 10 degrees of freedom. Three blocks with only

treatment or control villages are therefore dropped, resulting in a 10 percent reduction in the sample; nevertheless the "after" effect estimate is a similar 0.35 ($t = 2.13$) in this subsample.[21] The F-statistic of 1.76 has a p value of 0.091, weak evidence of heterogeneous treatment effects across blocks. As Allcott and Mullainathan suggest, if the treatment effect is detectably heterogeneous across blocks, than perhaps it is also plausible that it would be heterogeneous across districts.

A second approach is to explicitly model heterogeneity in the treatment effect. We matched the data on the 60 villages in Ahmednagar to information about those villages from the 2001 Indian census. This allowed us to check for interactions (strictly speaking, triple differences, because the main specification is a difference-in-differences): do any village level census variables predict larger or smaller treatment effects? We found no interactions with population density (triple difference $t = -0.54$), fraction of the population that belongs to a Scheduled Tribe ($t = -0.54$), or fraction of the population that belongs to a Scheduled Caste ($t = 0.62$). However, with a relatively small number of villages, this study probably did not have much power to detect such an interaction, if any exists;[22] nor was such an interaction built into the study design by *a priori* blocking. Thus, this approach offers little evidence about treatment effect heterogeneity.

3.5 Conclusion

We have analyzed a randomized controlled trial of a community sanitation program in Ahmednagar district of Maharashtra, India. The program was associated with a 0.3 to 0.4 standard deviation increase in children's height-for-age z-scores, or approximately 1.3

[21] This suggests another statistical significance test for the chapter's main result of an effect on height. In the spirit of Ibragimov and Müller (2010), estimating 11 treatment effects, one for each block, produces 11 coefficient estimates. Of these 11, 9 are positive. The probability of at least 9 positive numbers in 11 independent draws of numbers equally likely to be negative or positive can be computed exactly from a binomial distribution; it is 0.033, an alternative p-value for a median positive effect on height.

[22] Chapter two finds an interaction between population density and the effect of a sanitation program using variation at the district level across 500 districts.

centimeters in a four-year-old. Note that this was achieved without, in general, eliminating open defecation throughout villages, although in these low population density villages, it may have been eliminated from many children's immediate environments.

This important effect is about 1.5 to 2 times the size of the effect found by chapter two in a retrospective study of the average effect of the TSC throughout rural India. It is unsurprising that we find a larger effect. First, chapter two's result averages over heterogeneous and incomplete implementation, while we study a special experimental effort representing a collaboration between the state government and the World Bank. Second, our analysis only includes data from Ahmednagar; whatever process "selected" this district for the experiment may well be correlated with a large potential treatment effect.

Deaton (2012) encourages consumers of experimental evaluations to note the "experimental panel": the set of observations eligible to be randomized into the treatment or control group. Often, this set is not chosen according to a representative sampling strategy. Unusually, we have information about our experimental panel and about a larger set of villages that was originally planned to be an experimental panel. Although we cannot know what the experiment would have found in Nanded and Nandurbar, knowing this history and that these two districts have worse sanitation coverage than Ahmednagar and, in one case, a larger Scheduled Tribe population underscores that our experimental result is not independent of a context in place and time.

The reduction in the experimental panel highlights an important fact which no randomization is required to prove: the capacity of the Indian state is limited, and there remain many places which have not yet been reached by even the most basic sanitation coverage. Despite the effects of the Total Sanitation Campaign where it has happened, half of the Indian population defecates unsafely.

How well can the our findings in Ahmednagar be generalized? If reducing open defecation indeed can improve the height and health of children throughout India even approxi-

mately as well as it could in the villages experimentally studied, then this remaining half could represent an important opportunity to invest in India's human capital.

For policy, open defecation causes negative externalities, establishing a firm theoretical claim to public resources which is complemented by the empirical findings of this paper chapter. Economic theory may advise a resource-constrained government to address problems reflecting large market failures – such as externalities – before providing excludable and rival private goods. Because most curative care of non-communicable disease may not have such externalities,[23] this argument implies that sanitation should be a health policy priority.

[23]Of course, the market for curative health care involves other market failures, including asymmetric information (Arrow, 1963), although there is also evidence that many primary health care providers in India are poorly informed.

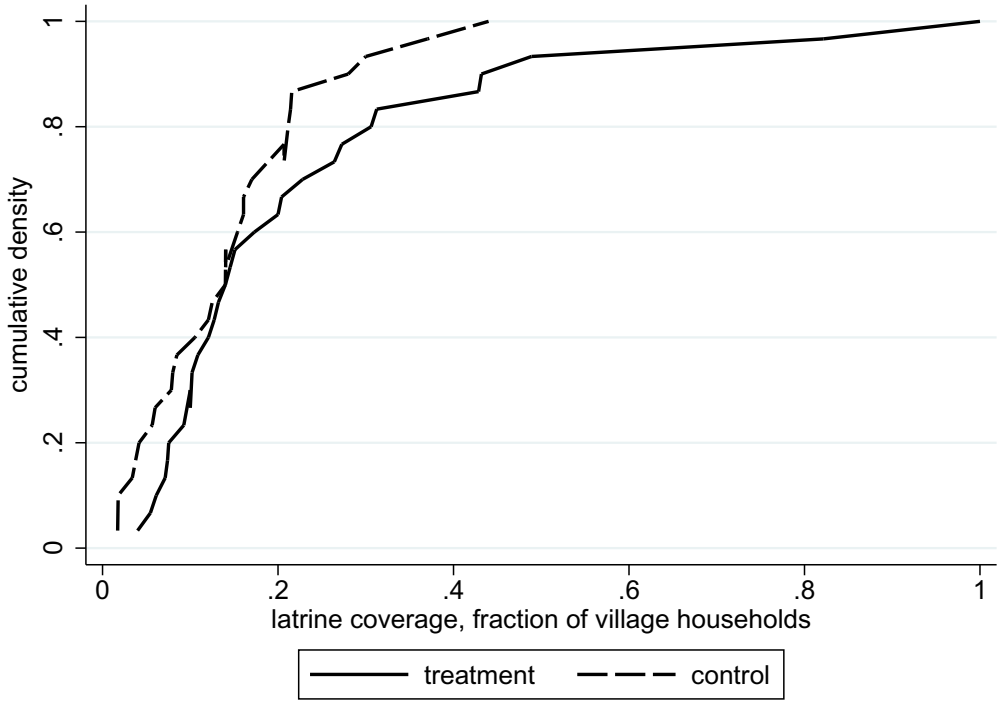


Figure 3.1: Distribution of village-level sanitation in Ahmednagar district, endline

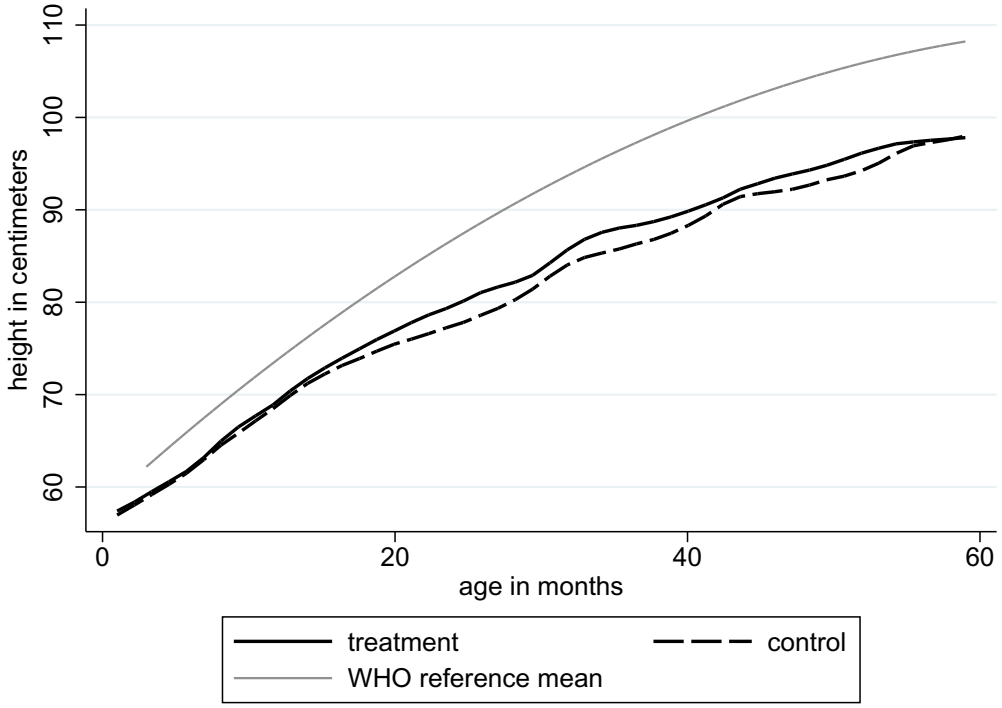
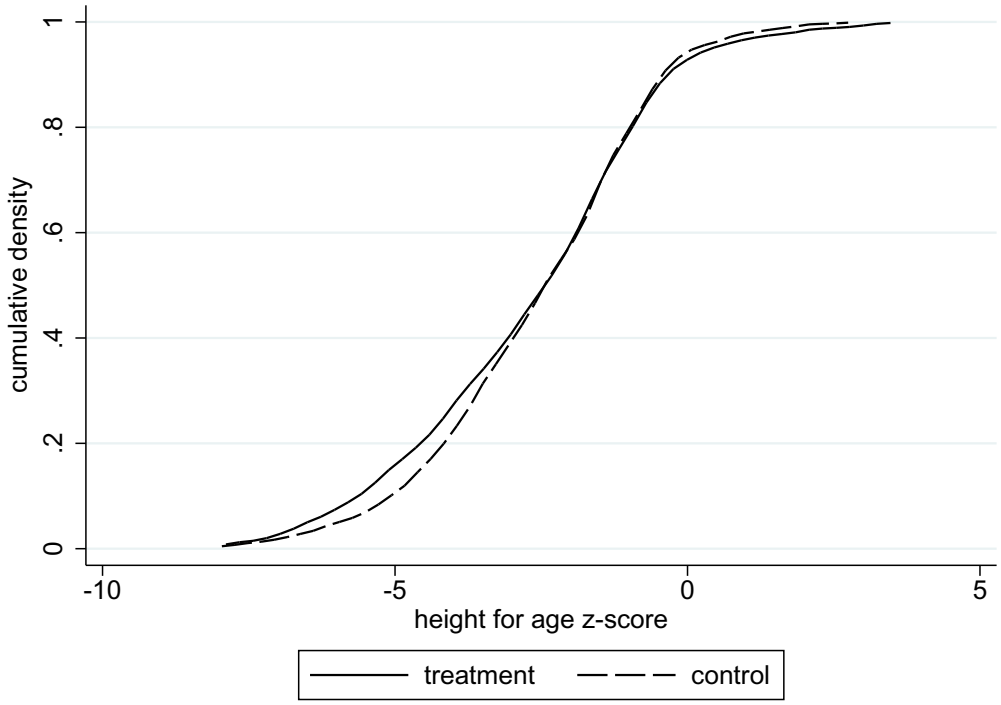


Figure 3.2: Height of children in Ahmednagar district by age, endline survey

(a) baseline survey



(b) endline survey

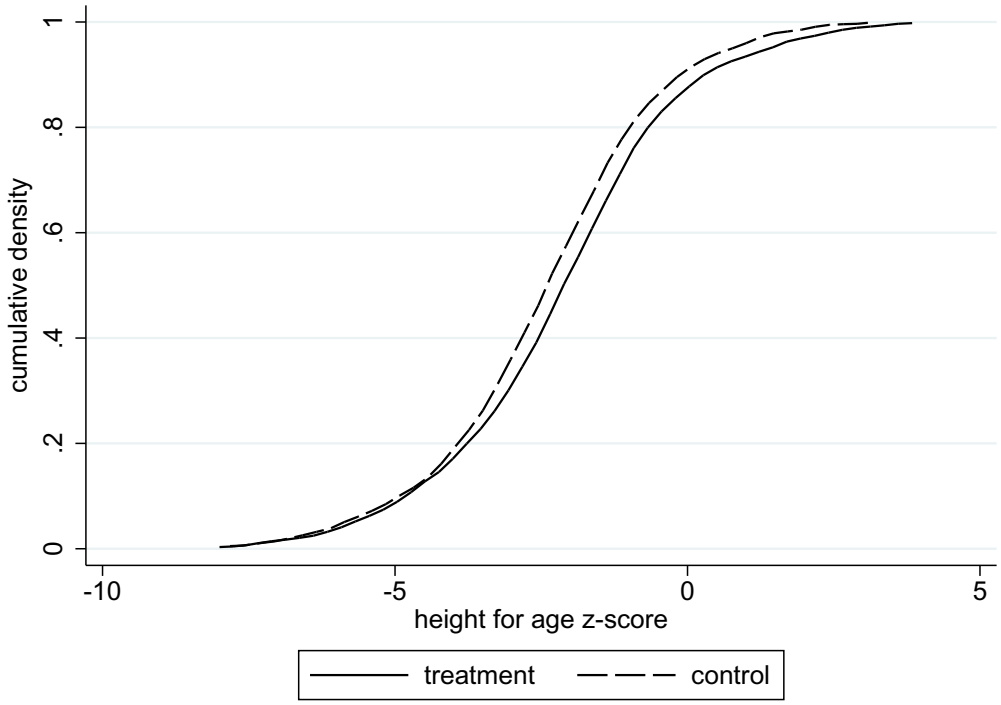


Figure 3.3: Distribution of height of children in Ahmednagar district

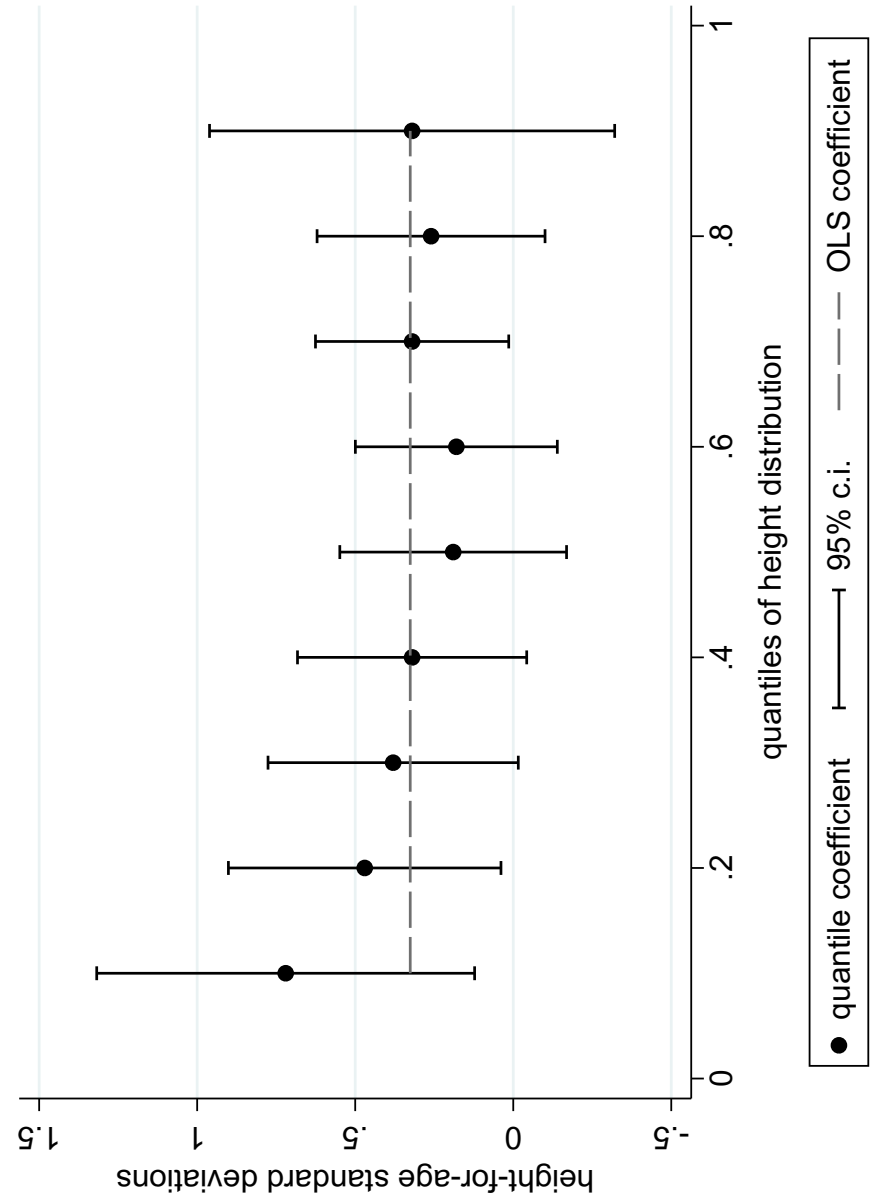


Figure 3.4: Quantile regression estimates of difference-in-differences, Ahmednagar district

Table 3.1: Comparison among studied districts

	source	Ahmednagar	Nanded	Nandurbar	rural Maharashtra	rural India
population, millions	2001 census	4.1	2.7	1.3	41.1	742.5
urban population %	2001 census	19.9	24.0	15.4	42.4*	27.8*
population density (per km^2)	2001 census	240	260	220	181-314	230-312
Scheduled Tribe %	2008 DLHS	12.7	16.9	71.4	23.6	23.1
Scheduled Tribe %	2001 census	7.5	8.8	65.3		
Scheduled Caste %	2001 census	12.0	17.3	3.2		
infant mortality rate	2001 census	44	61	64	53	73
open defecation %	2011 census	48.7	65.6	65.4	62.0	69.3
with toilet facility %	2008 DLHS	52.3	31.1	19.6	32.5	34.2
open defecation %	2001 census				81.8	78.1
electricity %	2011 census	75.1	74.5	58.3	73.8	55.3
modern housing materials %	2008 DLHS	39.3	50.4	7.3	16.8	19.6

* Fraction of all of population of Maharashtra and India that live in urban areas.

Table 3.2: Balance of baseline measurement sample means

	Ahmednagar district			Nanded and Nandurbar		
	control	treatment	t	control	treatment	t
height for age	-2.58	-2.68	-0.82	-3.70	-3.66	0.24
has vaccine card	0.95	0.94	-0.46	0.86	0.81	-1.46
fed breastmilk at birth	0.98	0.99	0.74	0.97	0.97	-0.13
months exclusively breastfed	4.80	5.21	1.09	5.75	5.95	1.10
total months breastfed	7.57	8.03	0.59	9.99	10.67	1.22
female	0.46	0.51	1.38	0.52	0.50	-0.95
age in months	37.76	37.37	-0.37	38.84	39.26	0.61
asset index 1 (first component)	-0.72	-1.03	-1.30	0.41	0.47	0.38
asset index 2 (second component)	0.06	0.06	0.01	-0.03	-0.03	-0.06
has separate kitchen	0.62	0.65	1.03	0.48	0.44	-1.52
has clock or watch	0.73	0.74	0.39	0.51	0.51	-0.09
n (children under 5)	1,686	1,754		3,967	3,953	
villages	30	30		60	60	

Table 3.3: Effects of the experimental program on height-for-age in Ahmednagar

	(1) OLS	(2) OLS	(3) vil. FE	(4) trunc.
round × dist FEs	✓	✓	✓	✓
age × sex		✓	✓	✓
village FEs			✓	
Panel A: Double difference, midline and endline				
treatment	-0.105	-0.0988		-0.101
	(0.129)	(0.129)		(0.130)
treatment	0.278†	0.236†	0.274*	0.241†
× midline	(0.154)	(0.140)	(0.136)	(0.141)
treatment	0.379†	0.418*	0.448*	0.427*
× endline	(0.211)	(0.195)	(0.190)	(0.195)
n (children)	3,432	3,432	3,432	3,432

	(1) OLS	(2) OLS	(3) vil. FE	(4) trunc.
round × dist FEs	✓	✓	✓	✓
age × sex		✓	✓	✓
village FEs			✓	
Panel B: Double difference, before and after				
treatment	-0.105	-0.0992		-0.102
	(0.129)	(0.129)		(0.130)
treatment	0.326*	0.324*	0.357*	0.330*
× mid. or end.	(0.160)	(0.146)	(0.141)	(0.147)
n (children)	3,432	3,432	3,432	3,432

Table 3.3, continued

	(1) OLS	(2) OLS	(3) vil. FE	(4) trunc.
round × dist FEs	✓	✓	✓	✓
age × sex		✓	✓	✓
village FEs			✓	
Panel C: Triple difference, midline and endline				
treatment	0.0412	0.0501		0.0559
	(0.172)	(0.172)		(0.192)
treatment × Ahm.	0.298	0.224	0.250	0.235
× midline	(0.237)	(0.232)	(0.227)	(0.250)
treatment × Ahm.	0.572*	0.609*	0.646*	0.640*
× endline	(0.264)	(0.256)	(0.249)	(0.273)
treatment	-0.147	-0.114		-0.122
× Ahmednagar	(0.214)	(0.212)		(0.232)
treatment	-0.0200	-0.000374	0.0154	-0.00300
× midline	(0.181)	(0.181)	(0.178)	(0.200)
treatment	-0.192	-0.211	-0.220	-0.226
× endline	(0.160)	(0.160)	(0.158)	(0.178)
n (children)	11,337	11,337	11,337	11,337

	(1) OLS	(2) OLS	(3) vil. FE	(4) trunc.
round × dist FEs	✓	✓	✓	✓
age × sex		✓	✓	✓
village FEs			✓	
Panel D: Triple difference, before and after				
treatment	0.0412	0.0500		0.0557
	(0.172)	(0.172)		(0.192)
treat. × Ahm.	0.431†	0.411†	0.443*	0.432†
× mid. or end.	(0.226)	(0.220)	(0.213)	(0.237)
treatment	-0.147	-0.114		-0.122
× Ahmednagar	(0.214)	(0.212)		(0.232)
treatment	-0.105	-0.104	-0.101	-0.113
× mid. or end.	(0.160)	(0.161)	(0.158)	(0.179)
n (children)	11,337	11,337	11,337	11,337

Table 3.4: Robustness of estimated effect on height to alternative extreme-value bounds

Panel A: Coefficients on treatment × after

	upper limit on height for age z-scores						
lower limit	3.0	4.0	4.5	5.0	5.5	6.0	mean
-9.0	0.336	0.328	0.333	0.301	0.296	0.252	0.309
-8.5	0.329	0.320	0.325	0.292	0.287	0.242	0.301
-8.0	0.365	0.356	0.361	0.331	0.326	0.281	0.338
-7.5	0.345	0.336	0.341	0.311	0.306	0.261	0.318
-7.0	0.327	0.318	0.324	0.294	0.290	0.244	0.301
-6.5	0.331	0.321	0.327	0.296	0.291	0.245	0.303
-6.0	0.229	0.218	0.226	0.193	0.188	0.142	0.201
mean	0.323	0.314	0.319	0.288	0.283	0.238	0.296

Panel B: t-statistics on treatment × after

	upper limit on height for age z-scores						
lower limit	3.0	4.0	4.5	5.0	5.5	6.0	mean
-9.0	2.05	1.99	2.04	1.70	1.63	1.36	1.82
-8.5	2.17	2.10	2.14	1.77	1.69	1.39	1.91
-8.0	2.62	2.52	2.55	2.15	2.06	1.73	2.31
-7.5	2.44	2.36	2.38	2.01	1.91	1.58	2.15
-7.0	2.36	2.25	2.29	1.89	1.80	1.48	2.04
-6.5	2.41	2.26	2.29	1.88	1.79	1.47	2.05
-6.0	1.73	1.59	1.64	1.28	1.21	0.90	1.42
mean	2.25	2.15	2.19	1.81	1.73	1.42	1.96

This specification corresponds with column 3 of panel B of table 3.

Bibliography

A'Hearn, Brian. 2002. "A restricted maximum likelihood estimator for truncated height samples." *Economics and Human Biology*, 2(1): 5–19.

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.

Allcott, Hunt and Sendhil Mullainathan. 2012. "External Validity and Partner Selection Bias." working paper 18373, NBER.

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.

Arrow, Kenneth J.. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review*, 53(5): 941–973.

Banerjee, Abhijit V.. 1997. "A Theory of Misgovernance." *Quarterly Journal of Economics*, 112(4): 1289–1332.

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*.

Bateman, O. Massee and Shelley Smith. 1991. "A comparison of the health effects of water supply and sanitaiton in urban and rural Guatemala." WASH Field Report 352, USAID.

Bateman, O. Massee, Shelley Smith, and Philip Roark. 1993. "A comparison of the health effects of water supply and sanitaiton in urban and rural areas of five African countries." WASH Field Report 398, USAID.

Bennett, Daniel. 2012. "Does Clean Water Make You Dirty? Water Supply and Sanitation in the Philippines." *Journal of Human Resources*, 47(1): 146–173.

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.

Blinder, Alan S. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources*, 8: 436455.

Bongartz, Petra and Robert Chambers. 2009. "Beyond Subsidies Triggering a Revolution in Rural Sanitation." *IDS In Focus*(10).

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): 499–532.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review*, 92(5): 1308–1334.

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.

Coffey, Diane, Reetika Khera, and Dean Spears. 2013. "Women's Status and Children's Height in India: Evidence from Rural Joint Households." working paper, Princeton University.

Currie, Janet. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature*, 47(1): 87–122.

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. 1997. *Analysis of Household Surveys*: World Bank.

——— 2007. "Height, health and development." *Proceedings of the National Academy of the Sciences*, 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.

——— 2012. "Searching for Answers with Randomized Experiments." presentation, NYU Development Research Institute.

Deaton, Angus and Jean Drèze. 2009. "Food and Nutrition in India: Facts and Interpretations." *Economic and Political Weekly*, 44(7): 42–65.

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.

Doblhammer, Gabriele and James W. Vaupel. 2001. "Lifespan depends on month of birth." *PNAS*, 98(5): 2934–2939.

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.

Fink, Günther, Isabel Günther, and Kenneth Hill. 2011. "The effect of water and sanitation on child health: evidence from the demographic and health surveys 19862007." *International Journal of Epidemiology*, 40(5): 1196–1204.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. "Decomposition Methods in Economics." *Handbook of Labor Economics*, 4a: 1–102.

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.

Geruso, Michael. 2012. "Black-White Disparities in Life Expectancy: How Much Can the Standard SES Variables Explain?" *Demography*, 49(2): 553–574.

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.

Günther, Isabel and Günther Fink. 2010. "Water, Sanitation and Childrens Health: Evidence from 172 DHS Surveys." Policy Research Working Paper 5275, World Bank.

Headey, Derek D.. 2013. "Developmental Drivers of Nutritional Change: A Cross-Country Analysis." *World Development*, 42: 7688.

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.

Ibragimov, Rustam and Ulrich K. Müller. 2010. "t-statistic Based Correlation and Heterogeneity Robust Inference." *Journal of Business and Economic Statistics*, 28: 453–468.

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

Jann, Ben. 2008. "A Stata implementation of the Blinder-Oaxaca decomposition." *Stata Journal*, 8(4): 453–479.

Jayachandran, Seema and Rohini Pande. 2012. "The Puzzle of High Child Malnutrition in South Asia." presentation slides, International Growth Centre.

Jeffery, Patricia, Roger Jeffery, and Andrew Lyon. 1989. *Labour pains and labour power: Women and childbearing in India*: Zed.

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): F52–F84.

Korpe, Poonum S. and William A Petri, Jr.. 2012. "Environmental enteropathy: critical implications of a poorly understood condition." *Trends in Molecular Medicine*, 18(6): 328–336.

Kov, Phyrum, Susanna Smets, Dean Spears, and Sangita Vyas. 2013. "Open defecation and child height: Evidence from changes in Cambodian provinces." working paper, World Bank WSP and rice.

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.

Liu, Hong, Hai Fang, and Zhong Zhao. 2012. "Urban-rural disparities of child health and nutritional status in China from 1989 to 2006." *Economics and Human Biology*.

Martorell, Reynaldo, Charles Yarbrough, Aaron Lechtig, Hernan Delgado, and Robert E. Klein. 1977. "Genetic-environmental interactions in physical growth." *Acta Pædiatr Scand*, 66: 579–584.

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.

McKenzie, David. 2012. "Beyond baseline and follow-up: The case for more T in experiments." *Journal of Development Economics*, 99: 210221.

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*.

Naandi Foundation. 2011. *HUNGaMA: Fighting hunger & malnutrition*: The HUNGaMA Survey Report.

Niehaus, Paul, Antonia Atanassova, Marianne Bertrand, and Sendhil Mullainathan. 2013. "Targeting with Agents." *American Economic Journal: Economic Policy*, 5(1): 206–38.

Oaxaca, Ronald. 1973. "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, 14: 693709.

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.

Panagariya, Arvind. 2012. "The Myth of Child Malnutrition in India." conference paper 8, Columbia University Program on Indian Economic Policies.

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.

Pritchett, Lant. 2009. "Is India a Flailing State?: Detours on the Four Lane Highway to Modernization." working paper, Harvard.

Pritchett, Lant, Salimah Samji, and Jeffrey Hammer. 2012. "It's All About MeE: Using Structured Experiental Learning ('e') to Crawl the Design Space." working paper, Princeton University Research Program in Development Studies.

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

Ramalingaswami, Vulimiri, Urban Jonsson, and Jon Rohde. 1996. "Commentary: The Asian Enigma." The Progress of Nations.

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.

Reimers, Cordelia W.. 1983. "Labor Market Discrimination Against Hispanic and Black Men." *Review of Economics and Statistics*, 65(4): 570–579.

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

Royston, Patrick, Douglas G. Altman, and Willi Sauerbrei. 2006. "Dichotomizing continuous predictors in multiple regression: a bad idea." *Statistics in Medicine*, 25: 127–141.

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.

Schmidt, I.M., M.H. Jorgensen, and K.F. Michaelsen. 1995. "Height of Conscripts in Europe: Is Postneonatal Mortality a Predictor?" *Annals of Human Biology*, 22(1): 57–67.

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

Spears, Dean and Sneha Lamba. 2012. "Effects of Early-Life Exposure to Rural Sanitation on Childhood Cognitive Skills: Evidence from Indias Total Sanitation Campaign." working paper, Princeton.

Steckel, Richard. 2009. "Heights and human welfare: Recent developments and new directions." *Explorations in Economic History*, 46: 1–23.

Tarozzi, Alessandro. 2008. "Growth reference charts and the nutritional status of Indian children." *Economics & Human Biology*, 6(3): 455468.

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.

Vogl, Tom. 2011. "Height, Skills, and Labor Market Outcomes in Mexico." Working Paper.

Waterlow, J.C.. 2011. "Reflections on Stunting." in Charles Pasternak ed. *Access Not Excess*: Smith-Gordon, Chap. 1: 1–8.

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.