Place and Child Health: The Interaction of Population Density and Sanitation in Developing Countries

Diane Coffey[1,2,3], Sabrina Haque[4], Payal Hathi[2], Lovey Pant[2], and Dean Spears[2,3,5]

December 2016. Forthcoming *Demography*

Diane Coffey (corresponding author)
e-mail: coffey@utexas.edu

[1] University of Texas, Austin. Department of Sociology & Population Research Center.

[2] r.i.c.e., a research institute for compassionate economics, India

[3] Indian Statistical Institute, Delhi, India

4 World Bank Water and Sanitation Program, Washington D.C.

5 University of Texas, Austin. Department of Economics.

Abstract A long literature in demography has debated the importance of place for health, especially children's health. In this study, we assess whether the importance of dense settlement for infant mortality and child height is moderated by exposure to local sanitation behavior. Is open defecation (i.e., without a toilet or latrine) is worse for infant mortality and child height where population density is greater? Is poor sanitation is an important mechanism by which population density influences child health outcomes? We present two complementary analyses using newly assembled data sets, which represent two points in a trade-off between external and internal validity. First, we concentrate on external validity by studying infant mortality and child height in a large, international child-level data set of 172 Demographic and Health Surveys, matched to census population density data for 1,800 subnational regions. Second, we concentrate on internal validity by studying child height in Bangladeshi districts, using a new data set constructed with GIS techniques that allows us to control for fixed effects at a high level of geographic resolution. We find a statistically robust and quantitatively comparable interaction between sanitation and population density with both approaches: open defecation externalities are more important for child health outcomes where people live more closely together.

Keywords Sanitation, Population density, Infant mortality, Child height, South Asia

1 Introduction

A long literature in demography explores the importance of place for health (Entwisle, 2007). In many cases, this has been characterized as a debate over the health consequences of living in urban settings versus rural settings (Woods, 2004; Dye, 2008; Sastry, 1997). Although many demographers who study the effects of urban residence on health in developed countries today find a strong urban advantage (Eberhardt et al., 2001; Hartley, 2004), historically, discussions of urban health have often begun with the history of poor sanitation and high infectious disease burdens that plagued the cities of now-rich countries while they were developing (Preston, 1975; Cutler and Miller, 2005).

In modern developing countries, there is active debate about what defines "urbanness" (Hugo et al., 2003; Dorélien et al., 2013) and when and why urban advantages in child and infant health exist (Fink et al., 2014; Günther and Harttgen, 2012; Van de Poel et al., 2007; Smith et al., 2005; Montgomery and Hewett, 2005; Jankowska et al., 2013). Bocquier et al. (2011) point out that urban advantages depend on the services and economic opportunities that a city provides, while Sastry (1996) points out that the effects of community-level variables on child health often depend on context; that is, that when exploring the effects of place on health, interactions are often important.

In developing countries today, dense settlement often implies a number of health advantages for children. It is correlated with more wealth, which buys better housing and food, and with more schooling, which leads to better educated mothers. Additionally, people in densely populated areas are more likely to have access to health services that matter for child survival and development, such as trained doctors, maternal care and medicines (Magadi et al., 2003; Matthews et al., 2010).

However, scholars have also hypothesized that one important reason why place matters for health in developing countries today, and why it mattered in developed countries historically, is variation in sanitation and the disease environment (Mosley and Chen, 1984; Preston and Haines, 1991; McGuire and Coelho, 2011). Recent research in economics, epidemiology and public health suggests that open defecation, the practice of defecating in the open without using a toilet or latrine, is an important cause of infant mortality and stunting in both rural and urban areas of developing countries (Humphrey, 2009a; Fink et al., 2011; Cameron et al., 2013; Spears, 2013).

In this paper we assess whether the importance of dense settlement for infant mortality and child height is moderated by exposure to local, or community-level, sanitation behavior. We ask: does sanitation interact with population density to produce these child health outcomes? Such an interaction would be consistent with facts and theories in the literature. If open defecation reduces human capital by releasing germs into children's environments, then it is plausible that the consequences of open defecation would be worse where people live more closely together and are more likely to encounter their neighbor's germs.

Documenting and measuring the magnitude of the interaction between open defecation and population density is important for several reasons. First, it moves beyond dichotomous rural and urban distinctions and clarifies the circumstances under which population density is positively associated with health, and the circumstances under which poor sanitation is particularly harmful. Second, it contributes to understanding the importance of externalities or "spillover effects" of sanitation: one household's toilet use or open defecation has consequences for neighboring households' children. Such externalities are recognized in public economics as a central rationale for policy action. Finally, documenting and measuring such an interaction could guide policy decisions. Open defecation is increasingly concentrated in South Asia, a region where even rural areas are very densely populated (Coffey et al., 2014).

We present two complementary analyses: the first establishes the broad importance of the interaction between sanitation behavior and population density for predicting infant mortality and child height in developing countries, and the second provides evidence to support the internal validity of this interaction. For the first analysis, we construct a new international dataset from 172 Demographic and Health Surveys (hereafter DHS) collected in 69 developing countries between 1990 and 2012. Child level health data are matched with estimates of community open defecation rates and census population density data for 1,800 sub-national regions. For the second analysis, we use Geographic Information System (GIS) codes to create a new data set of children in Bangladesh that allows us to identify the effect of the interaction of population density with local sanitation on child height. These new data allow our measure of population density to be more precise than is possible in the international data set, and they allow us to control for higher-resolution geographic fixed effects.

We motivate the international analysis by confirming the results of prior papers which find that urban children in developing countries are less likely to die in the first year of life than rural children. Using the dataset of 172 DHS, we find that part of this difference is explained by the fact that rural children are exposed to more open defecation, on average, than urban children. However, a positive interaction of urban place with local open defecation suggests that the urban survival advantage is less pronounced where open defecation is high. Further controlling for the interaction of population density and local sanitation clarifies that higher average population density in urban areas is the mechanism through which urban residence likely moderates the effect of sanitation on infant mortality.

We then focus directly on the population density–sanitation interaction and show that it is robust to a variety of respecifications. We also perform falsification tests to show that other variables related to socioeconomic status do not similarly interact with population density to predict infant mortality in these data. Finally, we plot the shape of the interaction between local open defecation and population density and find that it is steeper at higher population densities.

The second analysis seeks to further test the internal validity of the interaction between sanitation and population density in predicting child height. We use GIS codes to match children in the Bangladesh DHS to the population density for their area of residence using highly disaggregated census data. This allows us to construct an interaction of population density and local sanitation that provides a more precise measure of exposure to density of open defecation than we are able to use in the international dataset. We then regress child height on these more precise measures of exposure to density of open defecation using district and survey round fixed effects. As in Sastry and Hussey (2003), we use geographic fixed effects because they control for time-invariant properties of place at the level of the fixed effect, in this case the district.[1] The magnitude of the interaction that we identify in the Bangladesh dataset is quantitatively similar to what is predicted for Bangladesh by a semi-parametric model fit to the international data.

This paper proceeds in three sections. First, section 2 summarizes evidence from the literature about why poor sanitation would be expected to have a larger effect on infant mortality and child height where population density is higher. Section 3 describes the analysis and presents results from the international dataset. Section 4 describes the analysis and presents results from the Bangladesh dataset. Section 5 discusses the findings. We point out that although, taken at face value, our results might seem to recommend concentrating policy efforts on improving sanitation in urban areas, the distributions of sanitation coverage and population density in the world today show that many of the places on earth where open defecation is most densely practiced are actually classified as rural. In-

[1] We do not present multi-level models because, as explained by Sastry and Hussey (2003), these models require the assumption that the random effects that are used in the models be independent of measured covariates. This independence criterion is not met in this case; for example, more urbanized districts have higher population density, on average.

deed, our findings, combined with these empirical distributions, highlight the threats to child health posed by the enduring density of open defecation in *rural* South Asia.

2 Background: Population density, sanitation and disease externalities

Rural places have lower population density than urban places on average, but also have more open defecation than urban places and lower quality sanitation, on average. Although developing countries are making progress in improving sanitation, over one billion people still defecate in the open, without using a toilet or latrine (WHO and UNICEF, 2012). Increasingly, open defecation is concentrated in rural areas, but it is also becoming increasingly concentrated in countries with high rural population densities, such as Indonesia, Pakistan, and especially India, where the 2011 census finds that 90 percent of households without a toilet or latrine live in rural areas.

Open defecation is a practice with strong negative health externalities: it spreads infectious diseases such as diarrhea, polio, cholera and parasites. Greater population density could exacerbate these negative externalities by providing more opportunities for disease transmission. Although there are several examples of population density-health interactions in present-day developing countries in the literature,[2] and there is evidence from present-day developed countries,[3] discussion of the evidence that population density can intensify an epidemiological externality often begins with the history of urbanization in now-rich countries.

Much has been written about the lethal combination of population density and poor sanitation in 19th century London. To illustrate an exemplary use of observational statistics, Freedman (1991) recounts John Snow's investigation of the 1853-54 cholera epidemic. By tracing deaths to the supply of their households' water, he demonstrated the nature of the epidemic, and is widely credited for

[2]For example, Root (1997) finds that population density is correlated with child mortality across provinces of Zimbabwe. A study of typhoid in Dhaka showed that crowdednesss has a considerable impact on the transmission and distribution of the disease: areas with low risk of typhoid were those with the lowest population density and those with the highest risk had the highest population density (Corner et al., 2013). Ali et al. (2002) show that higher population density is associated with a greater risk of cholera in a rural part of Bangladesh. Grassly et al. (2006) describe challenges to polio eradication in densely populated Uttar Pradesh and Bihar: "high population density and poor sanitation can lead to more frequent infectious contacts and increase levels of excreted polio-virus in the environment."

[3]Brinkley (1997) and Coelho and McGuire (1997) discuss the influence of hookworm, spread by poor sanitation, on the population health and economic development of the American South. An observational study in rural Wisconsin in the U.S. found that a higher density of septic tanks was associated with an increased prevalence of diarrhea (Borchardt et al., 1979). Studies of the Tama River in Tokyo and the Cumberland River in Nashville showed that fecal bacteria concentrations, possibly originating from sewer overflows, were significantly affected by population density (Ham et al., 2009) and (Young and Thackston, 1999). An aggregated (or "ecological") study across three developed countries also found suggestive evidence that higher population density may be related to increased antibiotic resistance, because higher interpersonal contact can lead to the spread of resistant bacteria (Bruinsma et al., 2003).

establishing the infectious mechanism of the disease.

A large medical and epidemiological literature documents that poor sanitation continues to cause death and disease, particularly among children in developing countries today. 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. Humphrey (2009b) posits that chronic but subclinical "environmental enteropathy," caused by ingestion of fecal pathogens, may also lead to slowed growth. Lin et al. (2013) finds associations among fecal environmental contamination, enteropathy and child height in Bangladesh. Poor sanitation can also spread parasite infections, which, although rarely fatal by themselves, contribute to poor health and poor physical growth (Haque, 2007).[4] Several papers in economics have also identified important effects of sanitation-related diseases on anemia and early-life mortality (*e.g.* Galiani et al., 2005; Cutler and Miller, 2005; Watson, 2006; Coffey and Geruso, 2015), as well as effects on subsequent human capital accumulation (*e.g.* Bleakley, 2007; Baird et al., 2011; Hammer and Spears, 2016; Spears and Lamba, 2015).

Recent econometric studies suggest an interaction between sanitation and population density in predicting health and human capital outcomes across developing countries. As motivation for a study that seeks to explain differences in child height between India and sub-Saharan Africa, Spears (2013) observes that heterogeneity in open defecation density across developing countries can account for a large fraction of international differences in average child height. However, Spears (2013) does not focus on the internal validity of the sanitation-population density interaction. The following analyses are the first to use micro-level data from all available Demographic and Health Surveys and disaggregated fixed effects to quantify and verify the robustness of this interaction.

3 Population density, sanitation and child health in developing countries: Evidence from 172 DHS

In these analyses, we use a dataset of 172 DHS collected between 1990 and 2012 in 69 developing countries to assess whether the importance of dense settlement for infant mortality and child height in developing countries is moderated by exposure to local sanitation behavior. As motivation, we

[4] The relationship between the density of open defecation and child health outcomes likely depends on the kind of parasites that are present in a region, and the conditions under which they are most easily to be spread.

begin with a description of how urban place, sanitation, and population density predict infant mortality. We find that the urban infant survival advantage is importantly diminished after controlling for local sanitation, population density, and their interaction.

We then focus directly on the interaction of population density with local open defecation in predicting height and infant mortality. Although this multi-country analysis is not intended to precisely identify a causal effect, we demonstrate that the effect of population density on the sanitation-health gradient is quantitatively robust to model respecifications, including the introduction of a range of fixed effects and controls, suggesting that the interaction we document is unlikely to reflect omitted variables. To provide additional evidence that this relationship is not due to omitted variables, we conduct falsification tests that demonstrate that other measures of socioeconomic status do not similarly interact with population density to predict infant mortality. Finally, we model the shape of the dependence of the sanitation-mortality gradient, and the sanitation-height gradient on population density.

3.1 Data & summary statistics

These analyses combine data from two sources: population density from census or other aggregate demographic data, and sanitation, health, and other covariate data from DHS collected between 1990 and 2012. DHS are internationally comparable, nationally representative surveys collected in poor and middle income countries.[5] We append all available DHS to make a large dataset where each observation is an individual child. We merge to the child level data a new dataset on population density at the level of DHS sub-national regions (hereafter regions). For each of the over 1,800 regions, we manually matched the region to publicly available, published demographic data for the closest available year to the year of the survey. Appendix table A1 lists all of the countries and years in the international sample as well as the source of the region level data on population density.

Independent variable of interest. Our independent variable of interest is the interaction of the log of population density at the region level with local prevalence of open defecation near a child. We estimate local prevalence of open defecation near a child by estimating the fraction of the households in a child's primary sampling unit (PSU)[6] that defecate in the open rather than using a toilet or

[5]For more information about the use of DHS data, see Rutstein and Rojas (2006) or visit *www.dhsprogram.com*.

[6]DHS use two-stage random sampling. First, a PSU, which is either a rural village or a small set of urban blocks, is selected; second, households within the PSU are randomly selected.

latrine. We do this by computing the fraction of households in each PSU in the sample that report open defecation.[7] This is a local (or community-level) measure of exposure to open defecation, and not merely a property of the child's own household (Montgomery and Hewett, 2005). To isolate and emphasize the negative externality of neighbors' open defecation, we also control for whether a child's own household defecates in the open in all of the regressions we present.

Dependent variables. Our dependent variables are infant mortality and height-for-age. Infant mortality is a child level indicator, which we define for all live births that occurred at least one year before the date of the survey and no more than five years before the date of the survey.[8] Infant mortality is coded as 0 if the child survived her first year of life, and 1,000 if the child died within the first year. This scaling of the indicator by 1,000 makes our infant mortality estimates consistent with published population-level infant mortality rate (IMR) statistics. The second dependent variable is a child's height-for-age z-score.[9] A height-for-age z-score scales a child's height relative to a healthy population of that child's age and sex. We use the 2006 WHO international reference population of healthy children.

Summary statistics. Table 1 presents summary statistics about the international dataset. Panel A shows summary statistics for the dependent, independent, and select control variables in the sample as a whole; panels B and C show summary statistics for children living in below-median and above-median open-defecation PSUs, respectively.

Over 6 percent of children in the data died before their first birthday. The average child in our data is notably shorter than children in the healthy reference population. Children in above-median open-defecation PSUs are about a half of a standard deviation, on average, shorter than children in below-median open-defecation PSUs. About one-third of the average child's neighbors defecate in the open.

Population density varies widely in our sample, with an interquartile range from 31 to 239 people per square kilometer. Children living in above-median open defecation PSUs live in less population-dense regions, on average, than those in below-median open-defecation PSUs. Appendix figure A2

[7]Because the fraction of households in a PSU that defecates in the open is estimated from a sample, and not from data on every household in the neighborhood, this is a noisy measure of the true fraction of households in a child's local area who defecate in the open; this random measurement error will attenuate our coefficients, so any sanitation gradient we uncover may be a lower bound. We further discuss potential consequences of measurement error in appendix A.2.

[8]Figure A1, in the appendix, shows that our results are not sensitive to the choice of five years before the survey as a cutoff for inclusion in the sample.

[9]Following standard practice using these WHO z-scores, we omit any child beyond ± 6.

plots a kernel density estimate of the distribution of population density among children in our international sample. Throughout our analysis, we transform population density to a log scale. A normal distribution with the same mean and standard deviation is included for comparison; population density appears to match a lognormal distribution.

Although it is not used in the regressions (because it would be a country-year fixed effect, which we use as a control), we include GDP per capita from the Penn World Tables in table 1 for illustration. The median child in this dataset is poor; she is growing up in a country-year with a GDP per capita per day of $1.44. Finally, we also present summary statistics for some of the variables that we use in falsification tests and for some of the mother-level controls used in the regressions below. 28% of the children's neighbors have piped water, and 41% of them have electricity. PSU-average piped-water access and electrification are far lower for children living in PSUs with above-median open defecation rates. 61% of children had a mother who ever attended school and the median child's mother was 19 years old when she first gave birth.

3.2 Motivation: Urban place, sanitation and infant mortality

The literature reviewed in sections 1 and 2 suggests that, on average, in developing countries today, dense settlement confers an infant health advantage. But it also suggests that this advantage will be less pronounced where sanitation is poor. In this section, we motivate the analyses that follow by using the international dataset to present results from regressions of the form:

$$\text{mortality}_{ip} = \beta_1 \text{place}_p + \beta_2 \text{open defecation}_p + \beta_3 \text{place}_p \times \text{open defecation}_p + \alpha_c + \varepsilon_{ip}, \qquad (1)$$

where *mortality* for child i living in place p is scaled for infant deaths per 1,000; *open defecation* is open defecation in the child's local area (PSU); α_c is a country fixed effect; and *place* will be implemented either as a dummy for urban residence, as defined by the DHS,[10] as population density of the child's sub-national region, or with both in the same regression. All the variables are demeaned to facilitate comparability of coefficients across columns.

Table 2 presents descriptive regression results that build to the result of equation 1. We begin by estimating the within-country urban infant survival advantage in our data set. Column 1 finds that averaging over the combined data set, children in urban places, as defined by the DHS, are 16 per

[10]The DHS defines urban residence based on the definitions used by countries' national statistical offices.

1,000 more likely to survive their first year of life than children in rural places. Part of this apparently large urban advantage reflects the better sanitation environment in urban areas than rural areas. Column 2 adds local sanitation and shows that controlling for the better sanitation environment in cities diminishes the urban coefficient. However, this model assumes that open defecation has the same association with infant mortality in both urban and rural areas. Column 3 includes the interaction of urban and local sanitation, and finds that the coefficient on urban declines in absolute magnitude by almost two-thirds relative to the magnitude of its coefficient in column 1. Open defecation and urban residence interact, such that open defecation is more steeply associated with mortality in urban rather than rural places. The average urban child is only 2.4 per 1,000 less likely to die in infancy in places where everyone defecates in the open, compared with 8.0 per 1,000 – or more than triple the advantage – in places where nobody defecates in the open.[11]

Why does urban place interact with sanitation to predict infant mortality? We propose that it is the population density of urban places that leads to greater disease externalities. Therefore, columns 4 and 5 replicate the results in columns 1 and 3, this time replacing *urban* with population density. On average, higher population density places have slightly lower infant mortality than lower population density places. We find that open defecation is more steeply associated with mortality in more densely populated places. Population density is not itself associated with either a mortality advantage or a mortality disadvantage at the average level of open defecation.

Finally, the regression results in column 6 present a "horse race" which demonstrates that increased population density is indeed the reason why urban interact with sanitation to predict infant mortality. We include both the interaction of population density and local sanitation, as well as urban residence and local sanitation. Once the interaction of population density and open defecation is accounted for, there is no longer an apparent interaction between urban place and sanitation. Population density *per se* appears neither associated with greater nor lesser mortality, and the urban advantage documented in column 6 is only one-third as much as appeared to be the case in column 1.

Table 2 suggests that in developing countries, an interaction between sanitation and population

[11]Since the variables in the regression are demeaned, we compute the difference between urban and rural where local open defecation is zero by $-2.4 = -6.047 + 5.592 \times \left(1 - \overline{open\ defecation}\right)$, and the difference between urban and rural where local open defecation is 100% by $-8 = -6.047 - 5.592 \times \overline{open\ defecation}$ where $\overline{open\ defecation} = 0.35$, which is the mean of $\overline{open\ defecation}$.

density importantly moderates the relationship between place and early-life health and mortality. The analyses that follow sharpen our understanding of this interaction, and investigate its external and internal validity.

3.3 The interaction of sanitation and population density in 172 DHS

We have seen that the relationship between urban place and health depends importantly on population density and on open defecation. In this section we focus directly on establishing an interaction between population density and sanitation, and assess the robustness of the estimate.

3.3.1 Empirical strategy

For each dependent variable, we regress health on a linear interaction of local sanitation and population density, controlling for household sanitation and one of three levels of fixed effects α:

- *country*: for example, a fixed effect for India, pooling over the 1992, 1998, and 2005 DHS
- *survey*: a partition of *country*, for example, a fixed effect for India in each surveyed year
- *region*: a partition of *survey* into the sub-national region level at which population density is matched, for example, the Indian state of Bihar in 2005

Note that adding fixed effects means our identification is derived from heterogeneity within these regions. Depending on the question we seek to answer, this may be *over*controlling. For example, in the case of the region fixed effects, the difference in population density between regions within the same country may be of policy relevance.

Our regression specification is:

$$\begin{aligned} health_{ipsc} = \quad & \beta_1 local\ OD_{ipsc} \times \ln\left(density_{psc}\right) + \beta_2 \ln\left(density_{psc}\right) + \beta_3 local\ OD_{ipsc} + \\ & \beta_4 household\ OD_{ipsc} + X_{ipsc}\theta + \alpha_{psc} + \varepsilon_{ipsc}, \end{aligned} \tag{2}$$

where i indexes individual children, p is the region for which population density is matched, s indicates a DHS, and c is a country. X is an extensive set of controls which we use throughout the analysis of the international dataset. It includes six indicators for the child's household owning the six common DHS assets (electricity, radio, TV, motorcycle, car, refrigerator); indicators for sex, birth calendar month, and multiple births; year of birth entered linearly; indicators for first, second, or third birth order; an indicator for whether the child's mother attended school; and the mother's age at first birth entered linearly.[12] We also control for whether or not the child's own household defecates in the open. When child height is the dependent variable, we always add a vector of 120 age-in-months by sex indicators. Standard errors are conservatively clustered at the level of 172 DHS (thus, India's entire 2005 DHS is one cluster), except in specifications with country fixed effects, where standard errors are even more conservatively clustered at the country level.[13]

3.3.2 Results

Table 3 reports estimates of equation 2, for infant mortality in panel A and for height-for-age in panel B. It reports results using several combinations of fixed effects and controls. The result is quantitatively robust as the estimates remain in a stable range: a one log-unit increase in population density increases the change in infant mortality associated with moving from no neighbors defecating in the open to all neighbors defecating in the open by about 2 deaths per 1,000 live births, and increases the corresponding decline in height-for-age by about 0.04 of a height-for-age standard deviation.[14]

Results are similar if we use fixed effects for countries (64 for height and 69 for infant mortality) or if we instead use over 1,800 disaggregated fixed effects by region within each survey year, with or without a long vector of controls. Indeed, the regional fixed effects may represent over-controlling if part of what is important for child health in differences across region-years is differences in the density of open defecation across space and time. Two of the twelve coefficient estimates are not statistically significantly different from zero; we include them for completeness and note that their coefficients are of important magnitude and not statistically distinguishable from the other coefficient estimates. Moreover, this lack of statistical significance is only because we have conservatively clustered standard errors at country or country-year levels; if standard errors were clustered by sub-national region or survey PSU (as is common in use of DHS) then both of these coefficients would be highly statistically significant in our very large dataset.

[12] We are constrained to use variables that are available in all of the DHS.

[13] We do not use a constant term with this, or the other fixed effects models in the paper. In such a model, the constant term is absorbed into the set of fixed effects. (See Cameron and Trivedi (2010), p. 231.)

[14] We obtain these estimates simply by taking a visual approximation of the coefficients on the interaction in table 3. These are reported in the first row of panel A, for infant mortality, and panel B, for height-for-age.

3.4 Falsification: Measures of SES do not interact to predict infant mortality

In this section, we conduct falsification tests: we interact open defecation with other "placebo" measures of community socioeconomic status. If the interaction documented in table 3 merely reflects some unobserved spurious correlation between population density and health, rather than an effect of population density on the consequences of open defecation, then we would expect many other measures of community socioeconomic status to similarly apparently interact with population density.[15]

Figure 2 plots t-statistics on $\hat{\beta}_3$ from estimates of regression equation (3) with various community-level socioeconomic status variables substituted in place of sanitation, with and without a vector of controls X, including the household's own open defecation, as described above. Regressions take the following form:

$$\begin{aligned} mortality_{ipsc} = \quad & \beta_1 SES_{ipsc} + \beta_2 \ln\left(density_{psc}\right) + \\ & \beta_3 SES_{ipsc} \times \ln\left(density_{psc}\right) + \beta_4 household\ OD_{ipsc} + X_{ipsc}\theta + \alpha_c + \varepsilon_{ipsc}. \end{aligned} \tag{3}$$

In all cases, the SES variables are community (survey PSU) averages, computed from the household recode, as in our estimated local open defecation variable. So, for example, open defecation is the fraction of households in the PSU which defecate in the open, radio is the fraction of households in the PSU which have a radio, and bottom fifth is the fraction of the PSU whom the DHS asset index sorts into the bottom fifth of their survey round. The one exception is GDP, which is a country-year level variable.

The dotted lines in figure 2 indicate the threshold for statistical significance. Figure 2 shows that only local open defecation robustly statistically significantly interacts with population density to predict infant mortality, with and without controls.[16] This specificity of the sanitation-density

[15] Although we find little evidence that community level measures of SES interact with population density to predict infant mortality, there is evidence that other community level behaviors related to the spread of infectious disease do interact with population density. Table A5 shows the results of regressing infant mortality on interactions of community-level infectious disease behaviors, like the fraction of children with a BCG vaccine, measles vaccine, or the fraction of children sleeping under a bednet, with population density. Although some of these interactions are statistically significant, table A5 shows that the interaction of open defecation and population density is statistically significant even after controlling for these other interactions.

[16] A high fraction of the local area being in the bottom fifth of the country-year's asset index statistically significantly interacts with population density *without* controls, although not *with*; although it is beyond the scope of this paper, it is plausible that there is a special effect on health of poor people living densely close together. Note, however, that this cannot be an omitted variable in our results: our extended controls include indicators for the individual assets used to construct the asset index.

interaction increases our confidence that the result is indeed due to a greater effect of sanitation on height where population density is greater.

3.5 Extension: The shape of the sanitation-population density interaction

For tractability, the regressions in section 3.2 assumed a linear association between population density and the sanitation-health gradient: each log-unit increase in population density was assumed to be associated with the same steepening of the relationship between sanitation and health. However, with such a large dataset, we can model this relationship more flexibly to show the shape of the sanitation-population density interaction.

In this section, we allow the interaction between population density and the health-sanitation gradient to be a fifth-order polynomial. We use an odd-ordered polynomial to capture flexibility in the increasing relationship between population density and the sanitation-infant mortality gradient. We use an 5th order polynomial because of the statistical significance of these terms (F = 5.6; p < 0.01) and the failure of the extra 6th and 7th terms to be jointly significant additions to the model (F = 0.4; p = 0.79).

For both infant mortality and height-for-age, we estimate:

$$\begin{aligned} health_{ipsc} = \; & \beta_1 local\ OD_{ipsc} + \textstyle\sum_{j=1}^{5} \beta_{2,j} \ln \left(density_{psc}\right)^j + \\ & \textstyle\sum_{j=1}^{5} \beta_{3,j} local\ OD_{ipsc} \times \ln \left(density_{psc}\right)^j + \\ & \beta_4 household\ OD_{ipsc} + X_{ipsc}\theta + \alpha_{psc} + \varepsilon_{ipsc}. \end{aligned} \tag{4}$$

As before, we are estimating health outcomes for child i, in region p, in DHS s, and in country c. As described above, we introduce fixed effects α at country, survey, and region levels, in stages. We also include the same vector of extended controls X, as well as the household's own open defecation, as described above.

This functional form implies that the change in health associated with a change from 0% to 100% local open defecation is:

$$\frac{\partial \widehat{health}}{\partial local\ OD} = \hat{\beta}_1 + \sum_{j=1}^{5} \hat{\beta}_{3,j} \ln \left(density_{psc}\right)^i \tag{5}$$

Panel A of figure 1 plots the dependence of the infant mortality-open defecation gradient on population density; panel B plots the height-open defecation gradient as a function of population density. In both cases, the same six specifications that were used in section 3.3's table 3 are plotted: fixed effects at the country, survey, and region level, with and without an extended vector of controls. F-tests with eight degrees of freedom show that the higher-order interaction does not improve the fit – that is, that $\beta_{2,2}$ through $\beta_{2,5}$ and $\beta_{3,2}$ through $\beta_{3,5}$ are all zero – are rejected, for example with $F = 8.50$, $p < 0.0001$ in the case of country fixed effects with no controls.

Figure 1 shows that although adding controls and changing the fineness of fixed effects shifts the estimated function vertically, which changes the *level* of the sanitation-health gradient, the shape of the function – that is, the dependence of the health-open defecation gradient on population density – remains similar. Across model specifications, the association between open defecation and infant mortality, for example, is about twice as steep in places with the average population density of Bangladesh (or in the similarly dense, largely rural Indian states of Uttar Pradesh and Bihar) as it is in places with the average population density of sub-Saharan Africa. Moreover, the function curves convexly, so the effect of population density is even greater at higher levels of population density. Because the average population density in the Bangladesh data used in section 4 is especially high, these estimates predict a particularly steep sanitation-health gradient and large interaction with population density in that context.

4 Population density, sanitation and child height in Bangladesh

Section 3 showed that higher population density is robustly and uniquely associated with a steeper sanitation-health gradient. This section uses variation across time and place in local open defecation within Bangladesh – a country where open defecation has fallen sharply over recent decades – in order to provide further evidence for the internal validity of the sanitation-population density interaction.

Bangladesh is an apt case study to further interrogate the sanitation-population density interaction for two reasons. First, unlike many DHS, the Bangladesh DHS report GIS codes for primary sampling units (PSUs). This permits us to create a more precise measure of the density of open defecation to which an individual child is exposed than we were able to use in the international data, and to control for fixed effects at the district level, which is a much smaller geographic area than the region that was used in the international analysis. Second, Bangladesh experienced a rapid decline in open defecation over the period we study. According to WHO-Unicef statistics, national open defecation declined from 20.6 percent in 1999 to 3.9 percent in 2011 (UNICEF & WHO, 2012). As a result, much of the variation we use to identify the effect of the interaction of sanitation and population density on child height results from a reduction in the density of open defecation over time.

4.1 Data and summary statistics

We combine data from the 1999, 2004, and 2011 Bangladeshi DHS, as well as from two Bangladesh censuses, to investigate the relationship among open defecation, population density, and child height. To do this, we match the PSUs of children in the DHS to political boundaries using GIS codes.

There are three levels of political disaggregation within Bangladesh. Most coarsely, Bangladesh is divided into 7 divisons. Divisions are the sub-national region coded in DHS data; we refer to these as regions for consistency with section 3. Regions are divided into districts, which are not reported in the DHS. There are a total of 64 districts in Bangladesh; the average district has a population of about 2 million people. Districts are divided into sub-districts, which are then divided into Unions (rural), Wards (parts of cities), or Pourashava (towns), which we abbreviate UWP. The average UWP had 339,906 people in the 2011 census.

Each PSU in the Bangladesh DHS is accompanied by a GIS code, which is publicly available on request, which includes the latitude and longitude of the PSU.[17] We used ArcGIS 10 software and a polygon overlay technique to match PSUs from the DHS to districts and UWPs from the 2009 Local Government Engineering Department (LEGD) UWP-level map. After identifying each PSU's UWP, we matched it with a UWP-level population density from census data from Bangladesh Bureau of Statistics (2002) and Bangladesh Bureau of Statistics (2012) to create our independent variable of interest, the interaction of PSU-level open defecation with the log of UWP-level population density. The 1999 and 2004 DHS were matched to the 2001 population census of Bangladesh; the 2011 DHS was matched to the 2011 population census.[18] Thus, each PSU is matched to a highly disaggregated measure of population density. Because the DHS are repeated, nationally representative cross sections that do not form a panel of PSUs, it is often the case that a given UWP is not represented in more than one round of the DHS. Therefore, the smallest geographic unit for which we can include a fixed effect is the district.

[17] We drop two PSUs from the 2004 DHS where GIS information was not reported.

[18] In a small number of cases, area was not available from the census, so we computed population density by dividing census population by area from LEGD data.

Independent variable of interest. Our independent variable of interest is the interaction of the log of UWP-level population density with the fraction of households in a PSU that defecates in the open; this is the same for each child in a given survey round and PSU.

Dependent variable. The dependent variable in this analysis is the height-for-age z-score of children under five, using the WHO 2006 reference of healthy children. For the Bangladesh analysis, we no longer use infant mortality as a dependent variable. With a sample less than four percent as large as in the international analysis of section 3, we are unable to precisely identify effects on infant mortality, a low probability binary outcome, using district fixed effects. Sample size is less of a constraint for continuously-distributed, normalized height-for-age, which is routinely studied in samples of this size (e.g. Spears et al., 2013). The appendix presents evidence that supports an interactive effect of sanitation and population density on infant mortality. Appendix table A2 presents results for infant mortality that use fixed effects for region, rather than district fixed effects, and repeats falsification tests showing that there is no similar interaction with electrification or radio ownership.

Summary statistics. Table 4 reports summary statistics for the Bangladesh dataset. Observations are infants and children, so averages do not, in general, correspond to published summary statistics representative of the population of Bangladesh. For example, if young children are disproportionately found in poorer households, our summary statistics will present a worse picture of human development. Indeed, the summary statistics reflect a poor, mainly rural population with high mortality and low maternal nutrition. However, child height, infant mortality, sanitation, maternal nutrition, and electrification all show clear improvements over the three survey rounds.

4.2 Empirical strategy

We identify the association between local sanitation density and child height from cross-sectional and over-time variation within districts. The GIS matching described above allows us to use fixed effects that are approximately 10 times finer than the seven regional fixed effects used in the international analysis. We estimate regressions with district and survey round fixed effects for children under 5 years old of the following form:

$$\begin{aligned} height_{idt} = \quad & \beta_1 local\ OD_{idt} + \beta_2 \ln\left(density\right)_{idt} + \beta_3 local\ OD_{idt} \times \ln\left(density\right)_{idt} \\ & \beta_4 household\ OD_{idt} + X_{idt}\theta + A_{idt} \times sex_{idt} + year_{idt} + \delta_d + \gamma_t + \varepsilon_{idt}, \end{aligned} \tag{6}$$

where i indexes individual children, d indexes districts, and t indexes survey rounds. Standard errors are clustered at the district level (that is, pooling all survey rounds within a district). As with prior regressions in which height-for-age is the dependent variable, we include 120 age-in-months by sex indicators $A_{idt} \times sex_{idt}$. We also add fixed effects for the year in which a child was born, $year_{idt}$, to account for overall time trends. As before, we control for an indicator for whether the child's own household defecates in the open. δ_d is a district fixed effect and γ_t is a survey round fixed effect. This strategy allows us to control for everything about a child's district, for any potential time trends affecting height, as well as any potential survey round specific measurement issues.

In order to demonstrate the robustness of our result to individual and household regression controls, we add controls, X_{idt}, which are more comprehensive than those included in the international analysis, in stages:

- Birth demography: mother's age at the child's birth as a quadratic polynomial, indicators for multiple birth, indicators for calendar month of birth, and an indicator for being the first born to a mother.
- Household wealth: indicators for the household having electricity, a radio, a television, a bicycle, and a motorcycle or scooter.
- Maternal nutrition, anthropometry, & care: mother's height in centimeters; indicator for mother's literacy; indicator for breastfeeding beginning on the first day.

4.3 Results

Table 5 presents estimates of regression equation 6. We find that local sanitation statistically significantly and robustly interacts with local population density to predict average child height. Adding fixed effects and controls does little to change the magnitude of the coefficient on the interaction; none of the six estimates is statistically distinguishable from the others. These coefficients suggest that a doubling of population density is approximately associated with a 0.2 height-for-age standard deviation increase in the difference in average child height between places where there is no open defecation and where there is 100% open defecation.[19] The stability of the coefficient on the interaction suggests that it is unlikely to be driven by an omitted variable uncorrelated with all of these

[19] Because population density is in log, the coefficient estimate can be interpreted by saying that doubling the population density is associated with a $ln(2) \times \hat{\beta}$ difference in the dependent variable. In this case, $ln(2) \times 0.3 \approx 0.2$.

controls.

The average linear interaction in table 5 for Bangladesh is approximately 10 times the size of the international average linear interaction in table 3. This best linear approximation to the interaction is useful because it allows our fixed effects identification strategy and permits simple statistical significance tests with controls. However, figure 1 suggests that – over the entire global range of variation in population density – the interaction is not linear. Instead, the dependence of the health-sanitation gradient on population density appears to be steeper at greater population densities.

In international comparison, average population density in Bangladesh is very high. Bangladeshi children, therefore, would be on the far right side of panel B of figure 1, which predicts a particularly steep linearized interaction between population density and open defecation. Indeed, when we use the six models estimated in panel B of figure 1 to compute the relevant linear interaction gradients at the average population density for children the Bangladeshi sample, we find that the numerical predictions for the coefficient on the interaction range from -0.143 to -0.309. These magnitudes are larger than the global average linear interactions presented in table 3, which range from -0.023 to -0.074, and similar to what our fixed effects identification strategy finds within Bangladesh in table 5, which are coefficients that range form -0.261 to -0.455.

5 Discussion and conclusion

Our paper was motivated by the observation that an interaction between sanitation and population density importantly moderates the relationship between place and early-life health outcomes. The results presented in this paper sharpen our understanding of this interaction, and investigated its external and internal validity. In two separate analyses – representing two different points in a trade-off between external validity and internal validity – we find that poor sanitation is more detrimental for early life health where population density is greater, or stated differently, that population density does not have the same benefits for health where sanitation is poor. These results are biologically plausible because open defecation leads to environmental contamination with germs from feces, and these germs are more likely to cause disease where people are more likely to come in contact with them.

Although resolving longstanding debates about the health advantages or penalties of living in urban or densely-populated areas is well beyond the scope of this paper, our results suggest some clarifications about the importance of place for child health in developing countries. We have isolated that high population density and poor sanitation *in combination* are particularly threatening to early life health. Our results suggest that high density *without* poor sanitation is substantially less dangerous, such that the advantages of access to health care and other resources might dominate the disadvantages of disease externalities, yielding a net health benefit of living in dense cities (Leon, 2008). Additionally, urban settings with low population density may not be disadvantaged relative to rural settings with high population density.

Our result has an important implication for policy-makers: for a given level of open defection, concentrate attention on improving sanitation where population density is high, or at minimum include population density as a factor in allocation decisions. To emphasize, this does not exclusively or even necessarily recommend that sanitation policy attend to urban places. Population density is a continuous variable, and many parts of the developing world that are classified as rural have higher population densities than places classified as urban. The latest estimates of open defecation and population density in the developing world suggest an increasing concentration of open defecation in densely populated parts of rural India, which poses a significant threat to the health of children in these regions, despite their "rural" classification.[20]

[20]According to the Unicef-WHO statistics, open defecation is increasingly a South Asian, and particularly an Indian, problem. Although Bangladesh has drastically reduced open defecation, and Pakistan has seen marked improvements, there continues to be more open defecation in India than there is toilet or latrine use. 60% of open defecation in the world occurs in India, a country where 70% of households live in rural areas, and 70% of rural households defecate in the open. Primarily rural Indian states like Uttar Pradesh and Bihar, home to about 300 million people, have population densities that exceed 800 persons per square kilometer.

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.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. 2011. "Worms at Work: Long-run Impacts of Child Health Gains." working paper, Berkeley.

Bangladesh Bureau of Statistics. 2002. *Statistical Year Book of Bangladesh: 2001*: Bangladesh Bureau of Statistics, Planning Division, Ministry of Planning Government of Peoples Republic of Bangladesh.

——— 2012. *Statistical Year Book of Bangladesh: 2011*: Bangladesh Bureau of Statistics, Planning Division, Ministry of Planning Government of Peoples Republic of Bangladesh.

Bleakley, Hoyt. 2007. "Disease and Development: Evidence from Hookwork Eradication in the American South." *Quarterly Journal of Economics*, 122(1): 73–117.

Bocquier, Philippe, Nyovani Janet Madise, and Eliya Msiyaphazi Zulu. 2011. "Is there an urban advantage in child survival in sub-Saharan Africa? Evidence from 18 countries in the 1990s." *Demography*, 48(2): 531–558.

Borchardt, Mark A., Po-Huang Chyou, Edna O. DeVries, , and Edward A. Belongia. 1979. "Septic System Density and Infectious Diarrhea in a Defined Population of Children." *Environmental Health Perspectives*, 111(5): 742–748.

Brinkley, Garland L. 1997. "The decline in Southern agricultural output, 1860-1880." *Journal of Economic History*, 57: 116–138.

Bruinsma, N., J.M. Hutchinson, A.E. van den Bogaard, H. Giamarellou, J. Degener, and E. E Stobberingh. 2003. "Influence of population density on antibiotic resistance." *Journal of Antimicrobial Chemotherapy*, 51: 385–390.

Cameron, Adrian Colin and Pravin K Trivedi. 2010. "Microeconometrics using stata."

Cameron, Lisa, Manisha Shah, and Susa Olivia. 2013. "Impact evaluation of a large-scale rural sanitation project in Indonesia." Policy Research Workng Paper 6360, World Bank.

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.

Coelho, Philip R and Robert A McGuire. 1997. "African and European bound labor in the British new world: The biological consequences of economic choices." *Journal of Economic History*, 57: 83–115.

Coffey, Diane and Michael Geruso. 2015. "Sanitation, Disease, and Anemia: Evidence From Nepal." *r.i.c.e. working paper*.

Coffey, Diane, Aashish Gupta, Payal Hathi, Nidhi Khurana, Dean Spears, Nikhil Srivastav, and Sangita Vyas. 2014. "Revealed preference for open defecation." *Economic & Political Weekly*, 49(38): 43.

Corner, Robert J., Ashraf M. Dewan, and Masahiro Hashizume. 2013. "Modelling typhoid risk in Dhaka Metropolitan Area of Bangladesh: the role of socio-economic and environmental factors." *International Journal of Health Geographics*, 12(13): 1–15.

Cummins, Joseph R. 2013. "On the Use and Misuse of Child Height-for-Age Z-score in the Demographic and Health Surveys."Technical Report 201417.

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

Dorélien, Audrey, Deborah Balk, and Megan Todd. 2013. "What Is Urban? Comparing a Satellite View with the Demographic and Health Surveys." *Population and Development Review*, 39(3): 413–439.

Dye, Christoper. 2008. "Health and Urban Living." *Science*, 318: 766–769.

Eberhardt, Mark S et al.. 2001. *Health, United States, 2001: urban and rural health chartbook*: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Entwisle, Barbara. 2007. "Putting people into place." *Demography*, 44(4): 687–703.

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): 609Ű621.

Fink, Günther, Isabel Günther, and Kenneth Hill. 2011. "The effect of water and sanitation on child health: evidence from the demographic and health surveys 1986 - 2007." *International Journal of Epidemiology*, 40(5): 1196–1204.

——— 2014. "Slum Residence and Child Health in Developing Countries." *Demography*, forthcoming.

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

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

Grassly, Nicholas C., Christophe Fraser, Jay Wenger, Jagadish M. Deshpande, Roland W. Sutter, David L. Heymann, and R. Bruce Aylward. 2006. "New Strategies for the Elimination of Polio from India." *Science*, 314(5802): 1150–1153.

Günther, Isabel and Kenneth Harttgen. 2012. "Deadly Cities? Spatial Inequalities in Mortality in sub-Saharan Africa." *Population and Development Review*, 38(3): 469–486.

Ham, Young-Sik, Hiromi Kobori, and Masahisa Takasago. 2009. "Effects of combined sewer overflow and stormwater on indicator bacteria concentrations in the Tama River due to the high population density of Tokyo Metropolitan area." *Environmental Monitoring and Assessment*, 152: 459–468.

Hammer, Jeffrey and Dean Spears. 2016. "Village sanitation and child health: Effects and external validity in a randomized field experiment in rural India." *Journal of Health Economics*, 48: 135–148.

Haque, Rashidul. 2007. "Human intestinal parasites." *Journal of health, population, and nutrition*, 25(4): 387.

Hartley, David. 2004. "Rural health disparities, population health, and rural culture." *American Journal of Public Health*, 94(10): 1675.

Hugo, Graeme, Anthony Champion, and Alfredo Lattes. 2003. "Toward a new conceptualization of settlements for demography." *Population and Development Review*, 29(2): 277–297.

Humphrey, Jean H. 2009a. "Child undernutrition, tropical enteropathy, toilets, and handwashing." *The Lancet*, 374(9694): 1032–1035.

——— 2009b. "Child undernutrition, tropical enteropathy, toilets, and handwashing." *The Lancet*, 374: 1032 – 35.

Jaccard, James and Choi K Wan. 1995. "Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches." *Psychological bulletin*, 117(2): 348.

Jankowska, Marta M, Magdalena Benza, and John R Weeks. 2013. "Estimating spatial inequalities of urban child mortality." *Demographic research*, 28: 33.

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

Leon, David A. 2008. "Cities, urbanization, and health." *International Journal of Epidemiology*, 37(1): 4–8.

Lin, Audrie, Benjamin F Arnold, Sadia Afreen, Rie Goto, Tarique Mohammad Nurul Huda, Rashidul Haque, Rubhana Raqib, Leanne Unicomb, Tahmeed Ahmed, John M Colford et al.. 2013. "Household environmental conditions are associated with enteropathy and impaired growth in rural Bangladesh." *The American journal of tropical medicine and hygiene*: 12–0629.

Magadi, Monica Akinyi, Eliya Msiyaphazi Zulu, and Martin Brockerhoff. 2003. "The inequality of maternal health care in urban sub-Saharan Africa in the 1990s." *Population studies*, 57(3): 347–366.

Matthews, Zoe, Amos Channon, Sarah Neal, David Osrin, Nyovani Madise, and William Stones. 2010. "Examining the Şurban advantageŤ in maternal health care in developing countries." *PLoS medicine*, 7(9): e1000327.

McGuire, Robert Allen and Philip RP Coelho. 2011. *Parasites, pathogens, and progress: diseases and economic development*: MIT Press Cambridge, MA.

Montgomery, Mark R. and Paul C Hewett. 2005. "Urban poverty and health in developing countries: Household and neighborhood Effects." *Demography*, 42(3): 397–425.

Mosley, W Henry and Lincoln C Chen. 1984. "An analytical framework for the study of child survival in developing countries." *Population and development review*: 25–45.

Preston, Samuel H. 1975. "The Changing Relation between Mortality and Level of Economic Development." *Population Studies*, 29(2): 231–248.

Preston, Samuel H and Michael R Haines. 1991. "Fatal years: child mortality in late nineteenth-century America.."

Root, Graham. 1997. "Population density and spatial differentials in child mortality in Zimbabwe." *Social Science & Medicine*, 44(3): 413–421.

Rutstein, Shea Oscar and Guillermo Rojas. 2006. "Guide to DHS statistics." *Calverton, Maryland: ORC Macro*.

Sastry, Narayan. 1996. "Community characteristics, individual and household attributes, and child survival in Brazil." *Demography*, 33(2): 211–229.

——— 1997. "What explains rural-urban differentials in child mortality in Brazil?" *Social Science & Medicine*, 44(7): 989–1002.

Sastry, Narayan and Jon M Hussey. 2003. "An investigation of racial and ethnic disparities in birth weight in Chicago neighborhoods." *Demography*, 40(4): 701–725.

Smith, Lisa C, Marie T Ruel, and Aida Ndiaye. 2005. "Why is child malnutrition lower in urban than in rural areas? Evidence from 36 developing countries." *World Development*, 33(8): 1285–1305.

Spears, Dean. 2013. "How much international variation in child height can sanitation explain?" Policy Research Working Paper 6351, World Bank.

Spears, Dean and Sneha Lamba. 2015. "Effects of early-life exposure to sanitation on childhood cognitive skills: Evidence from India's Total Sanitation Campaign." *Journal of Human Resources*.

Spears, Dean, Arabinda Ghosh, and Oliver Cumming. 2013. "Open defecation and childhood stunting in India: an ecological analysis of new data from 112 districts." *PLoS One*, 8(9): e73784.

UNICEF & WHO. 2012. "Progress on drinking water and sanitation: 2012 update."

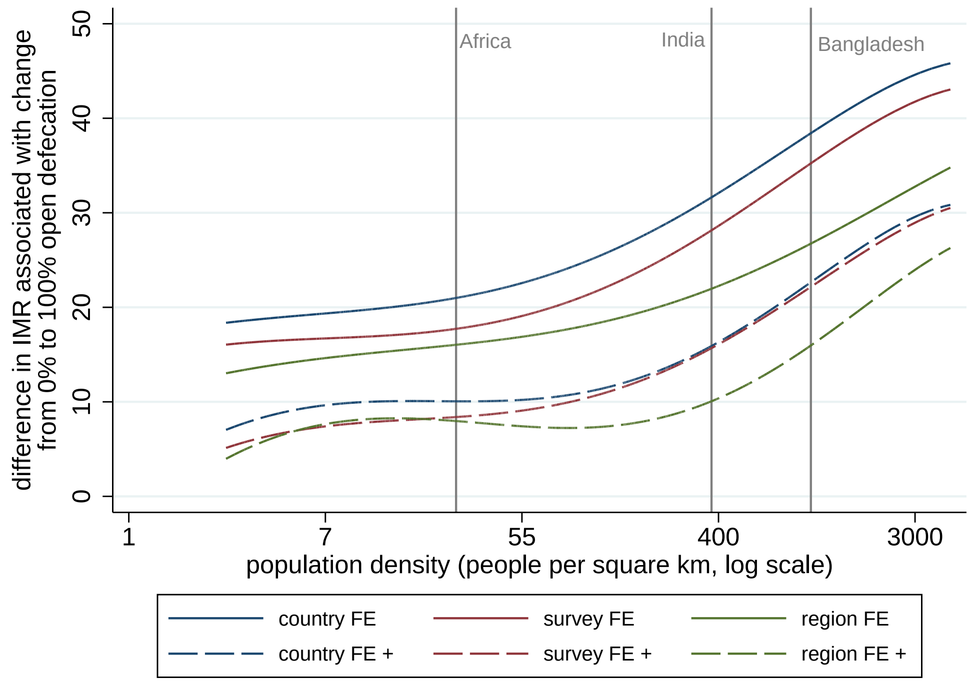
Van de Poel, Ellen, Owen O'Donnell, and Eddy Van Doorslaer. 2007. "Are urban children really healthier? Evidence from 47 developing countries." *Social Science & Medicine*, 65(10): 1986–2003.

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

Woods, Robert. 2004. "Urban-Rural Mortality Differentials: An Unresolved Debate." *Population and Development Review*, 29(1): 29–46.

Young, Katherine D. and Edward L. Thackston. 1999. "Housing Density and Bacterial Loading in Urban Streams." *Journal of Environmental Engineering*, 125: 1177–1180.

Panel A: Association between local open defecation and IMR



Panel B: Association between local open defecation and child height

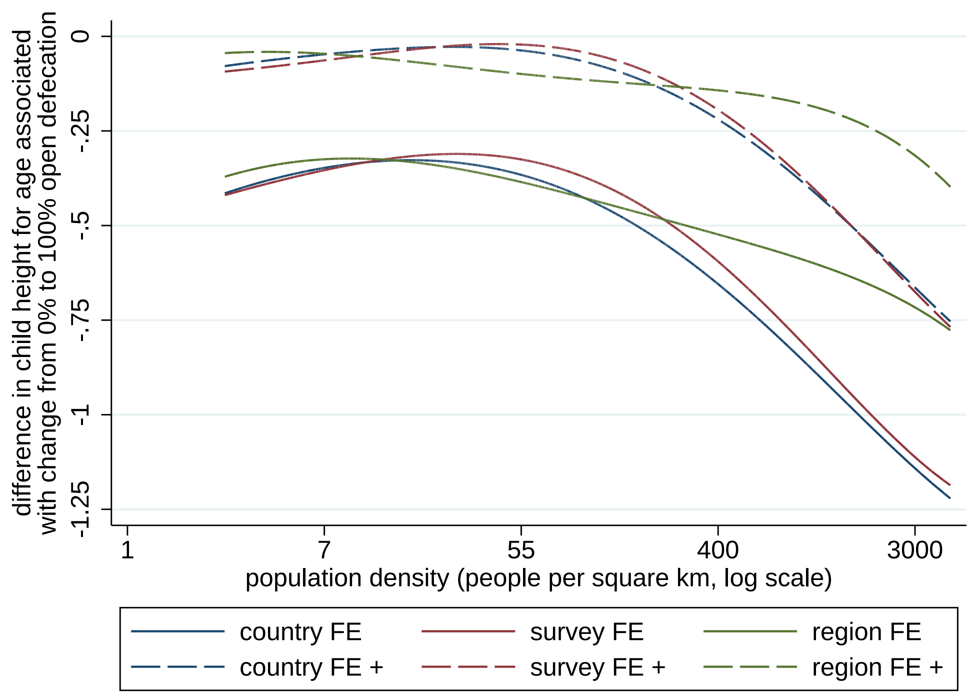


Figure 1: Dependence of sanitation gradient on population density, international sample

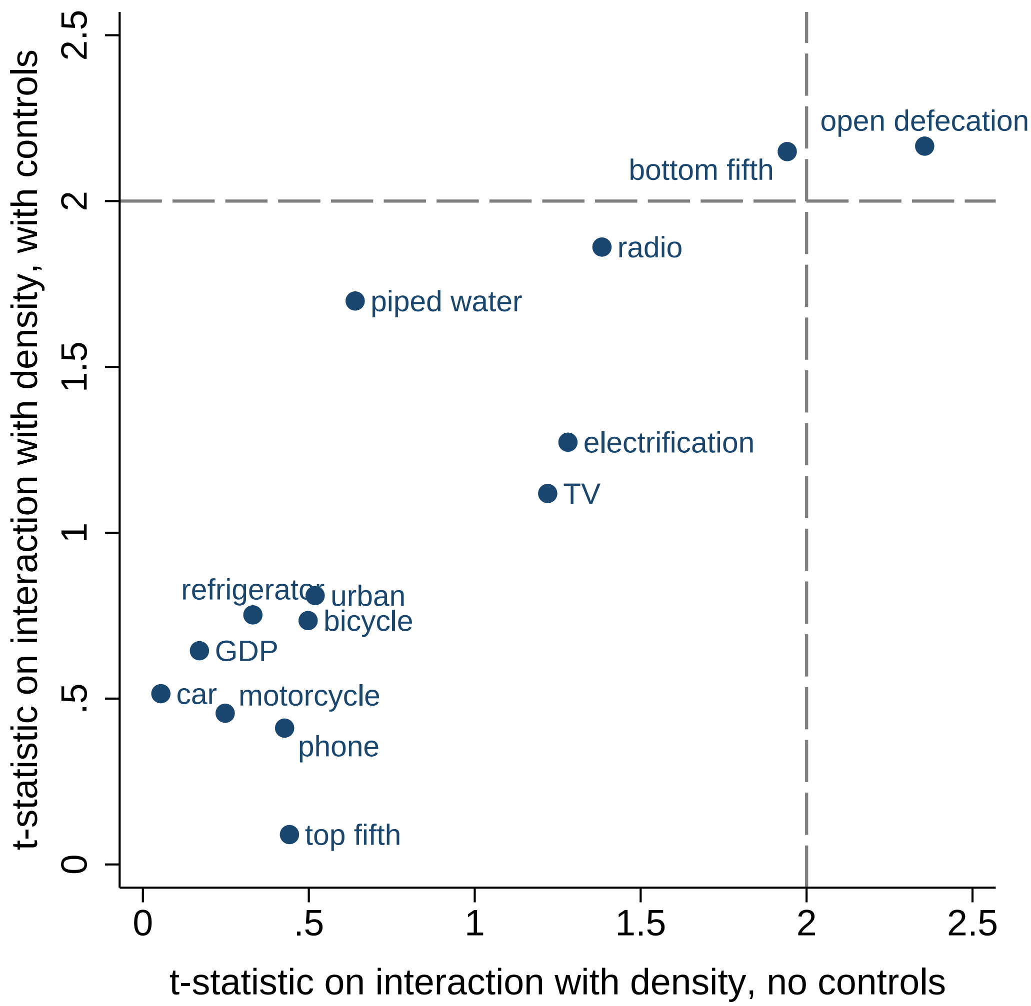


Figure 2: Among community-level SES measures, only open defecation interacts with population density to predict infant mortality, international sample

Table 1: Summary statistics, international sample

	mean	25th percentile	median	75th percentile
Panel A: Full sample				
infant mortality rate	62.24			
height-for-age	-1.49	-2.59	-1.53	-0.47
local open defecation	0.35	0.00	0.14	0.72
household open defecation	0.35	0	0	1
population density per km^2	443	31	81	239
ln(density)	4.48	3.43	4.39	5.47
GDP per capita (USD)	1,079	324	525	1,249
local piped water	0.28	0	0	0.57
local electrification	0.41	0	0.22	0.92
urban	0.33	0	0	1
mother ever attended school	0.61	0	1	1
mother's age at first birth	19	17	91	21
mother's height (cm)	130	126	130	134
n (IMR: live births)	1,112,465			
n (height: children under 5)	858,514			
Panel B: Below-median local open defecation				
infant mortality rate	50.98			
height-for-age	-1.31	-2.32	-1.28	-0.28
local open defecation	0.03	0	0	0.06
household open defecation	0.03	0	0	0
population desity per km^2	677	39	91	308
ln(density)	4.71	3.66	4.51	5.73
GDP per capita (USD)	1,379	360	771	1,718
local piped water	0.59	0.04	0.81	1.00
local electrification	0.41	0.00	0.25	0.88
urban	0.53	0	1	1
mother ever attended school	0.76	1	1	1
mother's age at first birth	20	17	19	22
mother's height (cm)	130	127	130	134
Panel C: Above-median local open defecation				
infant mortality rate	73.39			
height-for-age	-1.85	-2.95	-1.87	-0.78
local open defecation	0.67	0.40	0.72	0.95
household open defecation	0.66	0	1	1
population density per km^2	211	26	72	203
ln(density)	4.25	3.26	4.28	5.31
GDP per capita (USD)	780	324	441	783
local piped water	0.24	0	0	0.46
local electrification	0.14	0	0	0.12
urban	0.13	0	0	0
mother ever attended school	0.46	0	0	1
mother's age at first birth	19	17	18	21
mother's height (cm)	130	125	130	134

Observations are individual children born alive in the 10 years before the survey. Children are included in the summary statistics sample if they are in either the IMR or the height sample.

Table 2: Urban residence, population density, sanitation, and mortality, international sample

	(1)	(2)	(3)	(4)	(5)	(6)
	infant mortality, deaths per 1,000					
urban	-16.06***	-7.050***	-6.047***			-5.751**
	(1.502)	(1.614)	(1.753)			(1.760)
local open defecation		27.22***	28.28***		32.38***	28.71***
		(3.945)	(4.119)		(2.924)	(3.742)
urban $\times$ local open defecation			5.592†			4.472
			(3.256)			(3.288)
ln(density)				-2.121***	-0.331	0.0357
				(0.578)	(0.645)	(0.626)
ln(density) $\times$ local open defecation					3.321*	2.929*
					(1.381)	(1.366)
n (live births)	1,112,465	1,112,465	1,112,465	1,112,465	1,112,465	1,112,465

Standard errors are clustered by 172 DHS. All regressions include a country fixed effect. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Interacted variables are demeaned to preserve interpretation across columns.

Table 3: Local open defecation robustly linearly interacts with population density, international sample

Panel A: infant mortality is the dependent variable						
	(1)	(2)	(3)	(4)	(5)	(6)
fixed effects:	country	country	survey	survey	region	region
local open defecation	3.273*	2.271*	3.523**	2.772*	2.266*	1.581
$\times ln$ (density)	(1.390)	(1.049)	(1.178)	(1.077)	(1.060)	(1.071)
local open defecation	26.27***	12.61***	22.99***	11.71***	18.80***	8.715***
	(2.339)	(2.244)	(1.978)	(2.186)	(1.794)	(2.166)
ln (density)	-0.330	0.518	-0.316	0.390		
	(0.646)	(0.519)	(0.518)	(0.495)		
household OD	6.246***	3.102**	6.141***	3.455***	6.276***	3.808***
	(1.711)	(1.049)	(1.309)	(1.015)	(1.278)	(1.021)
urban		-1.709		-2.252		-2.222†
		(2.051)		(1.446)		(1.152)
extended controls		✓		✓		✓
n (live births)	1,109,116	942,350	1,109,116	942,350	1,109,116	942,350

Panel B: Child height-for-age is the dependent variable						
	(1)	(2)	(3)	(4)	(5)	(6)
fixed effects:	country	country	survey	survey	region	region
local open defecation	-0.0744*	-0.0445	-0.0677**	-0.0396*	-0.0394**	-0.0229†
$\times ln$ (density)	(0.0335)	(0.0275)	(0.0218)	(0.0192)	(0.0146)	(0.0116)
local open defecation	-0.493***	-0.115*	-0.457***	-0.102**	-0.437***	-0.114***
	(0.0465)	(0.0490)	(0.0325)	(0.0329)	(0.0236)	(0.0208)
ln (density)	0.0259†	-0.00212	0.0257**	-0.00168		
	(0.0150)	(0.0133)	(0.00957)	(0.00916)		
household OD	-0.183***	-0.0676***	-0.183***	-0.0718***	-0.185***	-0.0835***
	(0.0241)	(0.00840)	(0.0143)	(0.00664)	(0.0140)	(0.00657)
urban		0.135***		0.136***		0.122***
		(0.0360)		(0.0242)		(0.0191)
extended controls		✓		✓		✓
age-in-months$\times$sex	✓	✓	✓	✓	✓	✓
n (children under 5)	856,165	701,573	856,165	701,573	856,165	701,573

Standard errors are clustered by country in columns 1 and 2 and by DHS in columns 3 through 6. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Extended controls include six indicators for the child's household owning the six common DHS assets (electricity, radio, TV, motorcycle, car, refrigerator); indicators for sex, birth calendar month, and multiple births; year of birth entered linearly; indicators for first, second, or third birth order; an indicator for whether the child's mother attended school; and the mother's age at first birth entered linearly.

Table 4: Summary statistics, Bangladesh sample

	year		
	1999	2004	2011
height-for-age	-1.95	-1.92	-1.62
IMR	81.57	72.33	50.41
household open defecation	0.199	0.141	0.128
local open defecation	0.201	0.138	0.132
population density per km^2	4,983	4,344	4,466
ln (density)	7.23	7.17	7.29
mother's height (cm)	150	150	151
mother's age	22.72	22.59	22.43
local radio	0.33	0.32	0.08
local electricity	0.36	0.42	0.60
urban	0.27	0.31	0.31
n (height-for-age)	5,435	5,978	7,743
n (infant mortality)	12,517	12,817	16,902

Observations are individual children born alive. Children are included in the summary statistics sample if they are in either the IMR or the height sample.

Table 5: Open defecation interacts with population density to predict height, Bangladesh sample

	(1)	(2)	(3)	(4)	(5)	(6)
	height-for-age z-score					
local open defecation	-0.372*	-0.455**	-0.332*	-0.324*	-0.261†	-0.278*
× ln(density)	(0.176)	(0.152)	(0.163)	(0.149)	(0.139)	(0.137)
local open defecation	-0.654***	-0.768***	-0.624***	-0.590***	-0.364**	-0.331**
	(0.122)	(0.122)	(0.130)	(0.123)	(0.118)	(0.122)
ln(density)	0.045†	0.048†	0.055*	0.047*	-0.007	-0.003
	(0.023)	(0.026)	(0.024)	(0.023)	(0.018)	(0.019)
household open defecation	-0.227***	-0.223***	-0.214***	-0.193***	-0.079†	-0.034
	(0.043)	(0.042)	(0.041)	(0.041)	(0.041)	(0.040)
mother's height (cm)						0.038***
						(0.004)
age in months × sex	✓	✓	✓	✓	✓	✓
district FEs		✓	✓	✓	✓	✓
round & year of birth FEs			✓	✓	✓	✓
birth demography				✓	✓	✓
household wealth					✓	✓
maternal nutrition & care						✓
n (children under 5)	19,156	19,156	19,156	19,156	19,061	19,014

Standard errors clustered by 66 districts in parentheses. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$. For a complete list of control variables please see the text.

APPENDIX

A.1 Explanations of figures and tables

Figure A1 shows that our results are not sensitive to our choice of cutoff for inclusion in the international sample used in table 3. It plots coefficients on the interaction of population density and open defecation from column 1 of table 3, allowing the cutoff for inclusion in the sample to range from 5 to 15 years before the survey. The point estimate is essentially numerically unchanged. The confidence intervals are wider when births further back in time are included. This is consistent with greater measurement error in the independent variable for children born more years before the survey: children who were born long before the DHS experienced somewhat different early life conditions than what is measured by the DHS. However, the stability of the main result means that there is no reason to believe that our choice of cutoff biases the estimate of the interaction of population density and open defecation.

Figure A2 shows that the distribution of the log of population densities in our international sample, used for the analysis in section 3.3, follows approximately a normal distribution.

In order to verify the appropriateness of using the log of population density in our main regression results, rather than some other transformation of the data, figure A3 presents results of 11 estimates using Box-Cox transformations. We transform population density with Box-Cox parameter lambda at intervals of 0.1 from 0 (which is equivalent to our preferred log specification) to 1 (which is equivalent to population density being entered linearly). Because log likelihood is maximized between 0 and 0.2, and is lowest at the linear specification 1, we conclude from this figure that our log transformation is the most useful combination of model fit, theoretical appropriateness, and ease of interpretation. However, we further plot *t*-statistics on the interaction for the convenience of the reader, and note that our main result is qualitatively robust to this entire range of respecifications, with the interaction statistically significant at the 0.10 level in all cases, including intermediate values of lambda with no principled motivation, and at the 0.05 level in 6 of 11 cases. Therefore, this robustness check can increase confidence that our results are not spuriously due to any fragile log modeling choice.

Table A1 lists the data sources for the population density variable used in the analysis in section 3.3.

Table A2 uses the Bangladesh data set described in section 4.1 to show results of regressions of infant mortality on the interaction of population density and open defecation. These results complement the analysis shown in 4 of the text. Column 1 shows a regression of whether or not an infant died on PSU level open defecation, population density in his/her local area, and the interaction of these two variables. Column 2 adds fixed effects for region and finds a similar, positive, interaction of open defecation and population to what was reported in column 1. Columns 3 and 4 present falsification tests that find that the interaction of community-level electrification and population density does not similarly predict IMR, nor does the interaction of community-level radio ownership and population density.

Table A3 shows the results of a mechanism check that supports our main results. In table A3, we regress child weight-for-height (computed using the World Health Organization, 2006 norms), a measure of exposure to disease in the short and medium run, on the interaction of population density and open defecation. The coefficient is negative and numerically similar across specifications; it is more precisely estimated when we control for height (column 2). Just as it is important to control for age when using height-for-age as a dependent variable (Cummins, 2013), it is important to control for height when using weight-for-height as a dependent variable. This ensures that our estimation strategy is robust to any possible functional form of the mean height-weight relationship in the sample population that may differ from the WHO norms. Column 3 adds additional controls for the child's age, sex, and their interaction.

Table A4 shows the results of a robustness check that controls for the age of the child's mother at the time of the child's birth. Column 1 of table A4 reproduces the results of column 1 of table 3 in the main text. Column 2 additionally controls for the mother's age at the time of the child's birth. The coefficient on the interaction of population density and open defecation is of a similar magnitude, and it is statistically significant. Column 3 reproduces the results of column 2 of table 3, and column 4 adds a control for mother's age at the time of the child's birth. Again, the coefficient on the interaction of population density and open defecation is of a similar magnitude, and it is statistically significant.

Table A5 shows the results of a robustness check that adds interactions of population density and community-level infectious disease behaviors other than open defecation to regressions of the form presented in column 1 of table 3 in the main text. This table shows that community-level infectious disease behaviors, like the fraction of children with a BCG vaccine (column 2), or a measles vaccine (column 3), do interact with population density to predict infant mortality. This is not surprising: we would expect infectious diseases like BCG and measles to spread more easily in places where unvaccinated children live closer together than they do in places where children live farther apart. There is also suggestive, though not statistically significant, evidence that whether or not children under five sleep under a bed net (column 4) may also interact with population density to predict infant mortality. However, our main result, that the fraction of the PSU practicing open defecation interacts with population density to predict infant mortality, is nevertheless robust to controlling for the interaction of each of these measures of community-level infectious disease behavior with population density. The interaction of population density and open defecation is not merely proxying for the interaction of population density and other community-level infectious disease behaviors.

A.2 Measurement error in the independent variables

This section of the appendix considers the implications of measurement error for our finding that population density interacts with local open defecation to predict child health outcomes. One source of measurement error is that our data on open defecation come from a sample survey, rather than a census, so local open defecation to which a child is exposed is measured with error. Because the DHS chose a random sample of households in each PSU, this source of measurement error is classical measurement error. Classical measurement error leads to attenuation bias; that is, it biases coefficients toward zero. Therefore, measurement error in the open defecation variable would bias *against* finding a result.

The population density data present a second source of measurement error. We would like to be able to measure population density in a child's local area in the year of her birth, but our data on population density in the international sample are at the region level, and are from a single year, rather than the year of the child's birth. We do not think that using population density data from a year other than that of the child's birth is biasing in favor of our results because we see no systematic change in our results if births further back before the survey are included in the sample, other than increasing noise in the coefficient estimate (see figure A1). This is consistent with increasing measurement error, but inconsistent with bias. Nor do we believe that using data from the region, rather than the local level, in our international sample, would bias in favor of our result. Some evidence for this is that we find quantitatively similar robust estimates in Bangladesh as we do in our international sample; in the Bangladeshi case our population densities are measured with considerably less error because the geographic units are much smaller than DHS sub-national regions in our international analysis.

Finally, because our coefficient of interest is on an interaction, it is also useful to ask: what is the effect of measurement error in an independent variable on the estimated coefficient of an interaction? Jaccard and Wan (1995) have shown that the familiar attenuation bias effect of measurement error is extended to the case of interactions: we would expect the coefficient on the interaction to be attenuated towards zero.

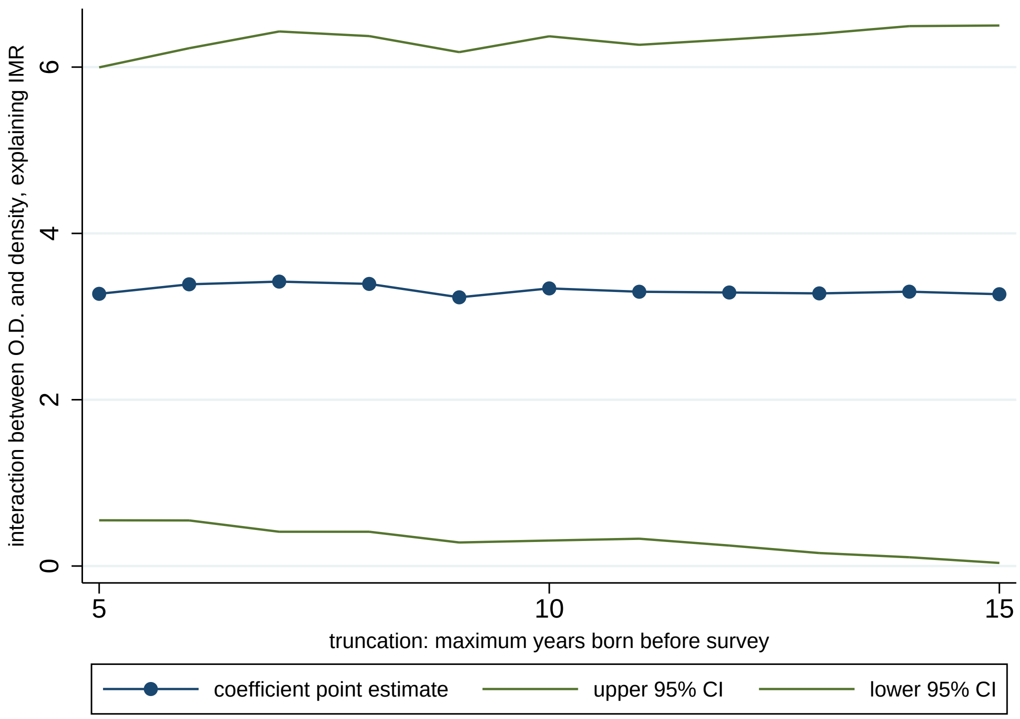


Figure A1: Coefficient on the population density-open defecation interaction does not depend on sample truncation

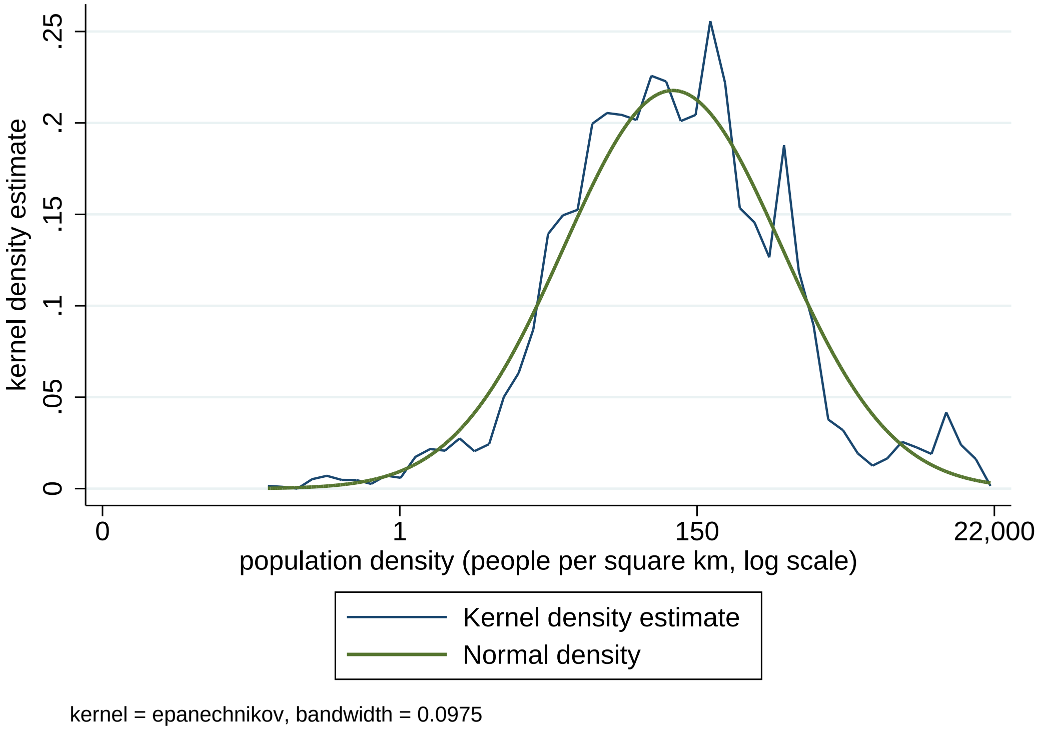


Figure A2: Distribution of population densities, international sample

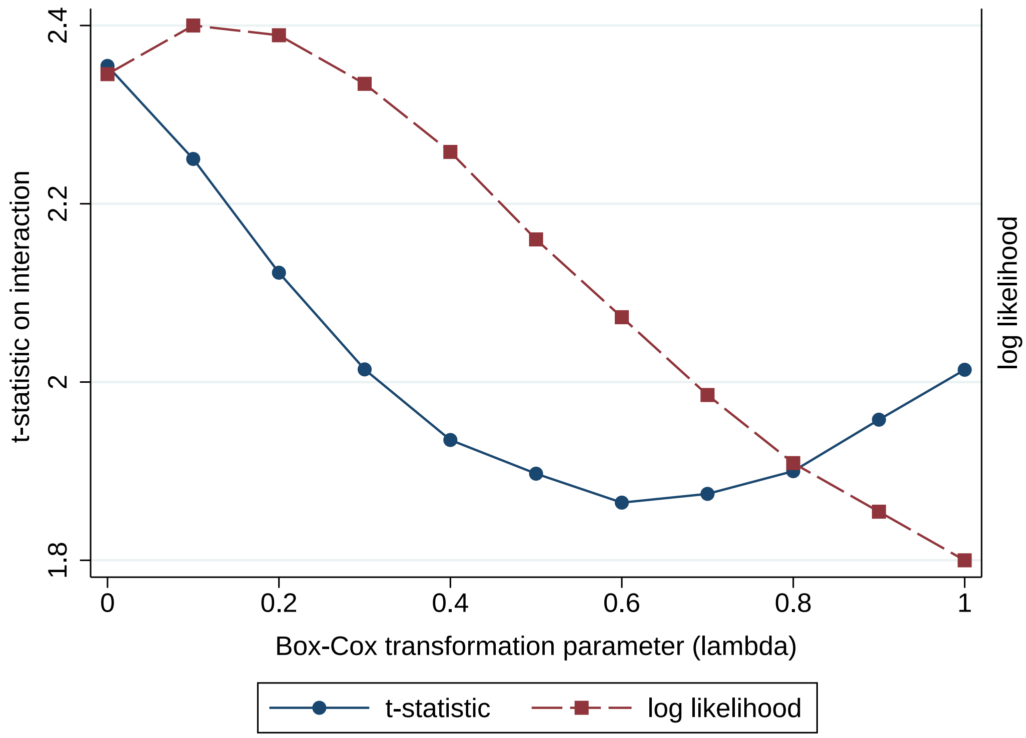


Figure A3: Box-cox transformation parameters and associated t-statistics for regression in column 1 of table 3

Table A1: International sample of 172 DHS

country	year	online source	ultimate source
Albania	2008	wikipedia	Institute of Statistics of Albania. 2011.
Armenia	2000	geohive	National Statistical Service. (2001 & 2011)
Armenia	2005	geohive	National Statistical Service. (2001 & 2011)
Armenia	2010	geohive	National Statistical Service. (2001 & 2011)
Azerbaijan	2006	geohive	State Statistical Committee.
Bangladesh	1993	wikipedia	Bureau of Statistics, Population Census Wing.
Bangladesh	1996	wikipedia	Bureau of Statistics, Population Census Wing.
Bangladesh	1999	wikipedia	Bureau of Statistics, Population Census Wing.
Bangladesh	2004	wikipedia	Bureau of Statistics, Population Census Wing.
Bangladesh	2007	wikipedia	Bureau of Statistics, Population Census Wing.
Bangladesh	2011	wikipedia	Bureau of Statistics, Population Census Wing.
Benin	1996	statoids	Troisième Recensement General de la Population et de l'Habitation.
Benin	2001	statoids	Troisième Recensement General de la Population et de l'Habitation.
Benin	2006	statoids	Troisième Recensement General de la Population et de l'Habitation.
Bolivia	1998	statoids	Instituto Nacional de Estadística, Table of department populations.
Bolivia	2003	statoids	Instituto Nacional de Estadística, Table of department populations.
Brazil	1991	geohive	IBGE , Brazil.
Brazil	1996	geohive	IBGE , Brazil.
Burkina Faso	1993	wikipedia	National Census. (2006)
Burkina Faso	1998	wikipedia	National Census. (2006)
Burkina Faso	2003	geohive	Institut National de la Statistique et de la Demographie.
Burkina Faso	2010	geohive	Institut National de la Statistique et de la Demographie.
Burundi	2010	geohive	ISTEEBU, Bujumbura, Burundi.
Cambodia	2000	geohive	National Institute of Statistics, Cambodia.
Cambodia	2005	geohive	National Institute of Statistics, Cambodia.
Cambodia	2010	geohive	National Institute of Statistics, Cambodia.
Cameroon	1991	geohive	National Institute of Statistics, Cameroon.
Cameroon	1998	geohive	National Institute of Statistics, Cameroon.
Cameroon	2004	geohive	National Institute of Statistics, Cameroon.
Cameroon	2011	geohive	National Institute of Statistics, Cameroon.
Central African Republic	1994	geohive	Census 2003, Central African Republic.
Chad	1996	statoids	Census of Chad. (1993)
Chad	2004	statoids	Census of Chad. (2009)
Colombia	1990	geohive	Departamento Administrativo Nacional de Estadística.
Colombia	1995	geohive	Departamento Administrativo Nacional de Estadística.

Continued on next page

Table A1: International sample of 172 DHS

country	year	online source	ultimate source
Colombia	2000	wikipedia	Census of Colombia. (2005)
Colombia	2005	wikipedia	Census of Colombia. (2005)
Colombia	2010	wikipedia	Census of Colombia. (2005)
Comoros	1996	geohive	Commissariat General du Plan, Union des Comores.
Congo	2007	geohive	http://www.cd.undp.org
Congo Brazzaville	2005	geohive	Centre National de la Statistique et des Études Économiques.
Congo Brazzaville	2011	geohive	Centre National de la Statistique et des Études Économiques.
Côte d'Ivoire	1994	citypopulation.de	Institut National de la Statistique.
Côte d'Ivoire	1998	citypopulation.de	Institut National de la Statistique.
Côte d'Ivoire	2011	citypopulation.de	Institut National de la Statistique.
Dominican Republic	1991	geohive	Oficina Nacional de Estadística.
Dominican Republic	1996	geohive	Oficina Nacional de Estadística.
Dominican Republic	1999	geohive	Oficina Nacional de Estadística.
Dominican Republic	2002	geohive	Oficina Nacional de Estadística.
Dominican Republic	2007	geohive	Oficina Nacional de Estadística.
Egypt	1992	statoids	Central Agency for Public Mobilization and Statistics.
Egypt	1995	statoids	Central Agency for Public Mobilization and Statistics.
Egypt	2000	statoids	Central Agency for Public Mobilization and Statistics.
Egypt	2005	statoids	Central Agency for Public Mobilization and Statistics.
Egypt	2008	statoids	Central Agency for Public Mobilization and Statistics.
Ethiopia	2000	geohive	CSA, Ethiopia.
Ethiopia	2005	geohive	CSA, Ethiopia.
Ethiopia	2011	geohive	CSA, Ethiopia.
Gabon	2000	geohive	Direction Generale de la Statistique et des Etudes Economiques.
Gabon	2012	geohive	Direction Generale de la Statistique et des Etudes Economiques.
Ghana	1993	statsghana	Ghana Statistical Service.
Ghana	1998	statsghana	Ghana Statistical Service.
Ghana	2003	statsghana	Ghana Statistical Service.
Ghana	2008	statsghana	Ghana Statistical Service.
Guatemala	1995	indexmundi	FAO and World Bank population estimates.
Guinea	1999	geohive	Institut National de la Statistique, Guinea.
Guinea	2005	geohive	Institut National de la Statistique, Guinea.
Guyana	2009	geohive	Statistics Guyana.
Haiti	1994	geohive	Institut Haïtien de Statistique et d'Informatique (IHSI), Haiti.
Haiti	2000	geohive	Institut Haïtien de Statistique et d'Informatique (IHSI), Haiti.
Haiti	2005	geohive	Institut Haïtien de Statistique et d'Informatique (IHSI), Haiti.
Haiti	2012	geohive	Institut Haïtien de Statistique et d'Informatique (IHSI), Haiti.

Continued on next page

Table A1: International sample of 172 DHS

country	year	online source	ultimate source
Honduras	2005	geohive	Instituto Nacional de Estadística, Honduras.
Honduras	2011	geohive	Instituto Nacional de Estadística, Honduras.
India	1992	censusindia.gov	Census of India. (1991)
India	1998	censusindia.gov	Census of India. (2001)
India	2005	wikipedia	Census of India. (2011)
Indonesia	2002	geohive	Biro Pusat Statistik.
Indonesia	2007	geohive	Biro Pusat Statistik.
Indonesia	2012	geohive	Biro Pusat Statistik.
Jordan	1997	geohive	Department of Statistics, Amman, Jordan.
Jordan	2002	indexmundi	FAO and World Bank population estimates.
Kazakhstan	1995	geohive	National Statistical Agency of Kazakhstan.
Kazakhstan	1999	geohive	National Statistical Agency of Kazakhstan.
Kenya	1993	statoids	Census of Kenya. (1999)
Kenya	1998	statoids	Census of Kenya. (1999)
Kenya	2003	geohive	Kenya National Bureau of Statistics.
Kenya	2008	geohive	Kenya National Bureau of Statistics.
Kyrgyz Republic	1995	geohive	National Statistical Committee, Kyrgyz Republic.
Lesotho	2004	geohive	Lesotho Bureau of Statistics.
Lesotho	2009	geohive	Lesotho Bureau of Statistics.
Liberia	2007	wikipedia	2008 National Population and Housing Census.
Madagascar	1992	statoids	Census of Madagascar. (1993)
Madagascar	1997	statoids	Census of Madagascar. (1993)
Madagascar	2003	statoids	Census of Madagascar. (1993)
Madagascar	2008	geohive	Institut National de la Statistique, Madagascar.
Malawi	1992	geohive	National Statistical Office, Malawi.
Malawi	2000	geohive	National Statistical Office, Malawi.
Malawi	2004	geohive	National Statistical Office, Malawi.
Malawi	2010	geohive	National Statistical Office, Malawi.
Maldives	2009	wikipedia	Census of Maldives. (2006)
Mali	1995	geohive	Institut National de la Statistique du Mali, Mali.
Mali	2001	geohive	Institut National de la Statistique du Mali, Mali.
Mali	2006	geohive	Institut National de la Statistique du Mali, Mali.
Moldova	2005	geohive	Department of Statistics and Sociological Analysis, Moldova.
Morocco	1992	statoids	Europa World Year Book. (2001)
Morocco	2003	geohive	Haut Commissariat au Plan, Morocco.
Mozambique	1997	geohive	INE, Mozambique.
Mozambique	2003	geohive	INE, Mozambique.
Mozambique	2011	geohive	INE, Mozambique.
Namibia	1992	indexmundi	FAO and World Bank population estimates.
Namibia	2000	geohive	Namibia Statistics Agency.
Namibia	2006	geohive	Namibia Statistics Agency.
Nepal	1996	geohive	Central Bureau of Statistics, Kathmandu, Nepal.
Nepal	2001	geohive	Central Bureau of Statistics, Kathmandu, Nepal.
Nepal	2006	geohive	Central Bureau of Statistics, Kathmandu, Nepal.

Continued on next page

Table A1: International sample of 172 DHS

country	year	online source	ultimate source
Nepal	2011	wikipedia	National Population and Housing Census. (2011)
Nicaragua	1998	geohive	INIDE, Nicaragua.
Nicaragua	2001	geohive	INIDE, Nicaragua.
Niger	1992	geohive	Institut National de la Statistique, Niger.
Niger	1998	geohive	Institut National de la Statistique, Niger.
Niger	2006	geohive	Institut National de la Statistique, Niger.
Nigeria	1999	indexmundi	FAO and World Bank population estimates.
Nigeria	2003	indexmundi	FAO and World Bank population estimates.
Nigeria	2008	indexmundi	FAO and World Bank population estimates.
Pakistan	1990	geohive	Pakistan Census Organisation, Pakistan.
Pakistan	2006	geohive	Pakistan Census Organisation, Pakistan.
Peru	1991	geohive	INEI, Peru.
Peru	1996	geohive	INEI, Peru.
Peru	2000	geohive	INEI, Peru.
Philippines	1993	statoids	Census 2000 of Philippines.
Philippines	1998	geohive	National Statistics Office, Philippines.
Philippines	2003	geohive	National Statistics Office, Philippines.
Philippines	2008	geohive	National Statistics Office, Philippines.
Rwanda	1992	geohive	National Institute of Statistics of Rwanda (NISR).
Rwanda	2000	statoids	Census of Rwanda. (2002)
Rwanda	2005	statoids	Census of Rwanda. (2002)
Rwanda	2010	geohive	National Institute of Statistics of Rwanda (NISR).
São Tomé and Príncipe	2008	geohive	Instituto Nacional de Estatística, São Tomé and Príncipe.
Senegal	1992	indexmundi	FAO and World Bank population estimates.
Senegal	1997	indexmundi	FAO and World Bank population estimates.
Senegal	2005	geohive	ANSD, Senegal.
Senegal	2010	geohive	ANSD, Senegal.
Sierra Leone	2008	geohive	Statistics Sierra Leone, Sierra Leone.
South Africa	1998	geohive	Statistics South Africa & The Local Government Handbook.
Swaziland	2006	geohive	CSO, Swaziland and the National Development Data Centre.
Tanzania	1991	indexmundi	FAO and World Bank population estimates.
Tanzania	1996	geohive	National Bureau of Statistics, Tanzania.
Tanzania	1999	geohive	National Bureau of Statistics, Tanzania.
Tanzania	2004	geohive	National Bureau of Statistics, Tanzania.
Tanzania	2010	geohive	National Bureau of Statistics, Tanzania.
Timor-Leste	2009	geohive	Direcção Nacional de Estatística, Timor-Leste.
Togo	1998	wikipedia	Direction Générale de la Statistique et de la Comptabilité Nationale.
Turkey	1993	indexmundi	FAO and World Bank population estimates.
Turkey	1998	indexmundi	FAO and World Bank population estimates.
Turkey	2003	indexmundi	FAO and World Bank population estimates.
Uganda	1995	geohive	Uganda Bureau of Statistics.

Continued on next page

Table A1: International sample of 172 DHS

country	year	online source	ultimate source
Uganda	2000	geohive	Uganda Bureau of Statistics.
Uganda	2006	geohive	Uganda Bureau of Statistics.
Uganda	2011	geohive	Uganda Bureau of Statistics.
Ukraine	2007	indexmundi	FAO and World Bank population estimates.
Uzbekistan	1996	indexmundi	FAO and World Bank population estimates.
Vietnam	1997	geohive	General Statistical Office, Vietnam.
Vietnam	2002	geohive	General Statistical Office, Vietnam.
Yemen	1991	indexmundi	FAO and World Bank population estimates.
Zambia	1992	geohive	Central Statistical Office, Zambia.
Zambia	1996	geohive	Central Statistical Office, Zambia.
Zambia	2001	geohive	Central Statistical Office, Zambia.
Zambia	2007	geohive	Central Statistical Office, Zambia.
Zimbabwe	1994	geohive	Central Statistical Office, Zimbabwe.
Zimbabwe	1999	geohive	Central Statistical Office, Zimbabwe.
Zimbabwe	2005	geohive	Central Statistical Office, Zimbabwe.

Table A2: Open defecation interacts with population density to predict infant mortality, Bangladesh

	(1)	(2)	(3)	(4)
	dependent variable: IMR (deaths per 1,000)			
local open defecation	$21.92^{\dagger}$	26.76*		
× ln(density)	(12.56)	(12.35)		
local open defecation	30.93**	25.73*		
	(10.82)	(12.54)		
local electrification			-1.08	
× ln(density)			(4.57)	
local electrification			-35.87***	
			(5.18)	
local radio ownership				3.64
× ln(density)				(5.98)
local radio ownership				27.80***
				(8.63)
ln(density)	-0.635	-0.006	$3.68^{\dagger}$	-3.15**
	(1.70)	(1.68)	(2.14)	(1.19)
household open defecation		$9.91^{\dagger}$	7.15	13.51**
		(5.15)	(4.48)	(4.39)
girl		-9.21***	-9.18***	-9.25***
		(2.47)	(2.47)	(2.48)
elapsed months, birth to survey		0.271***	0.275***	0.274***
		(0.041)	(0.041)	(0.041)
region fixed effects		$F = 9.58$	$F = 10.72$	$F = 10.07$
		$p = 0.0000$	$p = 0.0000$	$p = 0.0000$
n (live births)	41,852	41,852	41,852	41,852

Standard errors clustered by survey PSUs in parentheses. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

Table A3: Density of open defecation predicts child weight-for-height, international sample

	(1)	(2)	(3)
	weight-for-height		
local open defecation			
X *ln*(density)	-0.0359	-0.0385+	-0.0496*
	(0.0219)	(0.0218)	(0.0219)
local open defecation	-0.0485	-0.0468	-0.0416
	(0.0937)	(0.0939)	(0.0925)
ln(density)	0.0186*	0.0196*	0.0261**
	(0.00759)	(0.00762)	(0.00824)
household OD	-0.0683***	-0.0687***	-0.0898***
	(0.0161)	(0.0158)	(0.0177)
height (in cm)		-0.00492**	-0.0363***
		(0.00163)	(0.00674)
female			0.0444**
			(0.0139)
female X age-in-months			-0.00188***
			(0.000426)
Age-in-months			0.0285***
			(0.00503)
country Fes	✓	✓	✓
n	292011	292011	292011

Standard errors clustered by survey PSUs in parentheses. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

Table A4: Results are not sensitive to controlling for mother's age at child's birth, international sample

	(1)	(2)	(3)	(4)
	IMR			
local open defecation				
X *ln*(density)	3.273*	3.204*	2.271*	2.255*
	(1.390)	(1.360)	(1.049)	(1.041)
local open defecation	26.37***	26.39***	12.69***	12.57***
	(2.347)	(2.325)	(2.245)	(2.223)
ln(density)	-0.333	-0.344	0.516	0.528
	(0.645)	(0.634)	(0.518)	(0.518)
household OD	6.246***	6.281***	3.102**	3.068**
	(1.711)	(1.717)	(1.049)	(1.040)
mom's age at birth		-0.473***		-0.400*
		(0.109)		(0.182)
extended controls			✓	✓
country FEs	✓	✓	✓	✓
n	1109116	1109116	942350	942350

Standard errors clustered by survey PSUs in parentheses. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$.

Table A5: Results are robust to controlling for interaction of population density and other community-level infectious disease behaviors

	(1)	(2)	(3)	(4)
local open defecation	3.244*	1.979*	2.180*	3.184+
× *ln*(density)	(1.372)	(0.905)	(1.024)	(1.613)
local BCG vaccine		-4.932**		
× *ln*(density)		(1.540)		
local measles vaccine			-2.448+	
× *ln*(density)			(1.310)	
local bed net use				-2.523
× *ln*(density)				(2.048)
local open defecation	11.625+	6.326	4.664	0.258
	(6.218)	(4.587)	(5.142)	(6.352)
local BCG vaccine		-28.230***		
		(7.613)		
local measles vaccine			-47.606***	
			(6.819)	
local bed net use[a]				-5.982
				(9.016)
ln(density)	-1.468**	3.688*	1.004	-0.464
	(0.443)	(1.393)	(1.038)	(1.075)
household OD	6.165***	5.960***	5.887***	5.126***
	(1.739)	(1.664)	(1.634)	(1.167)
country fixed effect	✓	✓	✓	✓
n (live births)	1087033	1080212	1080207	375236

Standard errors clustered by survey PSUs in parentheses. Two-sided p-values: † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$ *** $p < 0.001$. [a]Fewer observations are included in column 4 because only certain African countries collected data on whether children under five slept under a bed net the night before the survey.