

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

Jeffrey Hammer and Dean Spears*

February 6, 2013

Abstract

Open defecation is exceptionally widespread in India, a country with puzzlingly high rates of child stunting. 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, an important marker of human capital. We further document sanitation externalities: an effect even on children in households that did not adopt latrines. Unusually, we also collected data in districts where the government planned but ultimately did not conduct an experiment, permitting analysis of the importance of the set eligible for randomization.

*Princeton University. Wallace Hall. Princeton, NJ 08540. 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, Diane Coffey, Juan Costain, Oliver Cumming, Angus Deaton, Paul Gertler, Avinash Kishore, 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 here are our own.

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, our paper 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.

Our paper 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 analysis of clustered randomized trials can respond to Deaton’s (2012) recent concern that outliers may inappropriately determine experimental conclusions.

This paper 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 may not. Finally, this paper studies the height of children, which is widely agreed to measure the long-term net nutritional status of children. Indian children are exceptionally short internationally, 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 food as a 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 (Spears, 2012c).

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.

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-level² 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 link-

¹Spears (2012c) notes that population density is also much greater in India than in Africa, providing more opportunities for contact with other people’s fecal pathogens.

²Some 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.

³Lisa 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.

ing open defecation to poor health and early life human capital accumulation. Checkley 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, recent well-identified retrospective econometric studies find an effect of a government sanitation program in rural India.⁴ From 2001 until its replacement with a new program in 2012, the Indian central government has operated a “flagship” rural sanitation program called the Total Sanitation Campaign (TSC). This program offered a partial construction subsidy for building household latrines, and most importantly provided for village-level community sanitation mobilization. This was especially encouraged by the Clean Village Prize, a cash incentive to the leaders of villages that eliminate open defecation. By matching survey and census data on health outcomes to administrative records and program rules, and by exploiting exogenous variation in the timing of program implementation and in a discontinuity in the function mapping village population to prize amount, Spears (2012a) estimates an average effect of the TSC on rural Indian children’s health. Averaging over implementation heterogeneity throughout rural India, Spears finds that the TSC reduced infant mortality and increased children’s height.⁵ 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.

⁴Additionally, 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).

⁵Similarly, in a country-year level study where 140 collapsed Demographic and Health Surveys form the set of observations, Spears (2012c) shows that international variation in open defecation explains much of the international variation in child stunting. In particular, the puzzle of profound Indian stunting (Ramalingaswami et al., 1996) could be importantly partially explained by India’s particularly widespread open defecation.

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, Maharashtra 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.

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 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.⁶ Section 3 presents and analyzes the results of the experiment. Section 4 discusses policy implications, considering treatment heterogeneity and the consequences of the government’s original decision to experiment. Section 5 concludes.

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

⁶This 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.

- August 2005: endline survey data collection

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

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 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 paper to evaluate (Spears, 2012a). Instead, this paper evaluates a modest experimental addition to the TSC in one district. Thus, whenever this paper refers to “the program” studied, we mean only this special, randomized sanitation promotion intervention. 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.⁷

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

⁷Presumably, 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.

(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.

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 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

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 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.⁸

2.4 Empirical strategy

The empirical strategy of this paper is built upon the random assignment of villages to treatment or control groups. Randomization produces unbiased estimates of average effects.

However, a randomized, controlled experimental intervention only occurred in Ahmednagar district. It would be difficult to know how often planned field experiments are canceled,⁹

⁸For this reason, Pritchett (2009) has described India as a “failing 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.

⁹Soon 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

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.

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.¹⁰

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

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).

¹⁰It 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.

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.

2.4.1 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 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, 2012b). Medical and epidemiological evidence, as well as econometric decomposition, suggests that widespread open defecation could be an important part of the explanation for Indian stunting (Humphrey, 2009; Spears, 2012c), and well-identified non-experimental evidence suggests that community sanitation can improve Indian children’s heights (Spears, 2012a).

Height of children under 5 is, therefore, the central dependent variable in our analysis.¹¹ 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

¹¹Although 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.

transform height into z -scores using the 2006 WHO reference population,¹² 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:

$$z_{ivt} = \beta_1 treatment_v + \beta_2 treatment_v \times midline_t + \beta_3 treatment_v \times endline_t + A_{ivt}\Gamma + \alpha_v + \gamma_t + \varepsilon_{ivt}, \quad (1)$$

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-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

$$z_{ivt} = \beta_1 treatment_v + \beta_2 treatment_v \times after_t + A_{ivt}\Gamma + \alpha_v + \gamma_t + \varepsilon_{ivt}, \quad (2)$$

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

¹²The 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).

assignment in the unexposed districts. Thus, we estimate

$$\begin{aligned}
z_{ivt} = & \beta_1 treatment_v + \beta_2 treatment_v \times midline_t + \beta_3 treatment_v \times endline_t + \\
& \beta_4 treatment_v \times Ahmednagar_v + \beta_5 treatment_v \times midline_t \times Ahmednagar_v + \\
& \beta_6 treatment_v \times endline_t \times Ahmednagar_v + A_{ivt}\Gamma + \alpha_v + \gamma_t \times \delta_v + \varepsilon_{ivt},
\end{aligned} \tag{3}$$

and

$$\begin{aligned}
z_{ivt} = & \beta_1 treatment_v + \beta_2 treatment_v \times after_t + \\
& \beta_4 treatment_v \times Ahmednagar_v + \beta_5 treatment_v \times after_t \times Ahmednagar_v + \\
& A_{ivt}\Gamma + \alpha_v + \gamma_t \times \delta_v + \varepsilon_{ivt},
\end{aligned} \tag{4}$$

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).¹³ 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. Additionally, we replicate our main result omitting each village in turn.

¹³For 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.

2.4.2 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 Spears (2012a).

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 Results

What did the community sanitation intervention achieve? As outlined in section 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.

3.1 Balance of observed baseline properties

Did the random assignment of villages to treatment and control groups achieve balance on observed baseline characteristics? That is, do the data give reason to be comfortable claiming that the groups are truly comparable, as one would hope if randomization were accomplished? Table 2 shows that the answer is yes, both for the district Ahmednagar where the program was implemented, and for 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.

3.2 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 midline 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, in general, only barely statistically significant. The exact two-sided p -value differs in 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), although note that a one-sided test is arguably preferable to this two-sided test, and would be higher-powered.

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,¹⁴ 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.

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 to the TSC throughout India that Spears (2012a) finds 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,¹⁵ 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 1 plots the CDFs of village latrine coverage for the treatment and control groups in Ahmednagar

¹⁴If 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.

¹⁵In the baseline data, about a quarter of latrine-owning households still practiced open defecation.

in the endline data.¹⁶ 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. Thus, even villages that fell far short of total sanitation had some improvement.

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

¹⁶To be clear, it is not the case that subtracting the curves gives a distribution of treatment effects.

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.

3.4 An effect on children’s height

Did the program increase children’s heights? Table 3 presents regression evidence from Ahmednagar that it did. The table reports results from 16 specifications in order to demonstrate 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 \times 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 \times time), midline and endline collapsed into “after,”
- **Panel C:** Triple difference (Nanded & Nandurbar included, treatment \times time \times 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 \times time \times 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 regression¹⁷ to demonstrate that this truncation has little effect. Section 3.4.4 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

¹⁷A’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.

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 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 difference¹⁸ (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.

3.4.1 Improvement in height, but not to WHO standard

How large is the estimated effect on children’s height? One way to understand the effect is to compare it with Spears’s (2012a) estimates of the effect of the government’s Total Sanitation Campaign throughout India. Averaging over incomplete and heterogeneous implementation,

¹⁸As 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 ($t = 1.38$).

Spears finds 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 than 24 months are about 2 standard deviations below the WHO reference mean, and the children in our study are even shorter. Figure 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.¹⁹ 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.

¹⁹Note 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.

3.4.2 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 (Spears, 2012a,c) 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 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 subset²⁰ (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.

²⁰The 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.

3.4.3 Differences throughout the height distribution

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

Panel A of figure 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 keeping in mind their baseline separation among shorter children). This suggests that improved sanitation helped relatively tall and short children both. 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 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 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.

3.4.4 Mismeasured ages and truncation are not important issues

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 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. 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 3.4, 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.

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.

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 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.²¹ If we did not know that the program simply was 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).

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 assess 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

²¹This 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.

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.²² 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;²³ nor was such an interaction built into the study design by *a priori* blocking. Thus, this approach offers little evidence about treatment effect heterogeneity.

²²This suggests another statistical significance test for the paper’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.

²³Spears (2012a) finds an interaction between population density and the effect of a sanitation program using variation at the district level across 500 districts.

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 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 Spears (2012a) in a retrospective study of the average effect of the TSC throughout rural India. It is unsurprising that we find a larger effect. First, Spears’ 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 approximately 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. Economic theory would 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 does not have such externalities,²⁴ this argument implies that sanitation should be a health policy priority, including before the provision of private health goods with small welfare consequences.

References

- A'Hearn, Brian.** 2002. “A restricted maximum likelihood estimator for truncated height samples.” *Economics and Human Biology*, 2(1): 5–19.
- Allcott, Hunt and Sendhil Mullainathan.** 2012. “External Validity and Partner Selection Bias.” working paper 18373, NBER.
- 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.
- Bhandari, Nita, Rajiv Bahl, Sunita Taneja, Mercedes de Onis, and Maharaj K.**

²⁴Of 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.

- 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.
- Bongartz, Petra and Robert Chambers.** 2009. “Beyond Subsidies Triggering a Revolution in Rural Sanitation.” *IDS In Focus*(10).
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. “Bootstrap-Based Improvements for Inference with Clustered Errors.” *The 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.
- 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.
- 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.
- 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.** 2007. “Height, health and development.” *Proceedings of the National Academy of the Sciences*, 104(33): 13232–13237.
- 2012. “Searching for Answers with Randomized Experiments.” presentation, NYU

Development Research Institute.

- 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.
- Humphrey, Jean H.** 2009. “Child undernutrition, tropical enteropathy, toilets, and hand-washing.” *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.
- Jayachandran, Seema and Rohini Pande.** 2012. “The Puzzle of High Child Malnutrition in South Asia.” presentation slides, International Growth Centre.
- Joint Monitoring Programme for Water Supply and Sanitation.** 2012. *Progress on Drinking Water and Sanitation: 2012 Update*: WHO and UNICEF.
- 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.
- Kumar, Santosh and Sebastian Vollmer.** 2012. “Does access to improved sanitation reduce childhood diarrhea in rural India.” *Health Economics*.
- 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.
- McKenzie, David.** 2012. “Beyond baseline and follow-up: The case for more T in experiments.” *Journal of Development Economics*, 99: 210221.
- 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*.
- 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.
- 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 Experiential Learning (‘e’) to Crawl the Design Space.” working paper, Princeton University Research Program in Development Studies.
- 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.
- 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.** 2012a. "Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from India's Total Sanitation Campaign." working paper, Princeton.
- 2012b. "Height and Cognitive Achievement among Indian Children." *Economics and Human Biology*, 10: 210–219.
- 2012c. "How much international variation in child height can sanitation explain?" working paper, Princeton University Research Program in Development Studies.
- Spears, Dean and Sneha Lamba.** 2012. "Effects of Early-Life Exposure to Rural Sanitation on Childhood Cognitive Skills: Evidence from India's 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): 455–468.
- Vogl, Tom.** 2011. "Height, Skills, and Labor Market Outcomes in Mexico." Working Paper.
- Watson, Tara.** 2006. "Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on U.S. Indian reservations." *Journal of Public Economics*, 90(8-9): 1537 – 1560.
- Zwane, Alix Peterson and Michael Kremer.** 2007. "What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review." *World Bank Research Observer*, 22(1): 1–24.
- Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S. Karlan, Richard Hornbeck, Xavier Giné, Esther Duflo, Florencia Devoto, Bruno Crepon, and Abhijit Banerjee.** 2011. "Being surveyed can change later behavior and related parameter estimates." *PNAS*, 108(5): 1821–1826.

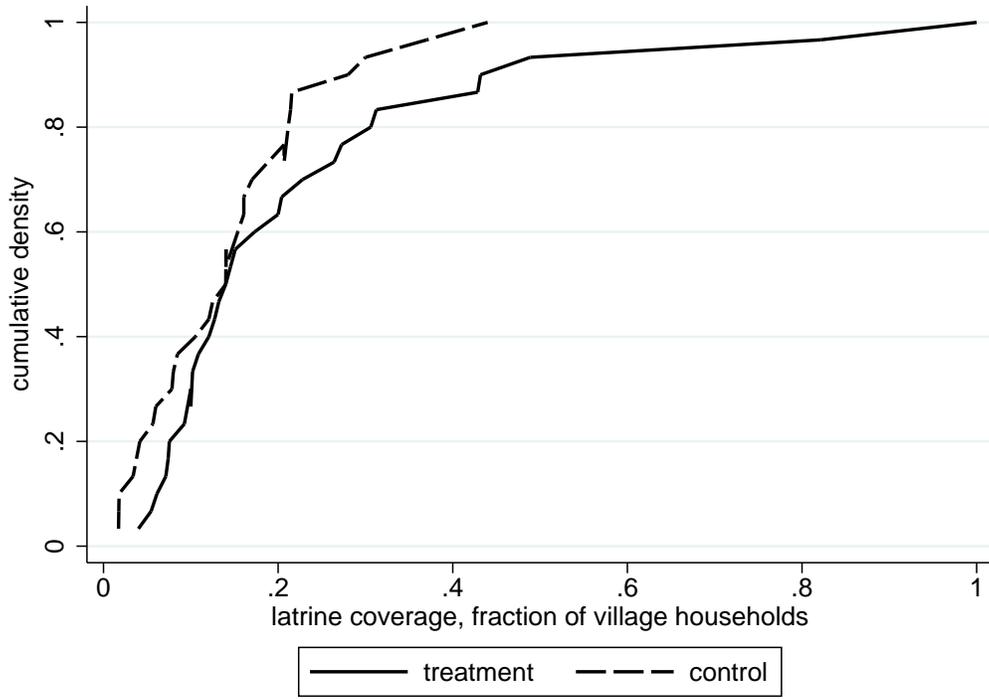


Figure 1: Distribution of village-level sanitation in Ahmednagar district, endline survey

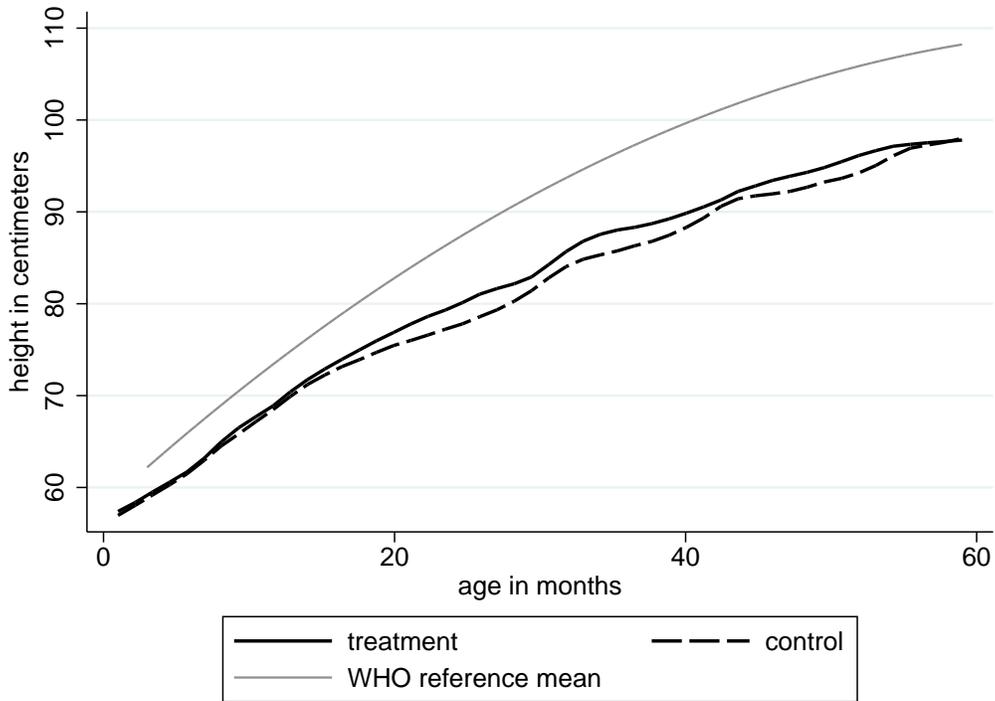


Figure 2: Height of children in Ahmednagar district by age, endline survey

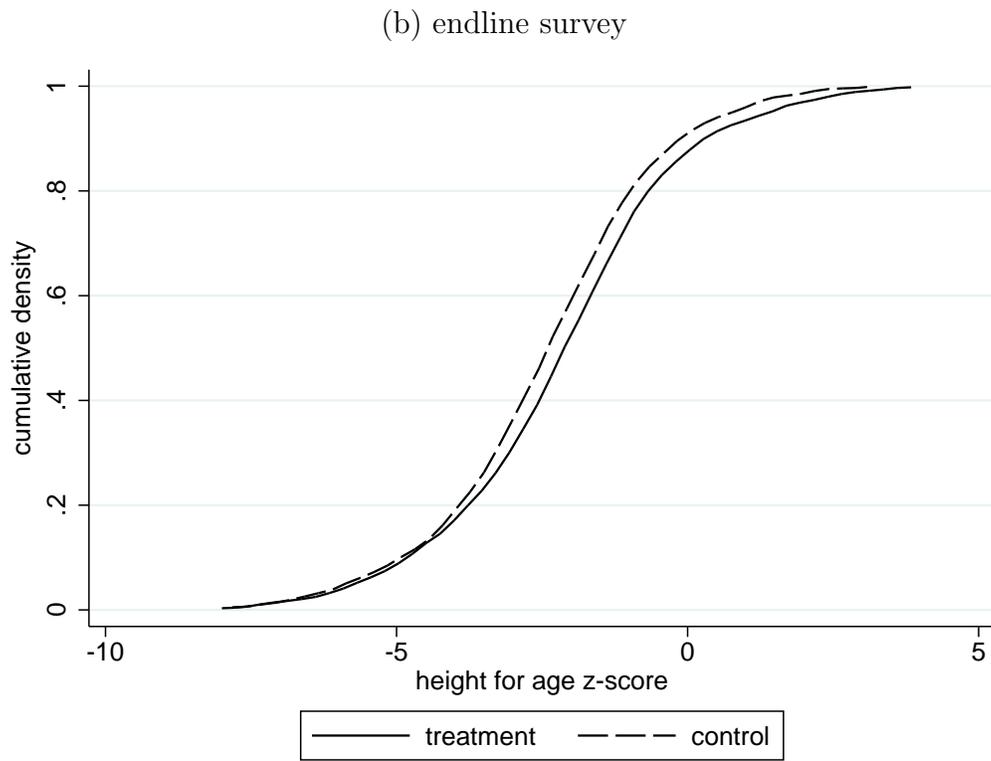
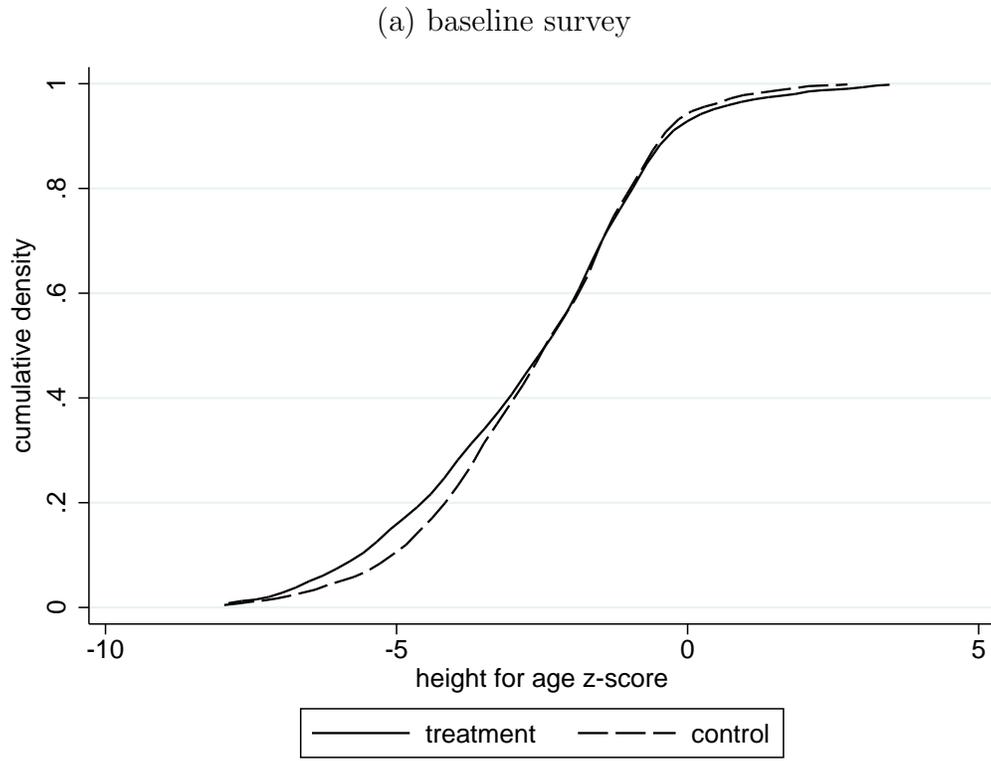


Figure 3: Distribution of height of children in Ahmednagar district

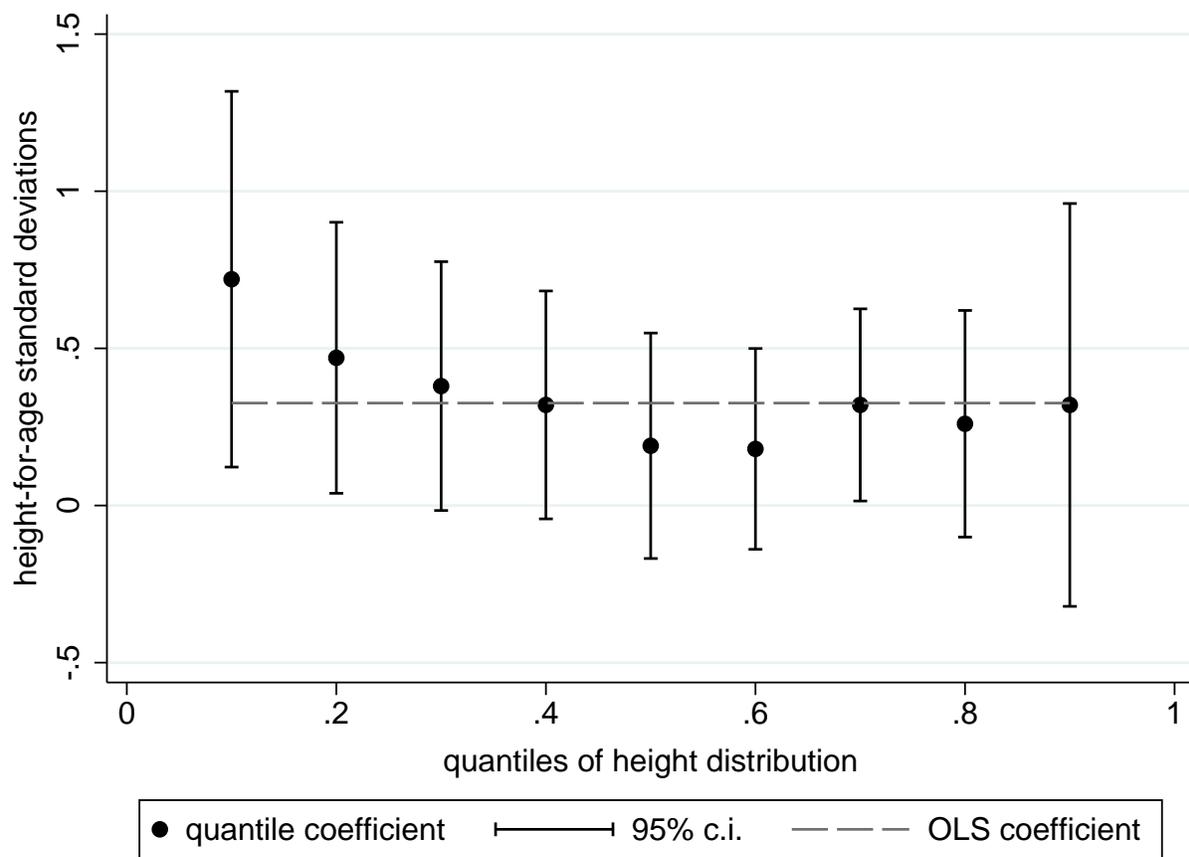


Figure 4: Quantile regression estimates of effect of program, Ahmednagar district

Table 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 (per 1,000)	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 2: Balance of baseline measurement sample means

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

Table 3: Effects of the experimental program on height-for-age in Ahmednagar

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
round \times dist FEs	OLS	OLS	vil. FE	truncated	OLS	OLS	vil. FE	truncated
age \times sex	✓	✓	✓	✓	✓	✓	✓	✓
village FEs			✓	✓			✓	✓
					round \times dist FEs			
					age \times sex			
					village FEs			
Panel A: Double difference, midline and endline								
treatment	-0.105 (0.129)	-0.0988 (0.129)	0.274 [†] (0.136)	-0.101 (0.130)	treatment	-0.105 (0.129)	-0.0992 (0.129)	-0.102 (0.130)
treatment \times midline	0.278 [†] (0.154)	0.236 [†] (0.140)	0.274* (0.136)	0.241 [†] (0.141)	treatment	0.326* (0.160)	0.324* (0.146)	0.330* (0.147)
treatment \times endline	0.379 [†] (0.211)	0.418* (0.195)	0.448* (0.190)	0.427* (0.195)	\times mid. or end.			
n (children)	3,432	3,432	3,432	3,432	n (children)	3,432	3,432	3,432
Panel B: Double difference, before and after								
treatment	-0.105 (0.129)	-0.0988 (0.129)	0.274 [†] (0.136)	-0.101 (0.130)	treatment	-0.105 (0.129)	-0.0992 (0.129)	-0.102 (0.130)
treatment \times midline	0.278 [†] (0.154)	0.236 [†] (0.140)	0.274* (0.136)	0.241 [†] (0.141)	treatment	0.326* (0.160)	0.324* (0.146)	0.330* (0.147)
treatment \times endline	0.379 [†] (0.211)	0.418* (0.195)	0.448* (0.190)	0.427* (0.195)	\times mid. or end.			
n (children)	3,432	3,432	3,432	3,432	n (children)	3,432	3,432	3,432
Panel C: Triple difference, midline and endline								
treatment	0.0412 (0.172)	0.0501 (0.172)	0.250 (0.227)	0.0559 (0.192)	treatment	0.0412 (0.172)	0.0500 (0.172)	0.0557 (0.192)
treatment \times Ahm.	0.298 (0.237)	0.224 (0.232)	0.250 (0.227)	0.235 (0.250)	treatment \times Ahm.	0.431 [†] (0.226)	0.411 [†] (0.220)	0.432 [†] (0.237)
treatment \times Ahm. \times midline	0.572* (0.264)	0.609* (0.256)	0.646* (0.249)	0.640* (0.273)	\times mid. or end.	-0.147 (0.147)	-0.114 (0.114)	-0.122 (0.122)
treatment \times Ahm. \times endline	-0.147 (0.214)	-0.114 (0.212)	0.0154 (0.178)	-0.122 (0.232)	treatment	0.214 (0.214)	0.212 (0.212)	0.232 (0.232)
treatment \times Ahmednagar	-0.0200 (0.181)	-0.000374 (0.181)	0.0154 (0.178)	-0.00300 (0.200)	\times Ahmednagar	-0.105 (0.160)	-0.104 (0.161)	-0.113 (0.158)
treatment \times Ahmednagar \times midline	0.181 (0.192)	-0.211 (0.160)	-0.220 (0.158)	-0.226 (0.178)	treatment	-0.105 (0.160)	-0.104 (0.161)	-0.113 (0.158)
treatment \times Ahmednagar \times endline	0.160 (0.160)	0.160 (0.160)	0.158 (0.158)	0.178 (0.178)	\times mid. or end.			
n (children)	11,337	11,337	11,337	11,337	n (children)	11,337	11,337	11,337
Panel D: Triple difference, before and after								
treatment	0.0412 (0.172)	0.0501 (0.172)	0.250 (0.227)	0.0559 (0.192)	treatment	0.0412 (0.172)	0.0500 (0.172)	0.0557 (0.192)
treatment \times Ahm.	0.298 (0.237)	0.224 (0.232)	0.250 (0.227)	0.235 (0.250)	treatment \times Ahm.	0.431 [†] (0.226)	0.411 [†] (0.220)	0.432 [†] (0.237)
treatment \times Ahm. \times midline	0.572* (0.264)	0.609* (0.256)	0.646* (0.249)	0.640* (0.273)	\times mid. or end.	-0.147 (0.147)	-0.114 (0.114)	-0.122 (0.122)
treatment \times Ahm. \times endline	-0.147 (0.214)	-0.114 (0.212)	0.0154 (0.178)	-0.122 (0.232)	treatment	0.214 (0.214)	0.212 (0.212)	0.232 (0.232)
treatment \times Ahmednagar	-0.0200 (0.181)	-0.000374 (0.181)	0.0154 (0.178)	-0.00300 (0.200)	\times Ahmednagar	-0.105 (0.160)	-0.104 (0.161)	-0.113 (0.158)
treatment \times Ahmednagar \times midline	0.181 (0.192)	-0.211 (0.160)	-0.220 (0.158)	-0.226 (0.178)	treatment	-0.105 (0.160)	-0.104 (0.161)	-0.113 (0.158)
treatment \times Ahmednagar \times endline	0.160 (0.160)	0.160 (0.160)	0.158 (0.158)	0.178 (0.178)	\times mid. or end.			
n (children)	11,337	11,337	11,337	11,337	n (children)	11,337	11,337	11,337

Standard errors clustered by village. Two-sided p values: † $p < 0.10$; * $p < 0.05$.

Panels A and B include only Ahmednagar; panels C and D also include Nanded and Nandurbar

Table 4: Robustness of estimated effect on height to alternative extreme-value bounds

Panel A: Coefficients on treatment \times 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 \times 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.