The nutritional value of toilets: How much international variation in child height can sanitation explain?

Dean Spears*

First circulated: December 2012
This version: June 2013

Abstract

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, a paradox called "the Asian enigma" which has received much attention from economists.

Could toilets help children grow tall, while disease externalities from poor sanitation keep children from reaching their height potentials? This paper provides the first identification of a quantitatively important gradient between child height and sanitation, which can statistically explain a large fraction of international height differences. 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.

The effect of sanitation on human capital is quantitatively robustly estimated across these strategies, and does not merely reflect wealth or other dimensions of development. Open defecation, which is exceptionally widespread in India, can account for much or all of the excess stunting in India.

---

*Centre for Development Economics, Delhi School of Economics, Delhi, India. dean@riceinstitute.org.
I am grateful for helpful suggestions from and discussion with Harold Alderman, Luis Andres, Radu Ban, Robert Chambers, Juan Costain, Joseph Cummins, Jean Drèze, Marianne Fay, Ariel Fiszbein, Andrew Foster, Richard Guerrant, Lawrence Haddad, Derek Headey, Peter Lanjouw, David Levine, J.V. Meenakshi, Doug Miller, John Newman; from seminar participants at Princeton, Tufts, the University of Virginia, the Delhi School of Economics, LBSNAA, the World Bank (in DC, Delhi, and Phnom Penh), and IFPRI; and especially from Anne Case, Diane Coffey, Angus Deaton, Michael Geruso, and Jeff Hammer. Errors and interpretations are my own. Portions of this paper have previously circulated as "Sanitation and open defecation explain international variation in childrens height: Evidence from 140 nationally representative household surveys," which this working paper supersedes.

1 Introduction

Physical height is a topic of increasing 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).

Could toilets help children grow tall, while disease externalities from poor sanitation keep children from reaching their height potentials? Sanitation has received little attention in economists' recent investigations of the puzzle of Indian malnutrition (*e.g.* Deaton, 2007; Tarozzi, 2008; Jensen, 2012; Jayachandran and Pande, 2012; Panagariya, 2013; Shah and Steinberg, 2013). Medical research documents that chronic childhood environmental exposure to fecal germs could be an important cause of stunting. Sanitation coverage is exceptionally poor in India, where over half of households defecate in the open without using a toilet or latrine, a much larger fraction than in other countries with similar incomes.

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, defecate in the open without using any toilet or latrine. The primary contribution of this paper is to document that much of the variation in child height among developing countries can be explained by differences in the prevalence of open defecation. I find quantitatively similar effects of sanitation on child height that are estimated from international heterogeneity across countries; that are identified from changes over time within Indian districts; and that are associated with differences among Indian and African rural localities.

This paper makes several contributions 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 (*e.g.* Pritchett and Summers, 1996), the importance of sanitation has received much less attention from economists. Moreover, I show that sanitation predicts child height even conditional on income and other dimensions of heterogeneous economic development or infrastructure. 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 the effect of quadrupling GDP per capita.

Second, this paper contributes to a resolution of the puzzle of the "Asian enigma" of Indian stunting, which has received much recent discussion in the economics literature. Differences in open defecation are sufficient to statistically explain much or all of the difference in average height between Indian and African children. Third, the paper 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 in the open 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. Finally, these findings 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 paper use complementary strategies to study the the relationship between height and open defecation, each focusing on a different dimension of heterogeneity. Section 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 3 compares children within one country, introducing fixed-effect panel methods to repeated cross-section data, in order to study differences within Indian districts over time (Deaton, 1985). Section 4 considers whether the India-Africa height gap can be explained by heterogeneity in local 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 5 considers whether estimates of of the association between height and sanitation in this paper and from the literature are sufficient to account for the India-Africa gap.

1.1 Open defecation causes stunting

A growing literature in economics documents that physical height has its origins in early life health (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, 2012b). Although a child's low height-for-age is often called "malnutrition," many scholars prefer "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" (Waterlow, 2011, 1), which is often narrowly interpreted as food. Indeed, two existing literatures indicate that infectious disease is an important component of malnutrition (Smith et al., 2013) and that early-life exposure to fecal germs in the environment reduces children's subsequent height: a medical literature documenting biological mechanisms, and recent econometric literature identifying a causal effect of poor sanitation on stunting.

First, medical and epidemiological literatures document at least two mechanisms linking environmental open defecation to poor health and early life human capital accumulation: diarrhea and chronic enteropathy. It has long been documented that diarrhea could cause stunting due to loss of consumed nutrients (*e.g.* Guerrant et al., 1992). Ingestion of fecal pathogens as a result of living near poor sanitation is well-known to cause diarrhea (Esrey et al., 1991). Checkley et al. (2008) use detailed, high-frequency longitudinal data from five countries to demonstrate effects of childhood diarrhea on subsequent height.

Another mechanism of sanitation's effect on stunting that has been more recently documented in detail in the medical literature but may be quantitatively more important is chronic but subclinical "environmental enteropathy" (Humphrey, 2009), also historically known as "tropical sprue." Environmental enteropathy is caused by repeated fecal contamination which, through an inflammatory response, increases the small intestine's permeability to pathogens while reducing nutrient absorption. Such malabsorption could cause malnutrition of various forms, stunting, and cognitive deficits, even without necessarily manifesting as diarrhea or otherwise observable illness (*see also* Petri et al., 2008; Mondal et al., 2011; Korpe and Petri, 2012). Lin et al. (2013) show that children in Bangladesh who are exposed to more fecal environmental contamination are more likely to exhibit biological markers of enteropathy, and in turn suffer impaired growth. In longitudinal data from field sites in eight countries, Kosek et al. (2013) show that environmental enteropathy is associated with subsequent deficits in growth.

In a second literature, recent econometric studies identify effects of sanitation on child height. Spears (2012a) studies a government sanitation program in rural India. From 1999 until its replacement with a new program in 2012, the Indian central government operated a "flagship" rural sanitation program called the Total Sanitation Campaign (TSC). Averaging over implementation heterogeneity throughout rural India, Spears finds that the where the TSC was active, it reduced infant mortality and increased children's height, on average. 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. Similarly, Hammer and Spears (2012) report a randomized field experiment by the Maharashtra government, 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. Section 5.1 considers the estimates of these causally well-identified studies in the context of this paper's results.

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

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 defecated in the open. 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 in the open.

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 in the open 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 in the open.

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[1] and including within households (Jayachandran and Pande, 2012). For example, Coffey et al.

[1] As one example, sanitation is unlikely to account for the slower cohort-to-cohort growth in women's height than men's documented by Deaton (2008).

(2013) find that in rural joint Indian households, children of higher birth order brothers are taller, on average, than children of lower birth-order 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, as even children at the top of social divides are impacted by the open defecation of their neighbors.

Therefore, it is already well-known that open defecation can be bad for children's health; that early-life disease is one factor that 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 links among these facts. The results indicate that sanitation is a statistically important predictor of differences in the height of children in developing countries and that open defecation can explain heterogeneity that has proven to be of lasting interest to economists and of significance to human development.

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 2.2 documents that, across country means, height is associated with open defecation, with little change after controlling for GDP. Next, section 2.2.1 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, section 2.2.2 verifies that other dimensions of infrastructure or well-being do not similarly predict child height. Section 2.2.4 observes that children would be more exposed to fecal pathogens where population is more dense, and finds that open defecation interacts with population density. Section 2.2.5 documents that the association between height and open defecation is steeper among older children, consistent with an unfolding effect of accumulating exposure. Finally, section 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.

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. For computing population density, area is taken from the United Nations Population Division. All other variables are from DHS surveys.

Using these data, the basic regression I estimate is

$$height_{cy} = \beta_1 open\ \ defecation_{cy} + \beta_2 \ln\left(GDPpc_{cy}\right) + \alpha_c + \gamma_y + X_{cy}\theta + \varepsilon_{cy}, \tag{1}$$

[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 2.2.3.

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 fixed effects α_c, year fixed effects γ_y, time-varying controls X_{cy} (including the average height of adult women in a DHS survey round, that is the mothers of the children studied), and the log of GDP per capita, all of which are added separately in stages. Standard errors are clustered by 65 countries.

2.2 Regression results

Figure 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. Circles are proportional to countries' populations. 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 reports estimates of regression 1 and will be referenced throughout this section of the paper. The main estimate of a linear decrease in height of 1.24 standard deviations associated with changing the fraction defecating in the open from 0 to 1 is qualitatively similar to Spears's (2012a) 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

[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.

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 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.

Another way to see that the effect of open defecation is not spuriously reflecting differences in wealth is to example wealth groups within India. Panel (b) of figure 2 adds average height and exposure to open defecation in wealth subsets of India's 2005 DHS (the solid circles) to the basic plot of country mean height and sanitation in the same 140 DHS country years as in figure 1 (the hollow circles). Included in published DHS data is a classification of households into wealth quintiles based on asset ownership. Average height of children within these five 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

[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.

families have electricity, a radio, a refrigerator, and a motorcycle or car. Even these relatively rich Indian children are shorter than healthy norms; this is expected, because 7 percent of the households living near even these rich children defecate in the open. 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 within India.

2.2.1 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 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. Note that panel (b) continues to demonstrate an association despite not including any data from India.

Returning to table 1, column 4 adds country fixed effects.[5] A control is also added for the average height of mothers of measured children; this is partially to control for omitted heterogeneity in the uterine environment and mothers' human capital, and partially to account for 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

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

uterine growth. DHS surveys are categorized into six global regions;[6] column 5 adds six region-specific linear time trends $\sum_r \delta_r year_y$ – one for each region r – to rule out that the effect is driven by spurious changes in specific parts of the world. Neither of these additions importantly changes the estimate of the coefficient, although adding so many controls increases the standard errors.

Table 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.

2.2.2 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 adds time-varying controls. These include female literacy (an important predictor of child welfare), fraction of deliveries in institutional settings (as a measure of the health care system), fraction of women who know about oral rehydration, fraction of children fully immunized, 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.

One component of infant and young child feeding is proper breastfeeding. If two controls, one for the fraction of infants ever breastfed and one the fraction of children exclusively breastfed for the first six months (the central recommendation of the WHO) are further

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

[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).

added as measures of the quality of infant nutrition, neither of these is statistically significant (t statistics of 0.24 and 0.42, respectively), so the estimate of the effect of open defecation loses precision. Although not reported in the table, the coefficient on open defecation in panel A with all fixed effects and controls becomes slightly *larger* in absolute value, -1.14 with a standard error of 0.78, and therefore a has one-sided p-value of 0.076. This estimate is not statistically significantly different from the other estimates.

Table 3 isolates 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. However, as shown in panel B, none of these "placebo" predictors matter for child height, conditional on sanitation and 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, food availability, or institutional delivery.

2.2.3 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 an average 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 or in column 1 of table 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.

2.2.4 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 Spears (2012a) finds 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 country-level population density per square kilometer times the fraction of people reporting open defecation. Figure 4 reveals that this measure of exposure to fecal pathogens (in logs, due to wide, skewed 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 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 panel (a) and $F_{2,57}$ statistic of 4.5 ($p = 0.0155$) in panel (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 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.

As a final verification of this prediction, panel (b) figure 4 replicates the plot of average child height against open defecation per square kilometer from panel (a), but replaces the three Indian observations with separate observations for India's states from its 2005-6 DHS. Some Indian states are as large as large countries, and have very high population density. Note that the countries' circles are all scaled larger than before because the large Indian circles have been removed. The general linear trend is unchanged, with an R^2 of 0.52; within India, states with dense populations and much open defecation have the shortest children.

2.2.5 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 5 plots the coefficients from estimating the basic equation 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.

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 5 estimates regressions in the form of equation 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, 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.[8]

[8] 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 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, 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,

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. However, we have seen that this does not fully capture the important difference in the disease environment between these regions: children in South Asia have more neighbors who defecate in the open and they are more exposed because they live nearer. Controlling, instead, for the number of people defecating in the open 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 alone, 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 4.3 considers the decomposition of this difference in more detail, using child-level height data.

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 sanitation in India have been slow. 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

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.

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 cross-sectional DHS surveys.

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.[9] 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.[10]

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. In particular, I regress child height on the fraction of households reporting open defecation at two levels of aggregation: districts and villages (or more precisely, rural primary sampling units (PSUs)).[11] The mean PSU studied here contains 10 children under 3 used in these regressions. 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

[9]The third and most recent NFHS does not include district identifiers.

[10]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.

[11]Note that the NFHS did not return to the same village PSUs both survey round. For readers familiar with the DHS, although the regressions are estimated using the birth recode, open defecation rates are merged in from the collapsed household recode, because while all households' open defecation contributes to the disease environment, not all households have a child.

and do not have latrines; section 4.1 considers the econometric implications of these negative externalities in more detail.

Therefore, the regression specification is:

$$height_{idvt} = \beta_1 open\ defecation^d_{dt} + \beta_2 open\ defecation^v_{dvt} + \alpha_d + \gamma_t + X_{idvt}\theta + A_{idvt}\vartheta + \varepsilon_{idvt}, \quad (2)$$

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.[12] 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.[13] As a further control, state-specific linear time trends will be added to this specification. Results are presented with and without controls and fixed effects to verify robustness.

Because this "panel" data is constructed out of repeated cross sections (Deaton, 1985), and not all households in a district or PSU were surveyed, sample mean open defecation rates are potentially noisy estimates of the true fraction of households practicing open defecation.

[12] Panagariya (2013) 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.

[13] 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).

This measurement error in the independent variable will attenuate estimates of the effect of open defecation on child height. This will be addressed in some specifications by weighting by the square root of the number of households used to estimate PSU open defecation, as recommended by Deaton (1985); or by instrumenting for PSU open defecation using district-level open defecation, and therefore identifying the effect from change over time within districts.

3.2 Regression results

Table 6 presents estimates of the effect of open defecation in equation 2, without district fixed effects in Panel A and with district fixed effects in Panel B. The estimates show a consistent effect of open defecation: 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, seen in the difference between columns 1 and 2, 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.[14] In particular, an effect is seen in columns 4, 5, and 7 which add state-specific linear time trends to the full set of controls.

The coefficient on village open defecation is smallest in absolute value with OLS fixed effects and controls, although it is not statistically significantly different from the other estimates; this could reflect the well-known attenuating bias of fixed effects.[15] Weighted

[14]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).

[15]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.

estimates in column 5 are similar to OLS estimates. Estimates instrumenting PSU open defecation with district open defecation are larger in absolute value than corresponding OLS estimates, which is consistent with attenuation bias in OLS.[16]

Would any village-level asset or indicator of infrastructure or well-being have the same apparent effect as village-level sanitation? Adding village electrification and water averages to the implementation of equation 2, column 3 of panel B (not reported in the text), 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 statistically insignificant individual t-statistics of 1.15 and -0.59, respectively, with a joint F-statistic of 0.73. Adding these two placebo controls to the fully controlled IV specification in column 7 of panel B changes the coefficient estimate on open defecation only from -0.822 to -0.809 ($s.e. = 0.488$); again neither electrification nor water is statistically significant, with a p-value from a joint χ^s test of 0.88.

Consistency of fixed effects estimates, which subtract level differences across groups, depends on a properly linear specification. Column 8 demonstrates that a quadratic term for village-level open defecation is not statistically significant, and indeed changes signs with and without district fixed effects. This result supports linear decomposition methods used in the next section.

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 these questions using pooled child-level data from the rural parts of nine DHS surveys: India's 2005-6 NFHS-

[16] I thank David Lavine for suggesting a third strategy: an application of Sullivan's (2001) Heteroskedastic Errors in Variables Estimator which weights each observation by an observation-specific reliability ratio (where the observation-specific "noise" in this case is the standard error of the first stage estimate of collapsed open defecation rates), used in a fixed effects application by Federman and Levine (2010). Using HEIV weights, I estimate an effect of PSU open defecation with district fixed effects and controls to be -0.357 (s.e. = 0.151), in place of the -0.353 in column 3 of Panel B.

3 and eight surveys from Africa in the 2000s.[17] In particular, the DHS surveys nearest 2005 (and balanced by year before and after) were selected from the five largest African countries available;[18] 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 4.1 verifies an association between village-level sanitation and height within the two regions. Then, section 4.2 considers an apparent 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 4.3 proceeds with the decomposition, applying linear and non-parametric approaches to explain the India-Africa gap.

4.1 Height and 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 indeed predicts children's height within each region. As in section 3, I use household-level DHS data to compute the fraction of households in a PSU reporting open defecation, which I treat as an estimate of village-level open defecation. This is matched to child-level height data. These 26,834 observations of individual children's height from India and 44,216 observations from Africa, each matched to its exposure to local (PSU) open defecation, will be the data used throughout this section 4 of the paper.

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

[18]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.

Figure 6 plots, separately for the African and Indian samples, non-parametric local regressions of child height on village open defecation, separating households that do and do not defecate in the open. 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. 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, whether or not their own households dispose of feces safely. Of course, some fraction of these correlations also reflects omitted variable bias.

How do India and Africa compare? The dashed vertical lines show that open defecation is much more common in the Indian than in the African data. Moreover, children in households that do not practice open defecation are shorter in Africa than in India at all levels of village open defecation. Finally, note that the association between open defecation and height is steeper in India than in Africa, which could reflect the difference in population density.

4.2 A paradox: International differences in well-being

Before estimating the decomposition, this subsection considers an apparent paradox in the economics literature on height. Deaton (2007) finds that international differences in height are not well explained by differences in GDP or infant mortality. How could it be height would not be correlated with wealth internationally, given that poor sanitation increases infant mortality (Spears, 2012a), and richer people are more likely to have toilets or latrines?

Figure 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.[19] In other words,

[19]In particular, it is well known that the relationship in the pooled data also depends upon the relationship among group 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

the slopes between income and height estimated using data from within each region in turn could be very different from the slope estimated using all the data, if the correlation of regional averages goes in a different direction.

In this case, Africa has higher mortality rates and less wealth than India, even though it has lower open defecation rates. Figure 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 (here, a count of common household assets included in the DHS). However, India has more open defecation, lower infant mortality, and more wealth, represented by the plotted circles. Health is not uni-dimensional and diseases are qualitatively different, so it is quite possible that Africa (with different burdens of HIV/AIDS and malaria and different diets) has more infant mortality than India at the same level of open defecation. Therefore, the pooled regressions are misleadingly flat. Consistently with Deaton's original null result for international height differences, pooling across regions shows no association between wealth or IMR and open defecation, even though open defecation is importantly associated with height.

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

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

$$\hat{\beta} = \sum_{s \in S} \lambda_s \hat{\beta}_s + \lambda_b \hat{\beta}_b, \tag{3}$$

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.

have a causal interpretation depends on the context and the sources of variation in independent variables, which is why this decomposition follows section 3's fixed effects estimates. This section estimates the fraction of the India-Africa height gap that can be statistically "explained" by differences in village-level sanitation, a main result of the paper.

4.3.1 Methods of decomposition

Three methods of decomposition are used. The first is a straightforward application of regression, as in table 5. I regress

$$height_{ivs} = \alpha India_s + \beta open\ defecation^v_{vs} + X_{ivs}\theta + \varepsilon_{is}, \tag{4}$$

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. The statistical significance of the change in $\hat{\alpha}$ is evaluated with a Hausman χ^2 test.

The second method is a weighted two-way Blinder (1973)-Oaxaca (1973) decomposition. In particular, having seen in section 4.2 that open defecation has different correlations within and across the Indian and African samples, I implement Reimers's (1983) and Jann's (2008) recommendation to first estimate

$$height_{ivs} = \beta_s open\ defecation_{vs} + X_{ivs}\theta_s + \varepsilon_{is}, \tag{5}$$

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

$$\left(0.5\hat{\beta}_{\text{India}} + 0.5\hat{\beta}_{\text{Africa}}\right)\left(\overline{open\ defecation_{v,\text{Africa}}} - \overline{open\ defecation_{v,\text{India}}}\right), \tag{6}$$

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.[20] In particular, the approach is to construct a counterfactual mean height of Indian children that matches the distribution of exposure to open defecation among African children:

- First, partition both samples into groups $g \in G(X)$, which share values or ranges of values of a set of covariates X. In this case, I construct 20 categories: an indicator for household open defecation interacted with 10 percentage point bins of PSU open defecation.

- Next, for each group g, for each region, compute $f(g\,|s) = \frac{\sum_{i \in g} w_{is}}{\sum_{g \in G(X)} \sum_{i \in g} w_{is}}$, the empirical density within each sample $s \in \{\text{India}, \text{Africa}\}$ in group g, using DHS sampling weights w_{is} for observation i in sample s.

- Finally, compute the counterfactual mean height of Indian children

$$\tilde{h}_{\text{India}} = \sum_{g \in G(X)} \sum_{i \in g} \frac{f(g\,|\text{Africa})}{f(g\,|\text{India})} w_i h_i, \tag{7}$$

  where h_i is the height-for-age z-score of child i in the Indian sample.

The unexplained gap is then $\tilde{h}_{\text{India}} - \bar{h}_{\text{Africa}}$.

4.3.2 Decomposition results

Table 7 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. Panel B reports the change in the unexplained difference when open defecation

[20]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.

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.

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.[21] 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 two 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.[22] Columns 5 and 6 control for a village-level estimate of infant mortality, computed from mothers' reported birth history. Notice that this *increases* the gap between India and Africa: children in India are shorter at the same level of infant mortality.

The counterfactual change in the India-Africa height gap upon accounting for open defecation is strikingly similar across methods and covariate specifications. In all cases, this change in the gap essentially matches or exceeds the original 0.14 standard deviation difference. In particular, the flexible non-parametric decomposition in Panel C might best accommodate any shape of the height-sanitation association. This approach shows the largest change, towards taller Indian children.

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

[22] Jayachandran and Pande (2012) 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 change in the gap similar to those when sanitation is controlled for in table 7. If, for example, IMR is controlled for before the decomposition, Indian first-borns are 0.132 standard deviations shorter than African first-borns, which increases by 0.16 to Indians being 0.025 taller.

5 Discussion: How much 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, 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?

5.1 A linear thought experiment

To answer the question above would require knowing the true average causal effect of open defecation on Indian children. 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.

Imagine a true linear causal effect, β, of sanitation on child height. Children in rural India are, on average, 0.142 standard deviations shorter than children in the rural African sample in section 4; open defecation is 31.6 percentage points more common in India. Therefore, the fraction of the rural India-Africa height gap that open defecation rates would explain would be whatever fraction β 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.

Table 8 collects estimates of the linear association between height and open defecation from this paper and others. Lin et al. (2013), applying epidemiological methods to a cross section in Bangladesh, find that children under four years old living in a clean sanitation environment are 0.54 height-for-age standard deviations taller than children living in a contaminated environment, after regression controlling for a wide list of DHS-style demographic and socioeconomic variables. Kov et al. (2013), whose results were circulated after and acknowledge the priority of this paper, use methods similar to section 3 to construct a panel out of two waves of Cambodia's DHS surveys. They find that Cambodian provinces which experienced greater improvements in sanitation from 2005 to 2010 also saw larger increases in average child height; their estimate of 0.47 is perhaps lower than estimates from South Asia due to Cambodia's much lower population density. Hammer and Spears (2012) report a field experiment that modestly improved sanitation coverage in Maharashtra; unsurprisingly, their instrumental variable treatment-on-the-treated estimate from a small experimental sample has a large confidence interval. Both their estimate and Spears's (2012a) may overstate the direct effect of latrines *per se* because the programs studied also promoted use of existing latrines; however, it is exactly latrine use, not ownership, that matters for child height.

What do these estimates together suggest? Like many regression estimates of the effects of inputs on human capital, some of these may be biased upwards. Collectively, however, they suggest that the best 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 India's high population density). If so, then sanitation could explain much or all of the difference in heights between Indian and African children. As in section 4.3.2, "explaining" over 100 percent of the gap is plausible because wealth differences predict that Indian children should be taller than African children.

Child height is increasingly well recognized as a critically important marker of development and human capital. Open defecation has a classic economic claim on public action due to negative externalities: 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. Indeed, early-life disease – and especially chronic disease due to fecal pathogens in the environment – appears to be an important determinant of height and stunting. If so, experimenting with methods to reduce open defecation may be an appropriate response to the Asian enigma.

References

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

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.

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

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

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.

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.

Deaton, Angus. 1985. "Panel data from time series of cross-sections." *Journal of Econometrics*, 30: 109–126.

——— 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.

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

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

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.

Federman, Maya and David I Levine. 2010. "Industrialization, pollution, and infant mortality." *Environment and Development Economics.*

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

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

Guerrant, RL, JB Schorling, JF McAuliffe, and MA de Souza. 1992. "Diarrhea as a cause and an effect of malnutrition: Diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration." *American Journal of Tropical Medicine and Hygiene*, 47(1): 28–35.

Hammer, Jeffrey and Dean Spears. 2012. "Effects of a village sanitation intervention on children's human capital: Evidence from a randomized experiment by the Maharashtra government." working paper, Princeton University.

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

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.

Jensen, Robert. 2012. "Another Mouth to Feed? The Effects of (In)Fertility on Malnutrition." *CESifo Economic Studies*, 58(2): 322–347.

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.

Kosek, Margaret, et al., and the MAL-ED network. 2013. "Fecal Markers of Intestinal Inflammation and Permeability Associated with the Subsequent Acquisition of Linear Growth Deficits in Infants." *American Journal of Tropical Medicine and Hygiene*, 88(2): 390–396.

Kov, Phyrum, Susanna Smets, Dean Spears, and Sangita Vyas. 2013. "Sanitation and Stunting in Cambodia." working paper, World Bank Water and Sanitation Program.

Lin, Audrie, Benjamin F. Arnold, Sadia Afreen, Rie Goto, Tarique Mohammad Nurul Huda, Rashidul Haque, Rubhana Raqib, Leanne Unicomb, Tahmeed Ahmed, John M. Colford Jr., and Stephen P Luby. 2013. "Household Environmental Conditions Are Associated with Enteropathy and Impaired Growth in Rural Bangladesh." *American Journal of Tropical Medicine and Hygiene*.

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.

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.*

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

Panagariya, Arvind. 2013. "Does India Really Suffer from Worse Child Malnutrition Than Sub-Saharan Africa?" *Economic and Political Weekly.*

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 and Lawrence Summers. 1996. "Wealher is Healthier." *Journal of Human Resources*, 31(4): 841–868.

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

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

Shah, Manisha and Bryce Steinberg. 2013. "Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital." working paper 19140, NBER.

**Smith, Michelle I., Tanya Yatsunenko, Mark J. Manary, Indi Trehan, Rajhab

Mkakosya, Jiye Cheng, Andrew L. Kau, Stephen S. Rich, Patrick Concannon, Josyf C. Mychalecky, Jie Liua, Eric Houpt, Jia V. Li, Elaine Holmes, Jeremy Nicholson, Dan Knights, Luke K. Ursell, Rob Knight, and Jeffrey I. Gordon. 2013. "Gut Microbiomes of Malawian Twin Pairs Discordant for Kwashiorkor." *Science*, 339(6119): 548–554.

Spears, Dean. 2012a. "Effects of Rural Sanitation on Infant Mortality and Human Capital: Evidence from a Local Governance Incentive in India." working paper, Princeton.

——— 2012b. "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.

Sullivan, Daniel G. 2001. "A Note on the Estimation of Linear Regression Models with Heteroskedastic Measurement Errors." working paper WP 2001-23, Federal Reserve Bank of Chicago.

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

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

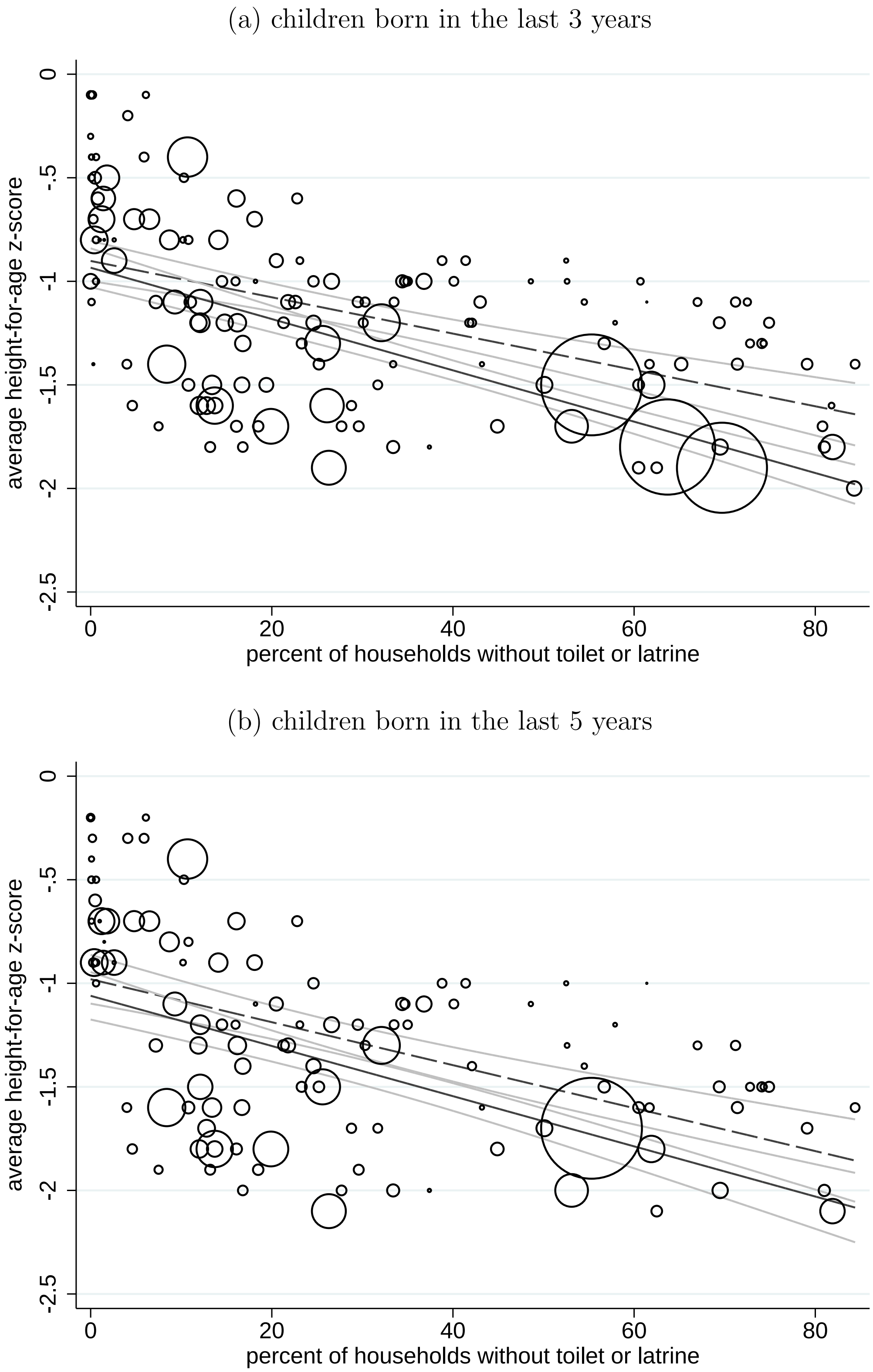


Figure 1: Open defecation predicts child height, across DHS survey round country-years

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

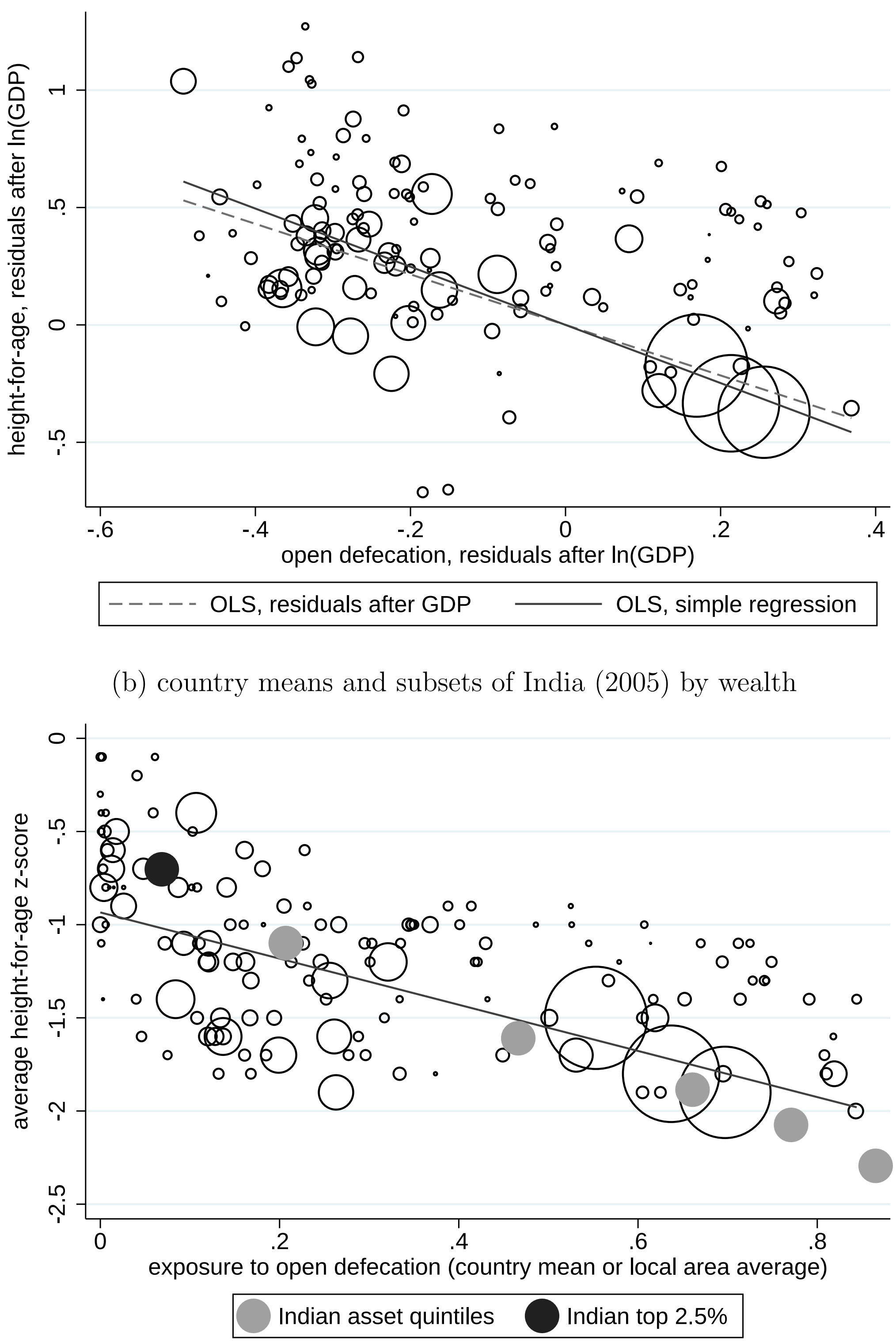
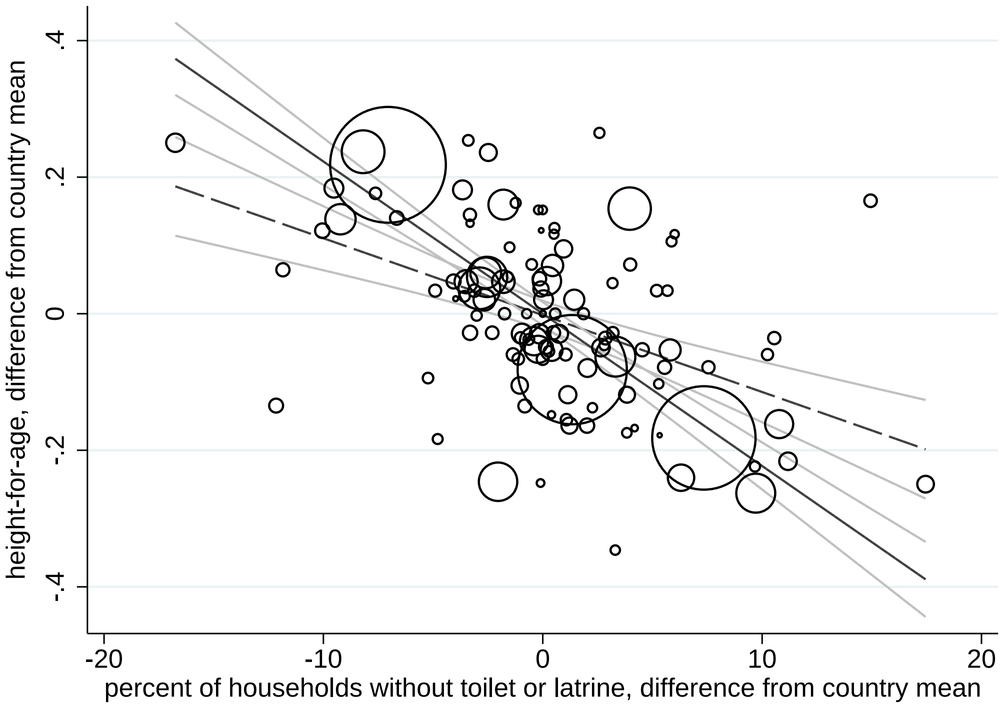


Figure 2: Wealth does not account for the association between child height and sanitation

OLS regression lines weight by country population; data from children under 3 years old. The "simple regression" line in panel (a) plots the slope of the uncontrolled regression of height on sanitation. Mean height of children in Indian wealth groups in panel (b) is plotted against the average rate of open defecation in the primary sampling units where they live.

(a) children born in the last 3 years



(b) children born in the last 5 years

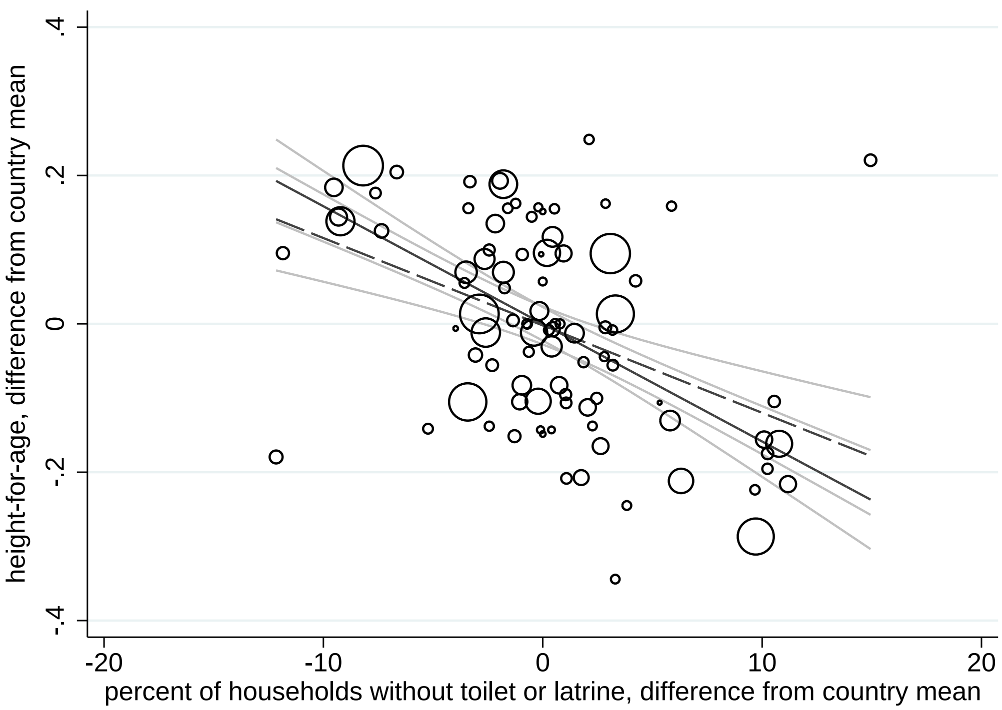
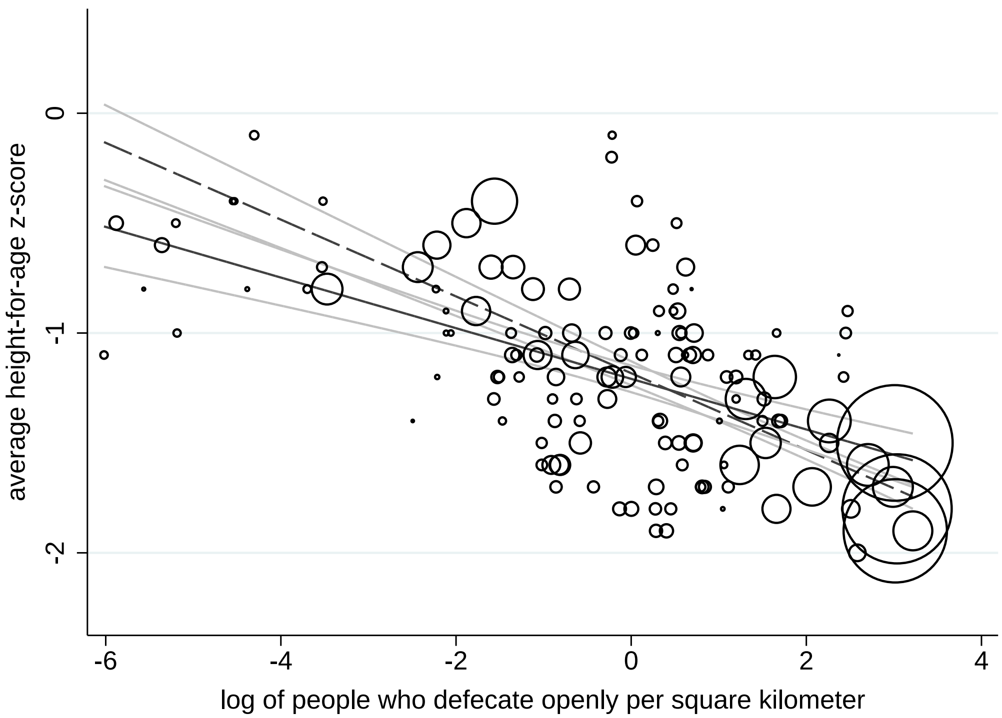


Figure 3: Deviation from country mean sanitation explains deviation from mean height
Solid OLS regression lines weight by country population; dashed lines are unweighted.

(a) DHS survey rounds (country-years) as observations



(b) DHS survey rounds and 2005 Indian states as observations

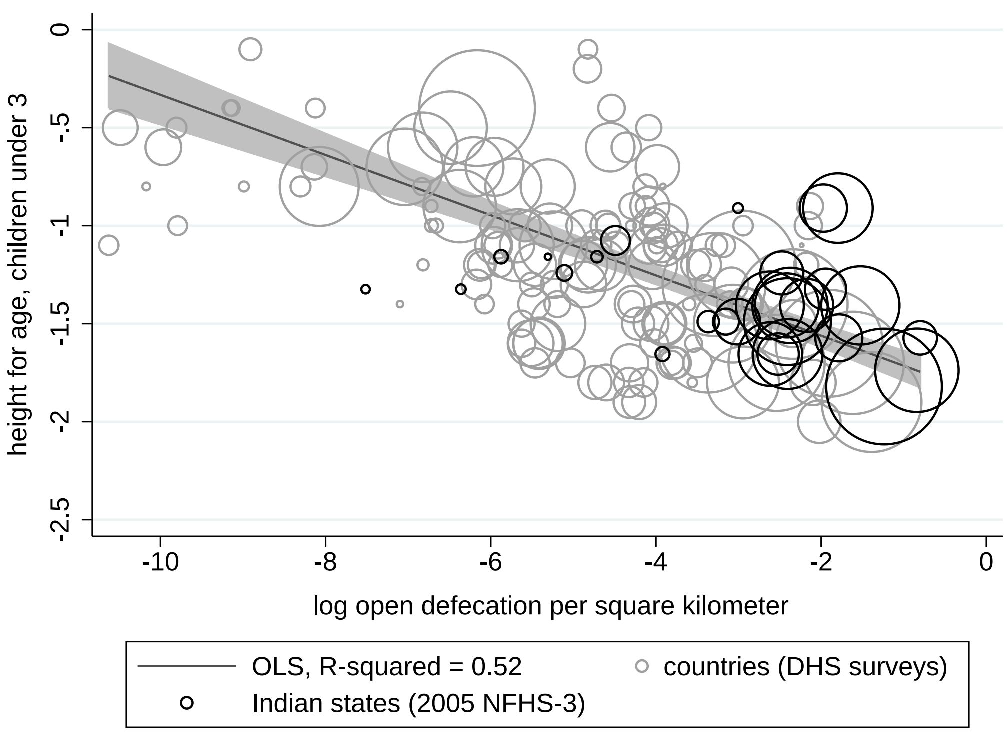


Figure 4: Open defecation interacts with population density to predict child height

Average heights of children under 3 years old. "Countries" in panel (a) are the same DHS survey rounds as in panel (a) of figure 1. In panel (b) the three country-year observations from India are removed and replaced with averages from Indian states from its 2005-6 DHS. Circles are scaled larger in panel (b) because the large India-wide observations are absent.

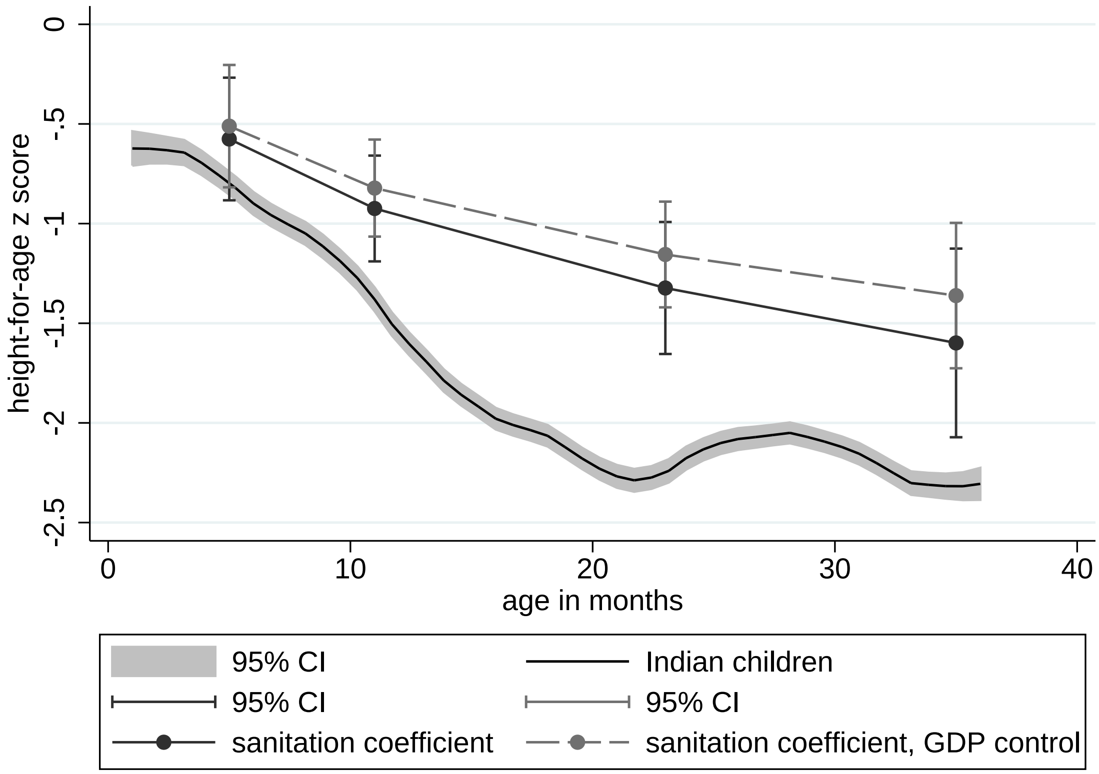
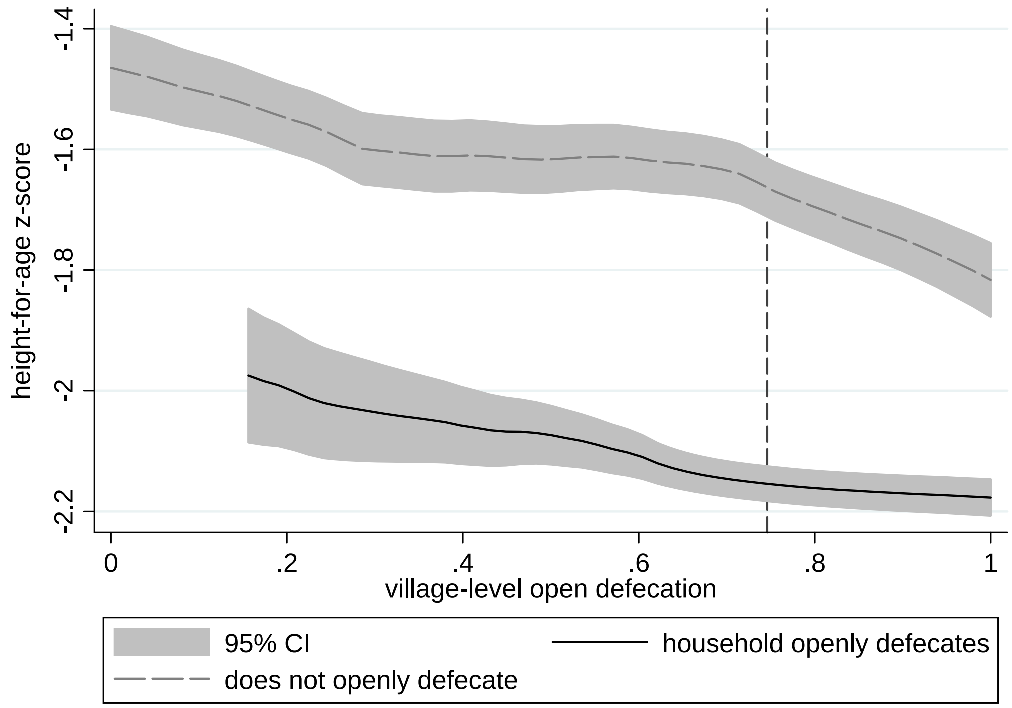


Figure 5: Open defecation is more steeply associated with child height at older ages
Confidence intervals are estimates of the coefficient from a regression of average country-level child height-for-age on open defecation, restricting the sample to four age categories. The curve plots the average height-for-age z-score of Indian children by age, for reference.

(a) India (NFHS-3)



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

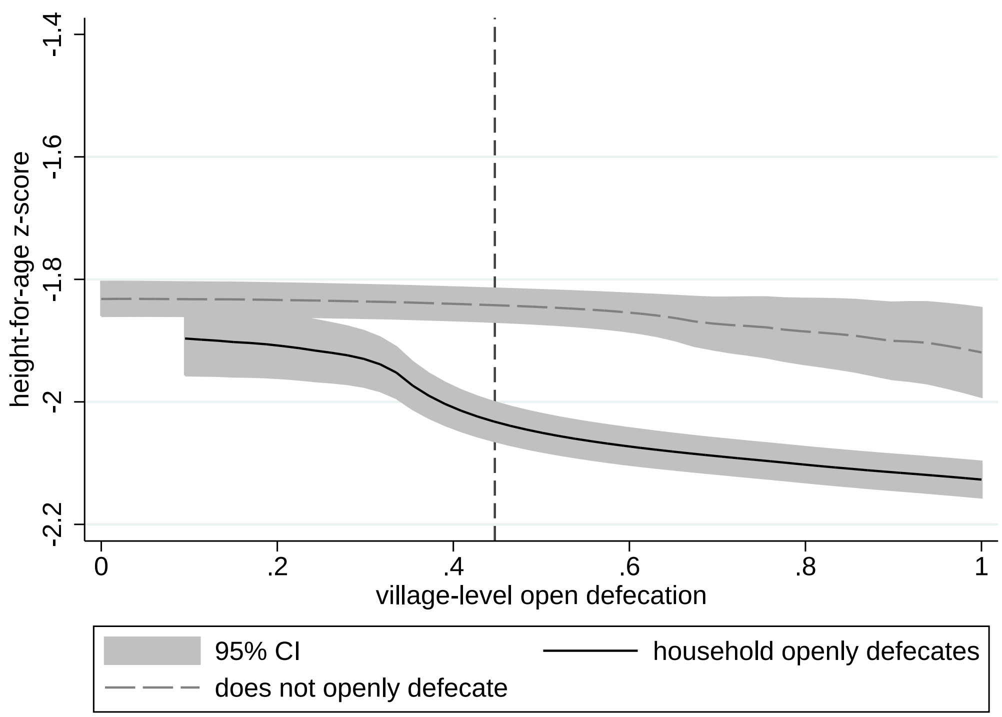


Figure 6: Negative externalities: Village-level open defecation predicts child height

Solid lines are local non-parametric regressions. Vertical dotted lines mark the overall mean open defecation fraction in these two rural samples.

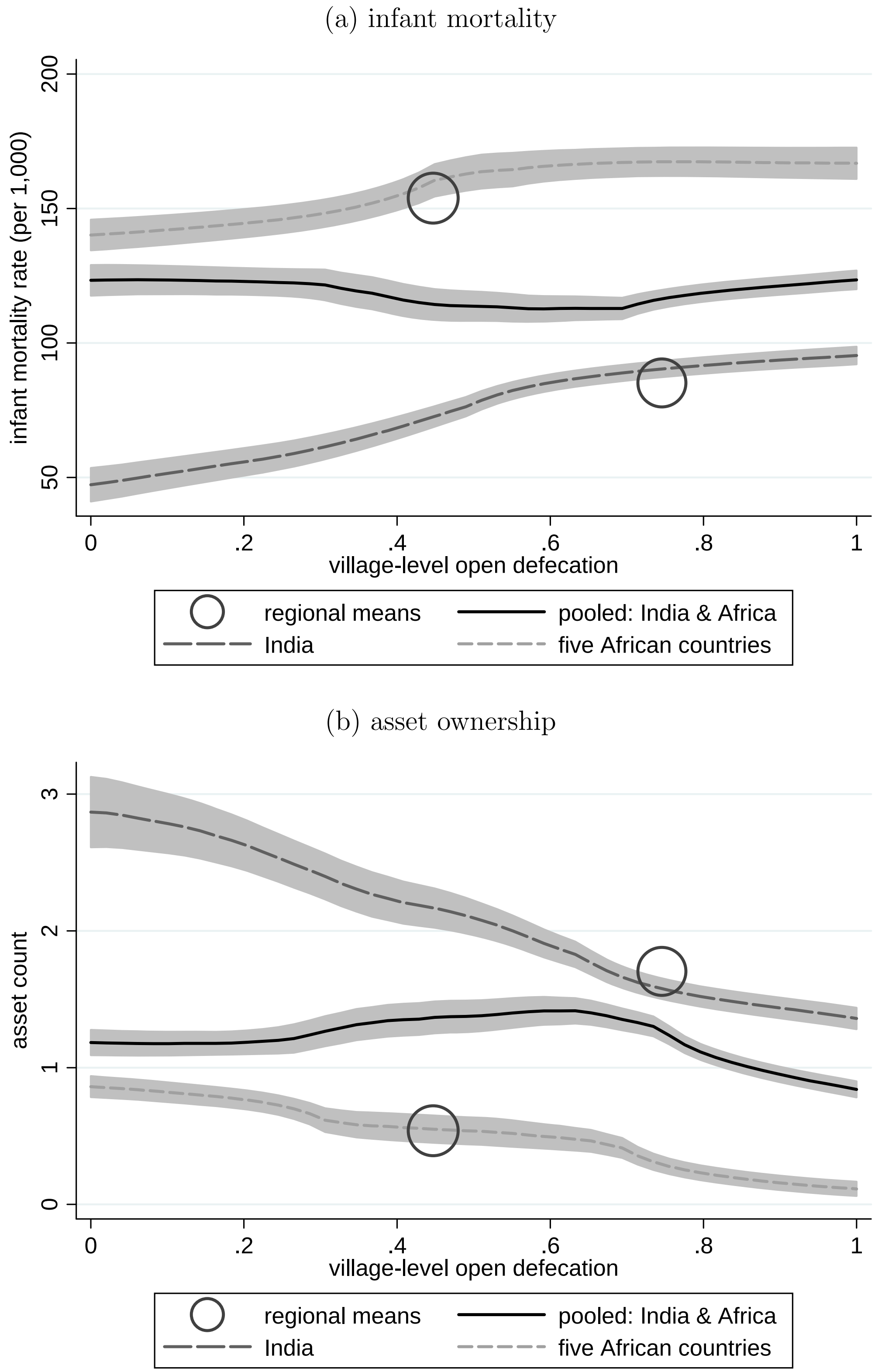


Figure 7: Simpson's paradox: open defecation and well-being across and within regions

Table 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†	-1.029†	-0.663***
	(0.226)	(0.158)	(0.156)	(0.434)	(0.583)	(0.552)	(0.181)
open defecation							-1.499*
$\times$ density							(0.631)
ln(GDP)			0.202**	0.512***	0.472**	0.817†	0.280***
			(0.0733)	(0.146)	(0.174)	(0.435)	(0.0525)
women's height				0.0130	-0.0564	-0.0540	0.0425**
				(0.0476)	(0.0904)	(0.137)	(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
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*
$\times$ 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, autocracy score, institutional delivery percent, and full immunization percent; see the text for more complete variable definitions.

Table 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: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Open defecation is a fraction 0 to 1.

Table 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	inst. delivery
Panel A: Unconditional regressions							
"placebo" variable	0.0143*	0.00478*	-0.00251*	0.00675*	1.070***	-0.0274*	0.015***
	(0.00367)	(0.00203)	(0.00120)	(0.00254)	(0.178)	(0.0117)	(0.002)
Panel B: Regressions conditional on open defecation and GDP							
"placebo" variable	0.00348	0.00382	0.0000230	-0.00271	-0.159	-0.0119	-0.001
	(0.00327)	(0.00268)	(0.000964)	(0.00226)	(0.138)	(0.00787)	(0.004)
open defecation	-0.880***	-1.077***	-1.002***	-0.918***	-1.003***	-0.843***	-0.981*
	(0.205)	(0.158)	(0.157)	(0.141)	(0.150)	(0.212)	(0.466)
ln(GDP)	✓	✓	✓	✓	✓	✓	✓
year FEs	✓	✓	✓	✓	✓	✓	✓
n (DHS surveys)	138	124	140	135	126	138	129
R^2	0.752	0.792	0.744	0.726	0.802	0.753	0.988

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 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
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 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: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 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 6: Effects of open defecation on child height: "panel" of Indian districts, NFHS-1 1992-3 to NFHS-2 1998-9

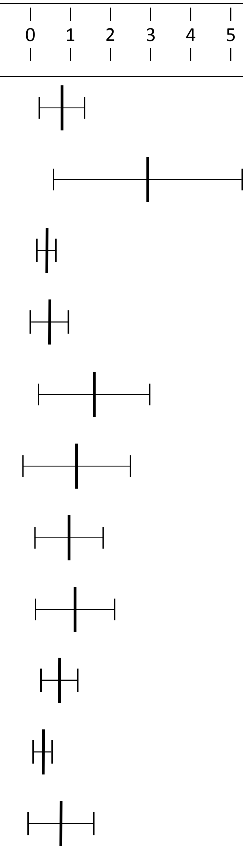
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
estimation:	OLS	OLS	OLS	OLS	WLS	IV	IV	OLS
Panel A: Without district fixed effects								
district open defecation	-0.779***	-0.279						
	(0.155)	(0.193)						
PSU open defecation		-0.537***	-0.523***	-0.447***	-0.486***	-0.838***	-0.574**	-0.774***
		(0.128)	(0.104)	(0.105)	(0.105)	(0.167)	(0.186)	(0.170)
PSU open defecation^2								-0.134
								(0.367)
Panel B: With district fixed effects								
district open defecation	-0.525†	-0.106						
	(0.307)	(0.321)						
PSU open defecation		-0.533***	-0.353**	-0.380***	-0.383***	-0.668†	-0.822*	-0.710***
		(0.125)	(0.117)	(0.113)	(0.113)	(0.394)	(0.418)	(0.179)
PSU open defecation^2								-0.548
								(0.337)
survey round FE	✓	✓	✓	✓	✓	✓	✓	✓
age-in-months × sex	✓	✓	✓	✓	✓	✓	✓	✓
control variables			✓	✓	✓		✓	
state time trends				✓	✓		✓	
n (rural children under 3)	23,588	23,588	23,588	23,588	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: † $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. OLS is ordinary least squares regression; WLS weights by the square root of the number of households sampled in the PSU to compute the PSU open defecation rate; IV instruments for PSU level open defecation with district level open defecation. Open defecation is a fraction 0 to 1. PSU stands for "primary sampling unit," here local village areas. 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. Only rural subsamples are used.

Table 7: Decomposition of rural India-Africa height difference: Fraction due to open defecation

	(1)	(2)	(3)	(4)	(5)	(6)
covariates:	none		birth demography		IMR (village-level)	
sanitation included:		✓		✓		✓
n (children)	71,048	71,048	71,048	71,048	71,048	71,048
Panel A: OLS regression, with an indicator for the Indian sub-sample:						
India indicator	-0.142	-0.001	-0.191	-0.049	-0.243	-0.099
	(0.026)	(0.026)	(0.026)	(0.027)	(0.026)	(0.028)
village-level sanitation		-0.480		-0.432		-0.418
		(0.035)		(0.035)		(0.035)
change in gap:	0.141		0.142		0.144	
fraction explained:	0.994		0.754		0.593	
test India dummies equal	$\chi^2_1 = 136$ $p = 0.00$		$\chi^2_1 = 130$ $p = 0.00$		$\chi^2_1 = 120$ $p = 0.00$	
Panel B: Oaxaca-Blinder decomposition, unexplained difference:						
unexplained difference	-0.142	-0.001	-0.187	-0.047	-0.243	-0.099
	(0.026)	(0.026)	(0.025)	(0.027)	(0.026)	(0.029)
change in gap:	0.141		0.140		0.144	
fraction explained:	0.994		0.749		0.593	
Panel C: non-parametric reweighting decomposition, counterfactual difference in means:						
difference in means	-0.143	0.061	-0.242	-0.022	-0.154	-0.005
change in gap:	0.204		0.220		0.149	
fraction explained:	1.43		0.908		0.967	

Table 8: What fraction of the India-Africa height gap can open defecation explain?

estimation strategy	source	slope		fraction explained
experiment, OLS	Hammer & Spears (2012)†	0.786		1.748
experiment, TOT IV	Hammer & Spears (2012)†	2.928		6.514
Cambodia, place & time FEs	Kov, *et al.* (2013)	0.469		1.042
Bangladesh, cross-section	Lin, *et al.* (2013)**	0.54		1.20
India's TSC, high	Spears (2012)†	1.592		3.542
India's TSC, low	Spears (2012)†	1.153		2.565
DHS means, FEs	Table 1	0.962		2.140
DHS means, controls	Table 1	1.029		2.287
DHS means, SA&SSA	Table 5*	0.721		1.604
India, district FEs	Table 6	0.380		0.844
India, district FEs & IV	Table 6	0.822		1.827

"Fraction explained" computes the fraction of the India-Africa height gap that would be explained if each row's estimate of the sanitation-height slope were true; the fraction is obtained by dividing the point estimate by 0.45. † These are probably overestimates of the effects of sanitation infrastructure because the programs they study also motivate use of existing latrines. *The "Table 5" estimate is not, in fact, from table 5, but is found by estimating the regression in column 3 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 who defecate in the open. ** Lin *et al.*'s estimate is weakly an underestimate of β because it is the difference in height linearly associated with the difference between a set of clean and contaminated sanitary environment, which will be some subset of the span between 0% and 100% open defecation. Lin *et al.* report two significant digits.