# Health and hunger: disease, energy needs, and the Indian calorie consumption puzzle

Josephine Duh and Dean Spears*

April 5, 2015

### Abstract

India's experience presents a puzzle at odds with basic facts of household economics: amidst unprecedented economic growth, average per capita daily calorie consumption has declined in recent decades. Does an improving disease environment explain the calorie decline? A diminished burden of infectious disease could lower energy needs by increasing absorption and effective use of calories. We document a robust effect of disease exposure – measured as infant mortality and as poor sanitation – on calorie consumption. Similar effects are found using multiple datasets and empirical strategies. Disease can account for an important fraction of India's calorie decline.

# 1    Introduction

Food consumption has long been central to household economics and the measurement of well-being. It has been documented at least since the 19th century that, particularly in developing countries, richer households consume more food on average than poorer households.[1] However, recent trends in India challenge our understanding of the basic facts of household economics and present a major puzzle. With rapid economic growth, the average Indian household has become richer over time. However, calorie consumption has been consistently declining over recent decades (Deaton and Drèze, 2009). From 1987-88 to 2004-05, average daily per capita calorie consumption decreased by about 130 calories, or 6 percent, and calories from cereals[2] fell by about 200 calories, or 14 percent.

What factors explain India's decline in calorie consumption amidst economic growth? This paper investigates one novel hypothesis: calorie consumption has fallen in part because slow improvements in India's disease environment have lowered energy intake needs. India has exceptionally poor sanitation: 60 percent of people worldwide who defecate in the open without using a toilet or latrine live in India.[3] Partially in consequence, India's infant mortality rate is higher than that of other countries with similar levels of national income per capita. Reductions in this considerable burden of disease may have allowed Indian consumers to retain, absorb, and use more of the food that they eat. Because many developing countries, including India, continue to suffer from much preventable disease, it is important to understand the economic impact of disease on average calorie needs.

This paper makes two contributions to the literature. To our knowledge, we are the first to estimate an importantly large, but plausibly sized, effect of the disease environment on average calorie consumption. Households exposed to a greater burden of infectious disease

---

[1]"Food consumption" refers to caloric intake as opposed to food variety.

[2]Cereals refer to: rice/rice products, wheat/wheat products, jowar/jowar products, bajra/bajra products, ragi/ragi products, maize/maize products, small millets/small millets products, and barley/barley products.

[3]The 1981 Census of India did not measure rural sanitation because rural open defecation was almost universal.

eat more. In addition, we present empirical evidence that improvements in the disease environment can account for a large fraction - possibly one-fifth or more - of the recent calorie consumption decline in India. Although none of our estimates suggest that disease can account for the entire calorie decline, our results advance an emerging literature highlighting the key role of disease externalities for nutrition (Smith et al., 2013; Spears, 2013).

Employing two empirical strategies and three separate datasets with complementary advantages, we estimate comparably-sized effects of the disease environment on calorie consumption. Indian districts that experienced larger declines in infant mortality from the mid-1980s to mid-2000s also saw larger decreases in average per capita calorie consumption. Similarly, the average household living in an area with worse sanitation and higher infant mortality consumed more calories than otherwise comparable households living in places with lower rates of open defecation and infant mortality. Effects are pronounced in areas where moree children suffer from diarrhoea; no similar effect is seen of fever or cough.

In the absence of an effect of the disease environment on nutritional needs, this result would appear paradoxical: generally, poorer people both suffer worse disease and eat less. Alternative causal explanations are challenged to account for the fact that eating more – which is typically associated with socioeconomic advantage – robustly coincides with exposure to disease externalities – which are typically associated with socioeconomic disadvantage.

## 1.1 Motivation: Three puzzles of nutrition and consumption

From the 1980s to the 2000s, average daily calorie consumption in India decreased by an economically important amount. As Deaton and Drèze (2009) summarise, "this decline has occurred across the distribution of real per capita expenditure, in spite of increases in real income and no long-term increase in the relative price of food" (42). Perhaps most puzzlingly, this decline occurred during an exceptional period of rapid economic growth in India: real GDP per capita grew at 4 percent per year from 1983 to 2005. Among a wide set of candidate explanations that they consider, Deaton and Drèze (2009) hypothesise that

3

"calorie requirements have declined due to lower levels of physical activity or improvements in the health environment." Eli and Li (2013) carefully consider the first suggestion and find that changing work requirements and physical activity are unlikely to account for more than one-third of India's calorie decline. Since much of the calorie decline therefore remains to be explained, we consider the effects of health: if an improved disease environment allows for better absorption and efficient use of caloric intake, how much of the decline could disease explain?

Although we do not directly address it in this paper, a related nutritional puzzle in India concerns anthropometric outcomes. People in India are exceptionally short in international comparisons, especially given their relatively high average incomes among developing countries (Deaton, 2007). It is particularly puzzling that people in India are shorter, on average, than people in sub-Saharan Africa who are poorer, on average; this fact is sometimes called the "Asian Enigma" (Ramalingaswami et al., 1996) and has received much recent attention from economists (Tarozzi, 2008; Jayachandran and Pande, 2013).[4] In a related finding that motivates our analysis, Spears (2013) demonstrates that children in populations exposed to more open defecation are shorter, on average; exceptionally poor sanitation in India can statistically account for India's deficit relative to Africa in child height. Economic historians have documented a large association between population height and the disease environment, as reflected in mortality rates Bozzoli et al. (2009). Hatton (2013), studying the historical increase in European height, concludes that "the most important proximate source of increasing height was the improving disease environment as reflected by the fall in infant mortality" (1).

A third motivating nutritional puzzle is the limited success of interventions that aim to improve nutritional outcomes (such as child height) through direct provision of food or nutrients. Numerous field experiments have documented limited success of nutritional

---

[4]Moreover, differences in economic growth across Indian states have not been associated with reductions in child malnutrition (Subramanyam et al., 2011; Coffey et al., 2013).

supplementation. Describing these as "nutrition-specific interventions," Bhutta et al. (2013) estimate that if 10 core nutrition-specific interventions were scaled-up to 90 percent, child stunting would fall by about one-fifth. If disease interacts with nutrient intake (*e.g.* Menon et al., 2013) such that a reduced burden of disease allows children's bodies to absorb and make more efficient use of nutrients from the intervention, then "nutrition-sensitive interventions" – such as improving the disease environment – may help to bridge the gap.

This paper builds on an important literature in economic history on nutrition and the demand for calories. Much of this literature studies Europe in the 18th through 20th centuries, where health consequences of sanitary environments have received considerable attention (Fogel, 1997; Preston and van de Walle, 1978). In a puzzle sharing similarities to the Indian puzzle that we study, Gregory Clark and Lindert (1995) show that food material supplies decreased in Britain from 1770 to 1850, despite growth in income. They allude to the possibility of changes in food demand related to disease but do not test for it directly. Given that the first treated water supply in London was not until 1829 and that the 1855 Metropolitan Water Act came after the sample period, disease may not be the primary driving force behind the demand shift in their study. Logan (2009) finds that calorie expenditure elasticities among 19th-century industrial workers were much greater than the elasticities among similarly poor or poorer people in developing countries today, suggesting that these historical workers were hungrier. One reason why the industrial workers were hungrier may be that, holding standard of living constant, sanitation and the disease environment was considerably worse historically, before widespread acceptance of the germ theory of disease.[5] Deaton (2006), reviewing Fogel and reflecting on this historical and anthropometric literature, expresses concern that "the synergism between economic growth and the growth of the size and the durability of the human body can turn in to an overemphasis on links between economic growth and health and an underemphasis on the role of disease and its prevention" (111).

---

[5]Disease externalities could also contribute towards explaining Logan's (2006) finding that the expenditure elasticity of calorie was greater for British than for American households in the late 19th century, if population density was greater in Great Britain.

If so, then this paper's estimate of effects of high disease burdens on calorie demand is an important step towards understanding a long-discussed but under-quantified mechanism.

## 1.2   Links between the disease environment and nutrition

Could a greater burden of disease increase calorie needs – either because nutrients eaten are not absorbed due to diarrhoea, parasites, or intestinal dysfunction, or because the body uses energy fighting disease (Stephensen, 1999)? As Deaton (2007) explains, anthropometric outcomes reflect "net nutrition," meaning nutrient intake *net* of losses to disease. If disease does increase nutritional requirements, then people in India would face an exceptional risk: more than half of households in India defecate in the open without using a toilet or latrine (WHO and Unicef, 2012), and open defecation is just one of many sources of infectious disease.

Although no prior paper in economics has sought to estimate an effect of the disease environment on calorie consumption, a substantial econometric literature documents effects of sanitation and water on health. Cutler and Miller (2005) document a large effect of water filtration and chlorination on mortality in major U.S. cities in the early 20th century. Similarly, Watson (2006) studied heterogeneous timing of public health investments – including sewer connections and septic tanks at U.S. Indian reservations. Watson found that a 10 percentage point increase in the fraction of homes receiving improved sanitation reduced infant mortality by 2.5 percent among Native Americans. Galiani et al. (2005) show that privatization of water supply in Argentina reduced child mortality by 8 percent.

Other papers in economics trace effects of sanitation onto nutritional outcomes and their eventual consequences for human capital. Bleakley (2007) documented that eliminating hookworm in the American South led to an increase in literacy and average incomes; worms and other parasites are a key mechanism by which disease could increase food intake needs. Baird et al. (2011), following up on Miguel and Kremer (2004) deworming experiment in Kenyan schools, found that children who received the deworming treatment grew up to be

adults who work more hours. Coffey (2013) finds that reductions in open defecation in Nepal are associated with increases in average haemoglobin levels, suggesting that poor sanitation – and resulting malabsorptive disease – could be one contributing cause of anaemia. Lastly, Spears (2012b) studied a government sanitation program in rural India. Exploiting heterogeneity in implementation throughout rural India, Spears finds that the where the program was active, it both reduced infant mortality and increased children's height, on average.

A growing biomedical literature on links between sanitation, disease, and nutrition is consistent with the possibility of an an effect of the disease environment on calorie demand, due to nutrient absorption and use. Research on the interaction between infection and nutrition has built upon early insights of Scrimshaw et al. (1968).[6] This literature on nutritional consequences of infectious disease has traditionally concentrated on diarrhoea (Guerrant et al., 1992; Checkley et al., 2008). However, recent research has concentrated on the possible importance of *environmental enteric disfunction* (EED), an inflammatory response of the intestines to chronic infection, resulting in reduced nutrient absorption (Humphrey, 2009; Mondal et al., 2011). Recent epidemiological studies have exploited novel methods to measure markers of EED, and have found strong associations among environmental sanitation, EED, and nutritional outcomes (Lin et al., 2013; Kosek et al., 2013). Although none of these studies have measured an economic *response* to these conditions, both diarrhoea and malabsorption due to EED as well as energy demands of fighting disease could be consistent with increased demand for calories where the disease burden is greater.

## 1.3   Outline

We conduct three complemeentary empirical analyses to estimate effects of the disease environment on calorie consumption. In our empirical analyses, we use two measures of disease externalities: infant mortality and open defecation. Infant mortality rates have long been

---

[6]Scrimshaw et al. (1968) write: "Infections so consistently worsen nutritional status that they must be taken into account in all clinical problems and public health programs that involve persons whose diet is inadequate or whose nutritional status is suboptimal" (59).

used as a measure of the disease environment by economic historians, and have been shown to correlate with anthropometric nutritional outcomes (Bozzoli et al., 2009; Hatton, 2013), although to our knowledge IMR has never previously been linked directly with increase calorie consumption. Open defecation is an important source of disease in India and can explain variation in child height internationally and within India (Spears, 2012a). Finding similar results with both explanatory variables contributes to our interpretation of our estimates as reflective of an effect of the disease environment.

First, in section 2, we combine two surveys of food consumption in India, one from 1987/88 and one from 2004/05, to create a district-level panel from repeated cross-sections. Having merged consumption data with infant mortality rates from the Indian Census, we employ a difference-in-differences identification strategy: districts that experienced sharper declines in infant mortality rates over these two decades also saw greater decreases in calorie consumption, conditional on overall household expenditure. Second, in section 3, we study a nationally-representative cross-sectional survey that uniquely combines data on food consumption, mortality, sanitation, and anthropometric outcomes. We show that households living in rural villages or urban blocks with higher rates of infant mortality (or open defecation) consume more calories, on average; we also confirm that adult women living in areas with more disease have lower body mass, even after accounting for differences in household expenditure and calorie consumption. Finally, in section 4, we use a unique survey dataset from 1983 which combines data on consumption and local sanitation with demography-specific employment information within households; this allows us to replicate our main result while verifying that variation in energy requirements for work are unlikely to drive our results. Section 5 quantitatively compares the estimates from these three approaches and applies decomposition methods in the spirit of Blinder-Oaxaca to estimate the fraction of the Indian calorie decline that can be explained by improvements in the disease environment.

# 2   Evidence from changes over time within districts

Did average per capita calorie consumption more steeply fall in districts where infant mortality rates more sharply declined between 1987-88 and 2004-05? In this section, we answer this question by combining two cross-sectional rounds of India's National Sample Survey (Rounds 43 and 61) in order to use a panel-based identification strategy (Deaton, 1985). We focus on these two survey rounds because the data allow us to identify the districts in which households live; we then match districts across years and merge the consumption data with infant mortality records from the Indian census. By focusing on within-district changes in calorie consumption and infant mortality, we are able to rule out confounding time-invariant factors, such as climate or average genetic potential body size of the local population. We first describe the data and our empirical strategy in section 2.1, and then, in section 2.2, we present the regression results.

## 2.1   Empirical strategy

Our outcome variable is daily per capita calorie consumption from the National Sample Survey (NSS). The NSS is a nationally representative household-based survey conducted by the National Sample Survey Office (NSSO) in the Ministry of Statistics and Programme Implementation. The NSSO annually fields the Consumer Expenditure Survey (CES) and Employment-Unemployment Survey (EUS) using a two-stage sample design covering all Indian states.[7]

We follow Deaton and Drèze (2009) in calculating calorie consumption from households' 30-day recall of food expenditures in the CES. Households are asked to report quantities of more than 200 items consumed from home production, market purchase, and free collection

---

[7]There are "thick" and "thin" rounds that correspond to the survey sample size: "thick" rounds, which include approximately 120,000 households in each survey, serve as the basis of India's consumption and poverty statistics, whereas "thin" rounds, approximately 40,000-50,000 households, help to monitor trends in-between "thick" rounds. Both Rounds 43 and 61 are "thick" rounds, and in our study, we use the CES to look at household food and non-food expenditures.

or gifts, and households are probed for total spending on each item purchased from the market. Using conversion factors from Nutritive Value of Indian Foods by Gopalan et al. (1989) provided by the NSSO survey reports, we are able to translate quantities into calories, and we divide by the household size to obtain average calories consumed per person.

Our explanatory variable that measures the disease environment is the district infant mortality rate (IMR). Because infants are highly sensitive to respiratory infections and intestinal diseases, the literature has frequently used IMR to proxy for the prevailing disease environment (Bozzoli et al., 2009; Hatton, 2013). IMR is defined as the number of deaths among babies less than 12 months old per 1000 live births. Because the NSS does not record information on infant mortality, we merge NSS data with district-level IMR from the Indian Census.[8]

Summary statistics of NSS Rounds 43 and 61 are presented in the columns 1 and 2 of Table 1. On average, daily calorie consumption per person decreased by 136 calories (or 6.3 perceent of average consumption level in 1987/88) between 1987/88 and 2004/05. The decline was mainly driven by a drop of 214 cereal calories (or 13.8 percent) per person.

As calorie consumption was falling, the disease environment and socioeconomic conditions were improving. In the late 1980s, the average Indian household lived in a district where IMR was 89 deaths per 1000 live births; by the mid-2000s, IMR decreased to 58 deaths per 1000. Over the same period, latrine coverage increased, which contributed to the reduction of disease (Spears, 2012b). Real[9] monthly per capita expenditures rose by 23 percent, and reported ownership of a TV increased by 14-fold. The fraction of illiterate women in the population was cut by nearly one-third and almost one-half for men.

To estimate the effect of the disease environment on calorie consumption, we use a fixed

---

[8]Because the Census years (1981, 1991, 2001) do not align with the interview dates for our panel, we interpolated IMR between rounds using different functional forms, namely linear and logarithmic, for each district. We found that the results were robust to the various interpolations, and in the main body of this paper, we present results assuming that the log of IMR is linear in time.

[9]We deflate nominal monthly per capita expenditures by the Consumer Price Index for Agricultural Laborers in rural areas (by month and year) and Consumer Price Index for Industrial Workers in urban areas (by month and year).

effects model that also allows us to control for household wealth, primary source of income, and education:

$$calories_{idt} = \beta_0 + \beta_1 IMR_{dt} + \beta_2 \ln{(MPCE)}_{idt} + X_{idt}\theta + \gamma_d + \delta_t + \varepsilon_{idt}, \tag{1}$$

where $i$ indexes households, $d$ districts, and $t$ years, here survey rounds. The outcome variable ($calories_{idt}$) is per capita consumption of total calories or cereal calories and the key explanatory variable is the district's infant mortality rate during the survey year ($IMR_{dt}$). To make use of the available information on households that may also correlate with calorie consumption, we conduct the analysis at the household level, and since IMR varies by district, we cluster standard errors at the district level.[10]

We add covariates ($X_{idt}$) in stages to demonstrate the stability of our result. We include urban residence, the household's monthly per capita expenditures (MPCE) in logs, caste and religion of the household head, literacy of adult male or female, and 34 categorical variables of the household's primary occupation. We include district ($\gamma_d$) and time ($\delta_t$) fixed effects. Note that, with district fixed effects, the coefficients are estimated based on changes within districts over time. Finally, we replicate all regressions with and without detailed semi-parametric controls for the count of household members of each sex in a set of age ranges, to verify that no demographic properties of households are responsible for our results.[11]

Our preferred results focus on calories from cereals rather than calories from all foods because we are interested in the effect of the disease environment on calorie needs: in most cases, cereals are the main and cheapest source of energy (Jensen and Miller, 2010). Nev-

---

[10]For robustness, we also conducted the analysis after aggregating to the district level, i.e. regress mean per capita calorie consumption on infant mortality rates and log of the district average monthly per capita expenditures with district-level observations. To ensure that our results are not driven by outliers, we trim the sample for districts with mean calorie consumption levels in the far upper and lower tails (top and bottom 1 percent by survey round). In Appendix Table 1, we show the results with a balanced panel, *i.e.*, only districts with observations in both rounds; this amounts to trimming approximately 6% of districts with non-missing values for IMR. The coefficient estimates for IMR are consistent with Table 2, and F-tests show that we cannot reject that the estimates are equal to 1.

[11]The set of age ranges are: 0-4 years old, 5-9 years old, 10-14 years old, 15-39 years old, 40-49 years old, 50-59 years old, and 60 or more years old.

ertheless, for robustness, we also show results using calories from all food groups as the outcome variable.

## 2.2 Effect of the disease environment on calories consumption

Figure 1 illustrates two central points of our paper. First, households that live in places with a greater burden of disease eat more calories, on average, despite their overall greater disadvantage. Second, changes in the disease environment can statistically account for much of the changes in calorie consumption over time. The graph plots the within-year associations between district IMR and household calorie consumption from cereals for each of the two time-periods as non-parametric local regressions. Plotted over these lines are the annual averages of calorie consumption and IMR. The larger vertical distance between the dots is the full 200 cereal calorie decline; the smaller vertical distance between the lines indicates that *at the same level of infant mortality* the difference between the two time periods in cereal calorie consumption is small. Therefore, the within-year gradient between disease and consumption can statistically explain a visible fraction of the calorie decline.[12]

Table 2 reports estimates of equation 1, verifying the statistical robustness and significance of the associations documented in the figure. Panels A and B are both included, for each regression specification, to verify that differences in demographic structure of households are not responsible for our results. For the main result in columns 1 through 7, the outcome variable is per capita calorie consumption of cereals; column 8 substitutes total calorie consumption. Across columns, we add controls for heterogeneity across households in wealth or occupation. Overall, as shown by the consistently positive estimate in the first row, districts with the largest declines in IMR also observed the biggest drops in consumption of calories from cereals or from all food groups. The results suggest that improvements in the

---

[12]We note that the level of cereal calorie consumption is steady around 1400 calories per person per day for households living in districts with IMR exceeding 70 deaths per 1000 live births in 2004/05. The cereal share of total calorie consumption is also fairly flat in the range of 70-75 percent in these districts in 1987/88 and in 2004/05.

disease environment led to lower energy consumption.

In column 1, we replicate the basic puzzle: daily per capita calories consumption of cereals fell by 213 kcal over 17 years. With the inclusion of IMR in the regression in column 2, we see that the unexplained gap is reduced by 63 percent (62 percent in Panel B with demographic controls), although it remains statistically significant. Households living in districts with higher IMR consumed more calories from cereals, and more calories overall. The positive sign on IMR is important: if high infant mortality is a marker for a poor disease environment and if wealthier areas generally have greater access to food, health care, and public services like sanitation, then the estimate would seem to go in the "wrong" direction – in the absence of the effect that this paper documents – because it implies that households in poorer disease environments eat *more* calories.

With fixed effects, we identify the coefficient on IMR ($\beta_1$) from district trends rather than levels; without these fixed effects, the model may be mis-specified to estimate a causal effect. Indeed, the magnitude of the estimate decreases, suggesting that cross-sectional differences partially account for variation in calorie consumption. Once district fixed effects are added in column 3, including further controls – such as household consumption, urban residence, average district consumption, and 34 occupation groups – changes the coefficient very little. In other words, all specifications based on within-district changes indicate that an additional infant death per 1,000 live births is associated with each person eating about one more calorie per day, on average.[13] Notably for our empirical strategy, adding state-specific linear time trends in column 6 does not reduce the coefficient estimate, suggesting that our result is not due to heterogeneity in spurious secular trends. These results suggest that improvements in the disease environment can account for some of the Indian calorie decline puzzle; in section 5, we will combine these estimates from others in the paper to assess the fraction of the decline that disease can explain.

---

[13]We cannot reject that $\beta_1 = 1$ in each specification represented in columns 3-8.

# 3    Evidence from local disease environments

We have seen that the districts where we observe greater improvements in the disease environment also experienced larger declines in average calorie consumption. This section exploits comparative advantages of the 2005 India Human Development Survey (IHDS) (Desai et al., 2009).

Although the IHDS is a cross-section – therefore, we cannot directly study changes over time – this disadvantage is balanced by three advantages that complement the panel analysis of NSS data in the previous section. First, the IHDS includes anthropometric measures of nutritional status for ever-married women 15-49 years old, children less than 5 years old, and children 8-11 years old. Second, the IHDS permits a wider range of controls for economic, social, demographic, and occupational characteristics of households. Third – and perhaps most importantly – we can use the IHDS to compute *local* measures of the disease environment by matching households to survey primary sampling unit (PSU) level estimates of infant mortality and sanitation coverage. Districts in India, studied in the previous section, are very large and contain much heterogeneity; PSU-level explanatory variables will more accurately capture the local disease environment to which households are frequently exposed.

This section proceeds in four parts. First, section 3.1 documents that households consume more calories if exposed to a larger fraction of local neighbours who defecate in the open, even controlling for overall consumption and a range of social and demographic factors. Next, section 3.2 replicates this result and that of the previous section by showing that households living in PSUs with more local infant mortality consume more calories. Then, section 3.3, presents a falsification test of the association between local morbidity and calorie consumption, showing that people who live near children with diarrhoea eat more, but that other types of disease predictably have no effect. Finally, in a check of the plausibility of a nutritional effect of the disease environment, section 3.4 demonstrates that adult women exposed to more open defecation have lower BMI, on average, even after accounting for

household expenditure and calorie intake.

## 3.1 Effect of sanitation on calorie expenditure

Do households exposed to more open defecation at the village or city sub-block level consume more calories on average? Exposure to germs in faeces may cause diarrhoea and other intestinal disease that diverts food from nutritional uses.

### 3.1.1 Empirical strategy

The empirical strategy of this section compares households exposed to different levels of local area open defecation at a fixed point in time. In particular, as an explanatory variable, we compute the fraction of households that defecate in the open instead of using a toilet or latrine for each PSU.[14] We estimate the association of this variable with calorie consumption as follows:

$$
\begin{aligned}
calories_{ip} = \quad & \beta_0 + \beta_1 \, local \ open \ defecation_p + \beta_2 household \ open \ defecation_{ip} + \\
& \beta_3 \ln{(MPCE)_{ip}} + \beta_4 urban_p + D_{ip}\theta_1 + Y_{ip}\theta_2 + S_{ip}\theta_3 + E_{ip}\theta_4 + \varepsilon_{ip},
\end{aligned}
\tag{2}
$$

where $i$ indexes individual household and $p$ indexes survey PSUs. As in section 2.2, we show that the results are robust to using either total calorie consumption or cereal calorie consumption as the outcome variable. The explanatory variable of interest is local area open defecation, a percent from 0 to 100. An indicator for a household's own open defecation is further included; this focuses the analysis on sanitation *externalities* while controlling for any average wealth difference between households who do and do not safely dispose of faeces.

Controls are added in stages to demonstrate robustness. In addition to household monthly consumption per capita and an indicator for urban residence, four vectors of controls are added. Demographic variables $D$ are household size and the number of children in the

---

[14]Spears (2013) and Kov et al. (2013) have shown in various contexts that this variable is associated with child height-for-age, which is commonly cited as an indicator of nutritional status.

household. Income sources, $Y$ are as assigned by the IHDS into eleven categories: cultivation, allied agriculture, agricultural labour, non-agricultural labour, artisan, petty trade, business, salaried, professional, pension/rent, and others. Social groups $S$ classifies households into one of eight groups: Brahmin, other higher caste, other "backwards" caste, Dalit, Adivasi (or "tribal"), Muslim, Sikh or Jain, and Christian. Finally, education $E$ is a set of indicators for the highest education level of an adult in the household and an indicator for having at least one literate household member.

As a robustness check, we will also include a specification where we control for local variation in the prices of rice and wheat, as reported to surveyors by interviewed households. These controls may not belong in a well-specified model: if a worse disease environment indeed increases demand for food at all prices, then disease will endogenously cause an increase in price. However, we include a specification with this control to verify that the result does not change. Similarly, we control for PSU average monthly per capita expenditure, and find similar results.

### 3.1.2 Results

Figures 2 and 3 provide initial evidence from the IHDS of a gradient between sanitation and total or cereal calorie consumption, respectively. Both graphs plot non-parametric regressions of average daily calories against overall household monthly consumption per capita. Unsurprisingly, the graphs slope upwards, as richer households eat more. In both graphs, households are split into three categories according to the local disease externalities to which they are exposed: households in which no household surveyed in their PSU defecates in the open, households living in PSUs where all households defecate in the open, and household living in PSUs at an intermediate level of open defecation. The graphs show that – at all levels of household consumption per capita – households exposed to more open defecation consume more calories, on average. The space between confidence intervals confirms that these differences are statistically significant, and the gap at all levels of economic status

16

suggests that the association between sanitation and calorie consumption does not merely reflect omitted wealth.

Are these differences robust to controls for demographic, income, social, and educational characteristics of the household? Table 3 indicates that they are. As we add covariates to the regression model, the sanitation-calories gradient remains stable at about 1 calorie per percentage point of local open defecation exposure.[15] To put this number in perspective, this would imply about a difference of 10 cereal calories per day associated with the 10 percentage point decline in open defecation in India between the 2001 and 2011 census rounds.

As before, this result may appear surprising because *poorer* people are more likely to defecate in the open, and *richer* people eat more calories on average. This can be seen in the flip of the sign of *household* open defecation from statistically significantly negative in column 1 of Panel B to statistically significantly positive in column 2, once the more precise control for household economic status is added. Additionally, the quantitative robustness of the main result to the sets of controls suggests that our finding is not a spurious reflection of heterogeneity in socioeconomic status or work requirements.

## 3.2   Effect of IMR on calorie expenditure

Can the results of section 2 be replicated using local, *i.e.*, village or urban sub-block level, infant mortality rates computed using IHDS in place of district IMR in the Indian census? In this sub-section, we use the IHDS to estimate equation 2 from section 3.1 with infant mortality substituted for sanitation as the key explanatory variable. In particular, we compute the fraction of live births who reportedly died before their first birthday, linearly scaled as a count of deaths per 1,000 live births, for each survey PSU.[16]

---

[15]This point estimate is smallest in column 6, but a $F$-test does not reject a null hypothesis that the coefficient is 1 even with this smaller estimate.

[16]On the one hand, this specification reduces measurement error in IMR exposure, relative to the district-level analysis, by matching the exposure more narrowly to a household's local environment. On the other hand, this specification increases measurement error by estimating each IMR statistic from a smaller sample. If the increase in measurement error dominates, this would cause attenuation bias, which would bias our estimate of the effect of IMR on calorie consumption towards zero.

Table 4 presents the results. The estimates are generally quantitatively consistent, albeit smaller in magnitude, with the district-level fixed effects results in Table 2; an extra infant death per 1,000 births is associated with 0.3 to 1.0 more calories consumed by each person each day. The smaller coefficient may reflect attenuation, since local IMR is computed from a small intra-PSU sample. Our preferred specification, column 3, includes all of the controls from Table 3. As a step towards replicating the district-level results, column 4 adds state fixed effects. Although we believe this is likely over-controlling because the disease environment importantly varies at the state level, we include it for robustness and note that a statistically significant gradient remains. Column 5 shows that IMR and local sanitation are both predictors of calorie consumption when included together, which is consistent with a multidimensional disease environment (Coffey et al., 2013). Finally, reiterating our concerns about the endogenous determination of prices, column 6 includes controls for cereal prices and finds coefficients qualitatively consistent with the other results.

## 3.3 Falsification test: Type of disease and calorie consumption

Section 3.2 showed that households living in local areas with higher infant mortality rates consume more calories, on average; this result is informative because IMR is an important and widely used measure of disease environments, but it is not specifically a measure of the type of disease most related to nutritional outcomes and needs. In this section, we report a similar analysis using different explanatory variables: the fraction of children under 5 in a local area who have suffered from diarrhoea, fever, or cough in the last month, as reported by their mothers. If our results are indeed driven by enteric morbidity reducing the absorption and use of nutrients in food, then we would expect to see an association between calorie consumption and diarrhoea, but not for other causes of disease.[17]

---

[17]As throughout this paper, this is an analysis of the local disease *environment*, including externalities. We can also compare calorie consumption in households whose own members have more or less diarrhoea. A one standard deviation (9 percentage point) increase in the fraction of household members having diarrhoea in the last month (where diarrhoea is measured as residuals after detailed demographic controls for age structure and sex) is associated with the households consuming 43 more calories per capita per day, on

Table 5 presents the results. There is an economically large and statistically significant association between local diarrhoea and calories consumption, but there is no association for fever or cough. A 10 percentage point increase in the fraction of a household's neighboring children suffering from diarrhoea is associated with an approximately 35 calories per person increase in daily consumption; estimates for fever and cough are of much smaller magnitude and are not statistically distinguishable from zero. These results are unchanged by controlling for household consumption, local area average consumption, and the demographic structure of the household (indicators for each count of number of persons, children, teens, married males, and married females). The specificity of this result is consistent with a causal effect of the enteric disease environment on calorie needs, which is precisely what would be most influenced by improvements in sanitation.

## 3.4 Effect of the disease environment on BMI

If the local disease environment increases calorie needs, then we might expect to see it reflected in measured nutritional status. Although height reflects early-life health and net nutrition and would not be expected to respond to the current disease environment, weight-for-height in contrast, reflects more recent net nutrition. Indeed, given the difficulty of meaningfully measuring diarrhoea morbidity with surveys, some researchers advocate using child weight as a proxy indicator of recent disease (Schmidt et al., 2011). Therefore, as a verification of the nutritional mechanism of our main results, we study the association between the local disease environment and the body mass index (BMI) of ever-married adult women, aged 20 to 45.

Do women exposed to more disease externalities weigh less, and if so, do the differences in body size merely reflect economic status or consumption? Figures 4 and 5 show an association between sanitation and average body mass across levels of household expenditure

average, controlling for log MPCE, urban residence, and non-parametric indicators for household size and for the count of children.

using the same sample-splitting strategy as in Figures 2 and 3. As Figure 4 unsurprisingly depicts, richer women in India weigh more, on average. However, at all levels of household consumption, women exposed to more local area open defecation weigh statistically significantly less. Figure 5 repeats this analysis, using household daily calorie consumption per capita in place of overall consumption. Strikingly, there is little apparent relationship between calorie consumption and women's weight.[18] As before, at all levels of household calorie consumption, women exposed to less open defecation weigh statistically significantly more.

Table 6 verifies the statistical significance and robustness of this result. Living in a PSU where nobody defecates in the open is linearly associated with being about one BMI point heavier, on average, than living in a PSU where everybody defecates in the open. This result is stable controlling for household overall and calorie consumption and for several vectors of controls. Column 7 demonstrates that the result is unchanged when the woman's height is added as a further control; height is a marker of early-life well-being and is in the denominator of BMI, so this control verifies that a mechanical correlation with height is unlikely.[19] These results are consistent with our overall interpretation of our findings: exposure to a more threatening disease environment increases calorie needs.

[18]Deaton and Drèze (2009) note that declining calorie consumption in India coexists with small improvements in child nutritional status. Spears et al. (2013) show that although there is a relationship between district level open defecation and child stunting rates in India, stunting rates are uncorrelated with district calorie consumption averages. We also observe that *household* calorie consumption may be a poor measure of the calorie consumption of married women of childbearing age, who often suffer low intrahousehold status in India.

[19]As a falsification test, we examine whether local sanitation coverage predicts women's heights, conditional on household wealth. Because most women in India move to their husband's family's home upon marriage, many women may face a different disease environment as adults than they did as children. As shown in appendix Figure 2, there is little correlation between height and contemporaneous rates of open defecation. This non-result is reassuring because adult height – as a marker of early-life health and cumulative health shocks (e.g. Case and Paxson, 2008) – should not systematically vary with current sanitation coverage unless healthier people are selected into better environments, even after accounting for differences in wealth.

# 4   Evidence from detailed occupational data

We have shown an effect of the disease environment on calorie consumption using variation within district over time and across villages or city sub-blocks in the cross-section. Households in worse disease environments consume on average more calories from cereals and all food groups. As noted by Deaton and Drèze (2009) and explored in detail by Eli and Li (2013), one candidate explanation for India's calorie decline is the reduction in energy requirements for work. Both lower work requirements and an improving disease environment are likely to have contributed to India's calorie decline, and this section asks whether changing patterns of work are an omitted variable in our estimates that drives the effect of the disease environment.

The analyses in sections 2 and 3 both included controls for urban or rural residence and for occupational categories. We now turn to the NSS 1983 cross-section because the survey collected information on calorie consumption, local sanitation coverage, and exceptionally detailed occupational categories for each household. Our main finding is that, even conditional on detailed controls for work, along with MPCE and household characteristics, households from villages or urban sub-blocks with worse sanitation consumed more calories from cereals and from all food groups. This suggests that, although differences in energy needs for income-generating activities may be independently important, they are not responsible for our results.

## 4.1   Empirical Strategy

In NSS Round 38 (1983), the same households were interviewed for the CES and EUS. In addition, households reported whether they use a toilet or latrine.[20] The dataset does not have district identifiers, and therefore, we cannot link it to our district panel nor can

---

[20]The analysis linking sanitation, calories consumption, and employment can only be done with this "thick" survey round. The question about household toilet use was dropped from future questionnaires for these surveys. After Round 50 (1993/94), the CES and EUS were no longer administered on the same set of households because the surveys took too much time.

we match district-level IMR from the Census. Instead, as in section 3, we compute PSU-level[21] sanitation coverage as a measure of local sanitation conditions to proxy for the disease environment.

Sample statistics are reported in column 4 of Table 1. Note that NSS Rounds 38 (1983) and 43 (1987/88) are similar according to the measures listed in the table. Per capita calories consumption of cereals and all food groups are 1,564 and 2,140 calories per person per day, which are approximately 14 calories higher and 32 calories lower than the averages in Round 43, respectively. Poverty is clear in the data: nearly 80 percent of households do not have a toilet, and food expenditures are almost 70 percent of the household budget.

We estimate the regression equation:

$$
\begin{aligned}
calories_{ips} &= \alpha_0 + \alpha_1 local\ open\ defecation_{ps} + \alpha_2 household\ open\ defecation_{ips} \\
&+ \alpha_3 ln\left(MPCE\right)_{ips} + \sum_{j=1}^{J} \omega_j employment_{ips}^j + X_{ips}\vartheta + \varsigma_s + \varepsilon_{ips},
\end{aligned}
\tag{3}
$$

where $i$ is an index for household, $p$ for PSU (i.e. village or urban sub-block), and $s$ for state-region. There are 77 state-regions in NSS Round 38, and state-regions are contiguous districts grouped by geographic features, population densities, and cropping patterns. The key explanatory variable is the percentage of households in the PSU without a latrine (local open defecation). We also control for whether the individual household has its own (or shares) a latrine and for the household's monthly per capita expenditures in logs. Additional covariates ($X_{ips}$) include household's caste, religion, type of flooring, source of drinking water, and type of cooking fuel. Flooring, water source, and cooking fuel proxy for household socioeconomic status and access to public goods.

Importantly, we control for work requirements in two ways. First, we control for demographic-specific work variables. We calculate the share of household members who fall into one of 38

---

[21]In the NSS, the Primary Sampling Unit (PSU) is a village for the rural sector and block for the urban sector. If the population per village/block exceeds one million inhabitants (i.e. 10 lakhs), then the unit is further divided into hamlet groups or sub-blocks.

age-sex-industry categories. There are seven age categories and five industry categories.[22] Second, we control for the household's primary occupation, which are using 3-digit NCO-1968 codes. In NSS 1983, there are 665 groups. The first method accounts for differences across households in demographic composition as well as the type of work across members within the household. The second method models the household as a single unit. To the extent that member-specific employment information might be noisy, the primary household-level occupation may be a preferred indicator. The results of both methods are shown in Table 7. With the inclusion of one or both of these work controls, identification of the effect of local sanitary conditions (captured by $\alpha_1$) comes from variation within state-region, across PSUs, holding constant household wealth and differences in occupation types, industries, or demographic composition.

## 4.2 Effect of sanitation on calories conditional on work

If the association of sanitation on calorie consumption were largely driven by spuriously correlated differences in energy needs for work activity, then once we controlled for the household's primary occupation, there would be no systematic relationship between latrine coverage and calories; in other words, $\alpha_1$ would fall to zero. The results in Table 7 reject this hypothesis and show that energy needs for work requirements do not explain why households in worse disease environments consume more calories per person.

Across the columns of Table 7, we include different combinations of control variables to predict per capita calories consumption of cereals (columns 1-6) or all food groups (column 7-8). Without any control variables, we find that there is a strong positive relationship

---

[22]The seven age categories are: (1) 0-4 years, (2) 5-9 years, (3) 10-14 years, (4) 15-39 years, (5) 40-59 years, (6) 60-99 years, and (7) missing age. The five industry categories are: (1) principal industry in agriculture/mining/construction, (2) principal industry in manufacturing/electricity/trade/business, (3) principal industry in services, (4) principal working status as domestic duties/unemployed/other/in school/too young or too old and subsidiary industry in agriculture/mining/construction/manufacturing/electricity/trade/business, and (5) principal working status as domestic duties/unemployed/other/in school/too young or too old and no subsidiary work reported. For age categories 1-3 and 7, we do not assign industry categories and group the members by age-sex only.

between local latrine coverage and calorie consumption; however, this coefficient is unlikely to represent a causal effect. Once we condition on MPCE, urban residence, and state-region fixed effects, the estimate decreases in magnitude but remains strongly significant. Moreover, it is quantitatively comparable to our earlier estimates from different empirical strategies.

How much of the apparent effect of sanitation on calorie consumption might actually reflect differences related to work? Juxtaposing columns 4 and 5 or columns 4 and 6, we see that there is not a large change in the coefficient estimate whether we control for state-region fixed effects or work variables, as captured by 38 demography-specific industry categories or 665 primary occupational categories. Including all work-relevant characteristics as explanatory variables in the regression (column 6), we still find that local sanitation is a strong and positive predictor of calories consumption. Relative to a person from a village where no one defecates in the open, a person from a village where 80 percent of his or her neighbours openly defecate consumed an additional 100 calories per day, on average. For calories from all food groups in columns 7-8, we confirm that our main result is robust to controls for detailed work controls alongside MPCE and household characteristics. Because these detailed work controls do not importantly change our estimates – either within this data set, or in comparison with the estimates in section 2 and 3 – it is unlikely that changes in work requirements are responsible for our results.

# 5   How much of the calorie decline could the disease environment explain?

Two empirical strategies using three separate datasets find quantitatively similar effects of the disease environment on calorie consumption in India. In light of these estimates, how much of the Indian calorie decline could be statistically accounted for by an improving disease environment? This section uses two complementary methods to estimate the fraction of the gap that can be explained.

First, we apply the regression results of this paper to linearly predict the change in calorie consumption associated with the observed change in the disease environment. Therefore, we calculate

$$percent\ explained = \hat{\beta}\frac{\Delta disease}{\Delta calories}. \tag{4}$$

The disease-calories gradient $\hat{\beta}$ is taken from various estimates from the regression tables in this paper. For completeness, we separately use change in IMR and change in sanitation, and change in cereal and total calorie consumption, all taken from Table 1 of summary statistics.

Results are reported in Panels A and B of Table 8. Both the change in infant mortality and the change in sanitation linearly predict a decrease in calorie consumption that would account for about 20 percent of the gap or more. To the extent that the changes in and effects of sanitation and IMR are independent of one another, the true total percent explained by improvements in the disease environment may be even greater.

Next, Panel C presents results from econometric decomposition analyses. These estimate the fraction of an average difference in an outcome between two groups that can be accounted for by differences in observable characteristics (Spears, 2013). Like any other analysis of observable data, a causal interpretation of a decomposition depends on the nature of the heterogeneity in the explanatory variables; decompositions such as these may overestimate the fraction causally explained if the associations that they use include omitted variable bias.

We use Oaxaca (1973)-Blinder (1973) decompositions that apply a similar linear method to that which is shown in equation 4. Different decomposition methods construct different estimates of the slope $\beta$. For robustness, we use two different estimates of $\hat{\beta}$: the simple regression slope from the pooled data, and an equally weighted average of the slopes estimated from within the two survey rounds, or points in time.

Additionally we use a non-linear decomposition that non-parametrically reweights the sample from the 1980s to match the distribution of exposure to disease in the sample from the 2000s sample (DiNardo et al., 1996).[23] In particular, we divide each sample into 16

---

[23] This approach has been recently used by Geruso (2012) to estimate the fraction of the U.S. mortality

infant mortality "bins" $b$ corresponding to intervals of 5 infant deaths per 1,000 live births. We then compute for each bin and each sample $w_b^{43}$ and $w_b^{61}$, the fraction of the sample in the 43rd and 61st survey rounds that are in bin $b$, using household survey weights. Each observation $i$ in the 43rd round is given a new weight

$$\tilde{\pi}_i = \pi_i \frac{w_{b(i)}^{61}}{w_{b(i)}^{43}}, \tag{5}$$

where $\pi_i$ is the survey weight of household $i$ computed by the NSS and $b(i)$ is the IMR bin to which household $i$ belongs. Finally, we compute a counterfactual calorie consumption mean for the 43rd round, if it had had the same distribution of exposure to IMR as the 61st round, as

$$counterfactual\ calories^{43} = \frac{\sum_i \tilde{\pi}_i calories_i^{43}}{\sum_i \tilde{\pi}_i}. \tag{6}$$

Panel C of Table 8 presents the results of these three decomposition methods, applied to the change between the 43rd and 61st NSS rounds in calorie consumption and cereal calorie consumption. Unlike the decomposition approaches in Panels A and B, these results are not based on causally narrow effect estimates. Nevertheless, changes in the disease environment over the two decades studied can statistically account for a substantial fraction of the decline in calorie consumption over this period. These results are consistent with our earlier observation that – if IMR and sanitation represent at least partially non-overlapping dimensions of the disease environment – "over 20 percent" may be only a lower bound on the part explained. However, no estimate here of the percent explained suggests that the disease environment can account for the entire calorie decline.

gap between black and white people that is due to socioeconomic status and by Spears (2013) to estimate the fraction of the gap in average child height between India and sub-Saharan Africa than can be accounted for by exposure to open defecation.

# 6 Conclusion

Over the past several decades, average calorie consumption in India has declined substantially. In this paper, we have presented and assessed the evidence for one candidate explanation: a gradually improving disease environment. Two complementary empirical strategies applied to different datasets estimate robust and quantitatively comparable effects of the disease environment on average calorie consumption. These estimates suggest that the disease environment could account for at least one-fifth of India's recent calorie decline.

Because India still faces an important burden of preventable infectious disease, these estimates suggest that the Indian economy may suffer a large cost of wasted calorie consumption. Any computation of such a cost is highly approximate. With this strong caveat, taking our linear regression results literally suggests that reducing open defecation rates in India from over 50 percent to zero could reduce per capita calorie needs by about 50 calories per day. If a poor person's calories have a marginal cost of about 0.02 cents apiece (1.25 international dollars of total consumption per day, with one third on food, and 2,000 calories), then eliminating open defecation would save about four dollars per person per year in food consumption, which is about one-tenth of one percentage point of GDP per capita. Again, there are many approximations in this figure: for example, some well-nourished people may feel no effect; different people have differently priced marginal calories; and this computation ignores potential improvements in other dimensions of the disease environment.[24] We include it merely to suggest that effects on calorie needs of the disease environment could add up to an important economic cost.

---

[24]Notably, however, this approximation of annual benefits – which ignores all benefits of improved sanitation other than reduced calorie needs – would save enough money to spend about $175 per latrine to build a latrine for each of the approximately 130 million households that the 2011 census reports defecates in the open, assuming that a latrine lasts at least five years.

# References

Baird, S., J. H. Hicks, M. Kremer, and E. Miguel (2011). Worms at work: Long-run impacts of child health gains. working paper, Berkeley. 1.2

Bhutta, Z. A., J. K. Das, A. Rizvi, M. F. Gaffey, N. Walker, S. Horton, P. Webb, A. Lartey, R. E. Black, and The Lancet Nutrition Interventions Review Group, and the Maternal and Child Nutrition Study Group (2013). Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? *Lancet 371*, 417–440. 1.1

Bleakley, H. (2007). Disease and development: Evidence from hookworm eradication in the American South. *Quarterly Journal of Economics 122*(1), 73–117. 1.2

Blinder, A. S. (1973). Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources 8*, 436–455. 5

Bozzoli, C., A. Deaton, and C. Quintana-Domeque (2009). Adult height and childhood disease. *Demography 46*(4), 647–669. 1.1, 1.3, 2.1

Case, A. and C. Paxson (2008). Stature and status: Height, ability, and labor market outcomes. *Journal of Political Economy 116*(3), 499–532. 19

Checkley, W., G. Buckley, R. H. Gilman, A. M. Assis, R. L. Guerrant, S. S. Morris, K. Mølbak, P. Valentiner-Branth, C. F. Lanata, R. 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. 1.2

Coffey, D. (2013). Open defecation and hemoglobin deficiency in young children. Conference presentation at Delhi School of Economics, Princeton Office of Population Research. 1.2

Coffey, D., A. Chattopadhyay, and R. Gupt (2013). Wealth and the health of children in India: A state-level analysis. working paper, IIPS. 4

Coffey, D., A. Deaton, J. Drèze, D. Spears, and A. Tarozzi (2013). Stunting among children: Facts and implications. *Economic and Political Weekly XLVIII*(34), 68–70. 3.2

Cutler, D. and G. Miller (2005). The role of public health improvements in health advances: The twentieth-century United States. *Demography 42*(1), 1–22. 1.2

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics 30*, 109–126. 2

Deaton, A. (2006). The great escape: A review of robert fogel's *The Escape from Hunger and Premature Death, 1700-2100. Journal of Economic Literature XLIV*, 106–114. 1.1

Deaton, A. (2007). Height, health and development. *Proceedings of the National Academy of the Sciences 104*(33), 13232–13237. 1.1, 1.2

Deaton, A. and J. Drèze (2009). Food and nutrition in India: Facts and interpretations. *Economic and Political Weekly 44*(7), 42–65. 1, 1.1, 2.1, 18, 4

Desai, S., A. Dubey, B. Joshi, M. Sen, A. Shariff, , and R. Vanneman (2009). India human development survey (ihds). computer file, University of Maryland and National Council of Applied Economic Research. 3

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044. 5

Eli, S. and N. Li (2013). Can caloric needs explain three food consumption puzzles? evidence from India. working paper, University of Toronto. 1.1, 4

Fogel, R. W. (1997). New findings on secular trends in nutrition and mortality: Some implications for population theory. *Handbook of Population and Family Economics*. 1.1

Galiani, S., P. Gertler, , and E. Schargrodsky (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy 113*(1), 83–120. 1.2

Geruso, M. (2012). Black-white disparities in life expectancy: How much can the standard ses variables explain? *Demography 49*(2), 553–574. 23

Gopalan, C., B. V. R. Sastri, and S. C. Balasubramanian (1989). *Nutritive Value of Indian Foods.* National Institute of Nutrition, Indian Council of Medical Research. 2.1

Gregory Clark, M. H. and P. H. Lindert (1995). A british food puzzle, 1770-1850. *Economic History Review 48*(2), 215–237. 1.1

Guerrant, R., J. Schorling, J. McAuliffe, and M. de Souza (1992). Diarrhea as a cause and an effect of malnutrition: Diarrhea prevents catch-up growth and malnutrition increases diarrhea frequency and duration. *American Journal of Tropical Medicine and Hygiene 47*(1), 28–35. 1.2

Hatton, T. (2013). How have europeans grown so tall? 1.1, 1.3, 2.1

Humphrey, J. H. (2009). Child undernutrition, tropical enteropathy, toilets, and handwashing. *The Lancet 374*, 1032 – 35. 1.2

Jayachandran, S. and R. Pande (2013). Why are Indian children shorter than African children? working paper, Northwestern. 1.1

Jensen, R. T. and N. H. Miller (2010). A revealed preference approach to measuring hunger and undernutrition. working paper 16555, NBER. 2.1

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

Kosek, M., et al., and the MAL-ED network (2013). Fecal markers of intestinal inflammation and permeability associated with the subsequent acquisition of linear growth deficits in infants. *American Journal of Tropical Medicine and Hygiene 88*(2), 390–396. 1.2

Kov, P., S. Smets, D. Spears, and S. Vyas (2013). Sanitation and stunting in cambodia. working paper, World Bank Water and Sanitation Program. 14

Lin, A., B. F. Arnold, S. Afreen, R. Goto, T. M. N. Huda, R. Haque, R. Raqib, L. Unicomb, T. Ahmed, J. M. C. Jr., and S. P. Luby (2013). Household environmental conditions are associated with enteropathy and impaired growth in rural Bangladesh. *American Journal of Tropical Medicine and Hygiene.* 1.2

Logan, T. D. (2006). Nutrition and well-being in the late nineteenth century. *Journal of Economic History 66*(2). 5

Logan, T. D. (2009). The transformation of hunger: The demand for calories past and

present. *Journal of Economic History 69*(2). 1.1

Menon, P., S. Cyriac, and A. Bamezai (2013). Diet quality, water, and toilets: What role for child undernutrition in india? Conference presentation at Delhi School of Economics, IFPRI. 1.1

Miguel, E. and M. Kremer (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica 72*(1), 159–217. 1.2

Mondal, D., J. Minak, M. Alam, Y. Liu, J. Dai, P. Korpe, L. Liu, R. Haque, and W. A. Petri, Jr (2011). Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clinical Infectious Diseases*. 1.2

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*, 693–709. 5

Preston, S. H. and E. van de Walle (1978). Urban french mortality in the nineteenth century. *Population Studies 32*(2), 275–297. 1.1

Ramalingaswami, V., U. Jonsson, and J. Rohde (1996). Commentary: The Asian enigma. The Progress of Nations. 1.1

Schmidt, W.-P., S. Boisson, B. Genser, M. L. Barreto, K. Baisley, S. Filteau, and S. Cairncross (2011). Weight-for-age z-score as a proxy marker for diarrhoea in epidemiological studies. *Journal of Epidemiology and Community Health 40*(6), 1678–1692. 3.4

Scrimshaw, N. S., C. E. Taylor, and J. E. Gordon (1968). *Interactions of Nutriton and Infection.* Geneva: World Health Organization. 1.2, 6

Smith, M. I., T. Yatsunenko, M. J. Manary, I. Trehan, R. Mkakosya, J. Cheng, A. L. Kau, S. S. Rich, P. Concannon, J. C. Mychalecky, J. Liua, E. Houpt, J. V. Li, E. Holmes, J. Nicholson, D. Knights, L. K. Ursell, R. Knight, and J. I. Gordon (2013). Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science 339*(6119), 548–554. 1

Spears, D. (2012a). working paper, Princeton. 1.3

Spears, D. (2012b). Effects of rural sanitation on infant mortality and human capital: Evidence from a local governance incentive in India. working paper, Princeton University Research Program in Development Studies. 1.2, 2.1

Spears, D. (2013). The nutritional value of toilets: How much international variation in child height can sanitation explain? working paper, Princeton University Research Program in Development Studies. 1, 1.1, 14, 5, 23

Spears, D., A. Ghosh, and O. Cumming (2013). Open defecation and childhood stunting in India: An ecological analysis of new data from 112 districts. *PLoS ONE 8*(9). 18

Stephensen, C. B. (1999). Burden of infection on growth failure. *Journal of Nutrition 129*(2), 534S–538S. 1.2

Subramanyam, M. A., I. Kawachi, L. F. Berkman, and S. V. Subramanian (2011). Is economic growth associated with reduction in child undernutrition in india? *PLOS Medicine.* 4

Tarozzi, A. (2008). Growth reference charts and the nutritional status of Indian children. *Economics & Human Biology 6*(3). 1.1

Watson, T. (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. 1.2
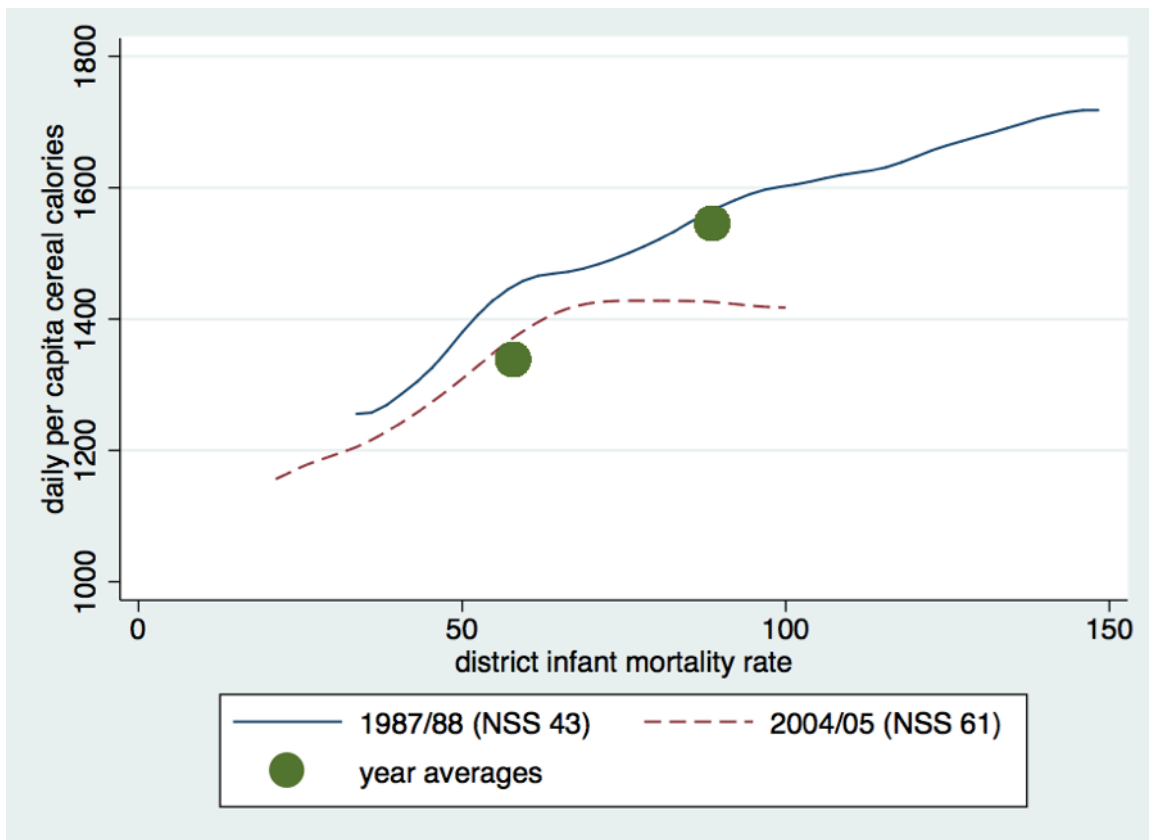
FIGURE 1.  Cereal calorie consumption and infant mortality
(NSS 1987/88 and 2004/05)
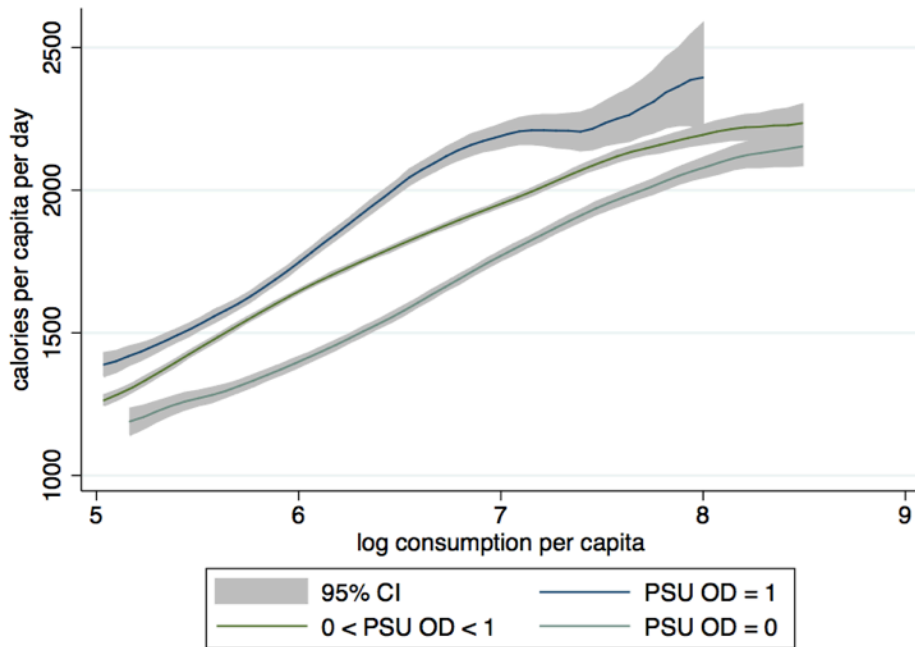
FIGURE 2.    Calorie consumption and monthly per capita
expenditures by local sanitation coverage
(IHDS 2005)
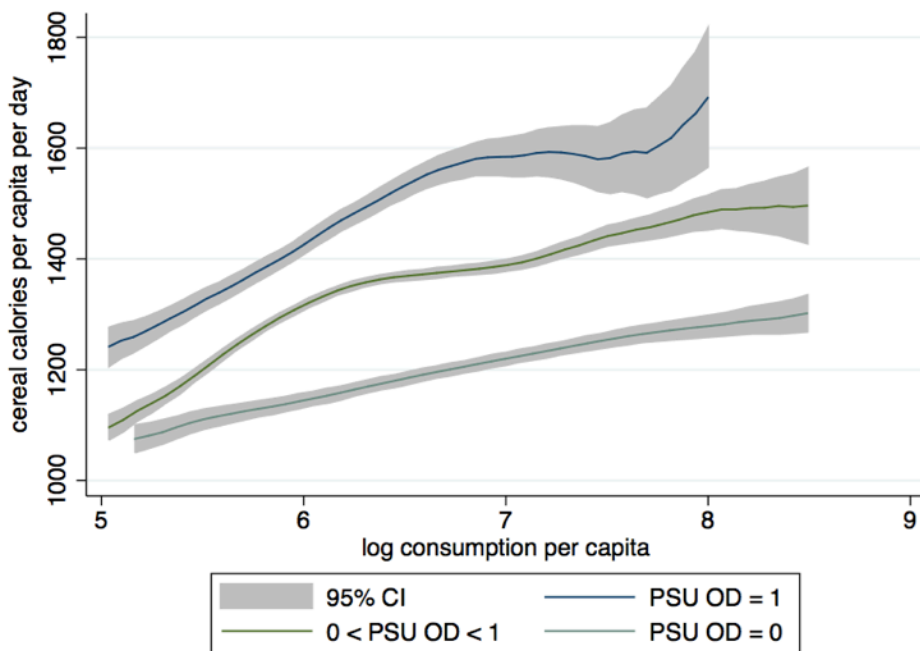


FIGURE 3.    Cereal calorie consumption and monthly per capita
expenditures by local sanitation coverage
(IHDS 2005)
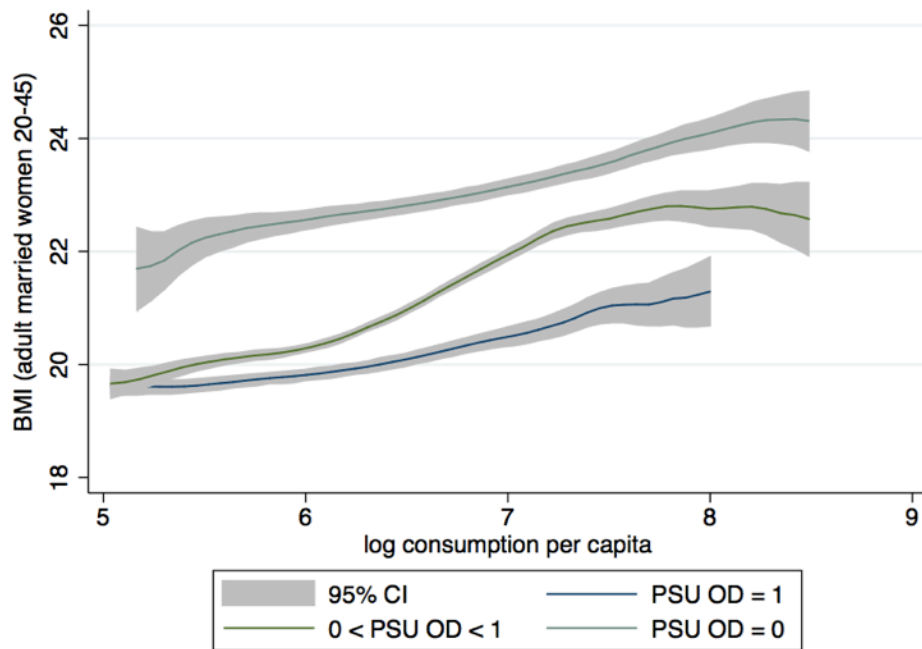
FIGURE 4.    Adult women's BMI and monthly per capita
expenditures by local sanitation coverage
(IHDS 2005)
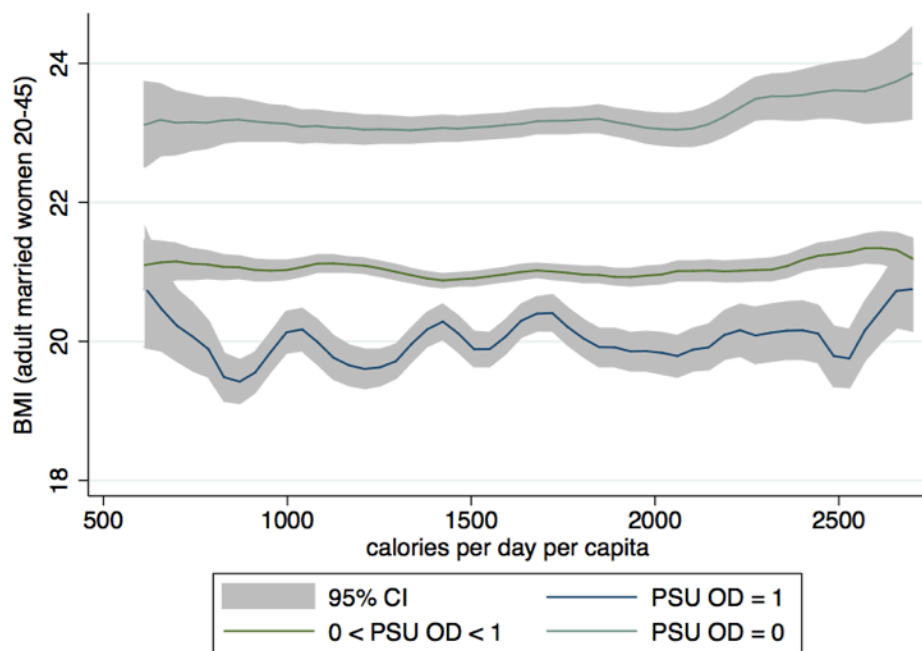


FIGURE 5.    Adult women's BMI and calorie consumption by local
sanitation coverage
(IHDS 2005)

**TABLE 1. Sample Statistics for National Sample Survey Rounds 38/43/61 and India Human Development Survey 2005**

| | NSS Round 43 (1987-88) | NSS Round 61 (2004-05) | IHDS (2005) | NSS Round 38 (1983) |
|---|---|---|---|---|
| Per capita consumption of cereal calories | 1550 | 1337 | 1254 | 1564 |
| Per capita consumption of all calories | 2172 | 2046 | 1787 | 2140 |
| District IMR (per 1000 live births) | 89 | 58 | 45 | . |
| % Households without toilets in district (NSS 43 & 61) or PSU (IHDS & NSS 38) | 80.7 | 60.9 | 58.4 | 79.2 |
| Urban | 0.225 | 0.248 | 0.290 | 0.213 |
| MPCE (in 1987-88 Rupees) | 174 | 214 | 799 | 116 |
| Food expenditure as share of HH budget | 0.723 | 0.666 | 0.555 | 0.735 |
| Self-employed in agriculture (NSS) or Cultivation (IHDS) | 0.331 | 0.297 | 0.246 | 0.362 |
| Agricultural labor | 0.317 | 0.285 | 0.155 | 0.217 |
| Has a TV | 0.024 | 0.391 | 0.486 | . |
| Has a motorized vehicle | 0.042 | 0.151 | 0.161 | . |
| No adult female is literate | 0.606 | 0.415 | 0.483 | . |
| No adult male is literate | 0.308 | 0.197 | 0.236 | . |
| N (households) | 115114 | 113123 | 38227 | 108330 |

Notes: Samples exclude households in the top and bottom 1% of cereal calories distribution. MPCE stands for monthly per capita expenditures. For Round 38, sample only includes households that matched in both in the Consumer Expenditure Survey and Employment-Unemployment
^MPCE in 1983 Rs

**TABLE 2. Larger declines in calorie consumption in districts that experienced bigger reductions in infant mortality**
**(National Sample Survey district panel, 1987/88 to 2004/05)**

| | (1) Cereals | (2) Cereals | (3) Cereals | (4) Cereals | (5) Cereals | (6) Cereals | (7) Cereals | (8) Cereals | (9) All |
|---|---|---|---|---|---|---|---|---|---|
| Dependent variable: | | | | | | | | | |
| **Panel A: Without detailed household demography controls** | | | | | | | | | |
| IMR, district-level | | 4.340*** | 1.313*** | 0.975** | 0.849* | 1.203** | 0.870* | 0.811* | 1.741** |
| | | (0.397) | (0.496) | (0.472) | (0.478) | (0.569) | (0.476) | (0.479) | (0.724) |
| Round 61 | -213.197*** | -79.586*** | -174.209*** | -214.514*** | -242.329*** | 1.192 | -193.798*** | -195.347*** | -75.605*** |
| | (10.236) | (15.548) | (18.544) | (17.885) | (22.271) | (23.223) | (17.895) | (17.994) | (26.243) |
| **Panel B: With detailed household demography controls** | | | | | | | | | |
| IMR, district-level | | 4.661*** | 1.252*** | 0.974** | 0.823* | 1.151** | 0.865* | 0.798* | 1.504** |
| | | (0.401) | (0.482) | (0.463) | (0.467) | (0.559) | (0.466) | (0.468) | (0.686) |
| Round 61 | -225.040*** | -85.491*** | -197.387*** | -224.589*** | -258.015*** | 15.865 | -198.868*** | -201.442*** | -128.781*** |
| | (10.585) | (15.903) | (18.385) | (17.806) | (22.035) | (22.863) | (17.684) | (17.817) | (25.600) |
| District fixed effects | | | X | X | X | X | X | X | X |
| ln(real MPCE) (1987/88 Rs) | | | | X | X | X | X | X | |
| Urban residence | | | | X | X | X | X | X | |
| District average MPCE | | | | | X | | | | |
| State-specific time trends | | | | | | X | | | |
| Household controls | | | | | | | X | X | |
| Occupational controls | | | | | | | | X | |
| Observations | 228,234 | 228,234 | 228,234 | 228,234 | 228,234 | 228,234 | 228,234 | 228,234 | 228,234 |

Notes: Standard errors in parentheses are clustered at the district level (408 clusters). Infant mortality rate is number of children who die before reaching 12 months per 1000 live births. "Cereals" refers to rice, wheat, coarse grains, cereal substitutes, and the products of these items. The dependent variable is per capita calories consumption of either cereals or all food groups and is adjusted for meals eaten away from home. "Round 61" is an indicator that equals 1 if households surveyed in 2004/05. MPCE stands for monthly per capita expenditures and was deflated by the CPIAL (rural households) or CPIIW (urban households) to 1987/88 Rupees. Household controls include indicators for: scheduled caste, scheduled tribe, Muslim, other religion (non-Hindu and non-Muslim), no literate female adult in household, and no literate male adult in household. Occupational controls are composed of 34 categories based on NCO-1968 codes. Households in the top and bottom 1% of [staple or all] calories distribution for each sector-round are dropped from the estimating sample. ***p<0.01,**p<0.05,*p<0.10

**TABLE 3. Higher average calorie consumption in villages/urban sub-blocks with lower sanitation coverage (India Human Development Survey, 2005)**

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| **Panel A: Total Calories** | | | | | | | |
| % Households without toilet in PSU | 2.050*** | 1.663*** | 1.912*** | 1.649*** | 1.486*** | 0.580** | 1.310*** |
| | (0.256) | (0.276) | (0.283) | (0.283) | (0.272) | (0.259) | (0.309) |
| Household open defecation | -76.380*** | 84.173*** | 49.178*** | 45.216*** | 24.665** | 39.60*** | 67.43*** |
| | (13.778) | (12.286) | (12.335) | (12.006) | (11.221) | (11.97) | (12.10) |
| Log of MPCE | | 480.811*** | 409.943*** | 414.592*** | 440.704*** | 448.5*** | 460.4*** |
| | | (9.595) | (10.025) | (10.650) | (11.172) | (9.947) | (9.364) |
| Urban residence | | -289.764*** | -264.927*** | -226.863*** | -220.256*** | -234.6*** | -240.8*** |
| | | (18.786) | (19.045) | (19.459) | (18.673) | (16.50) | (19.40) |
| N (households) | 39673 | 39673 | 39673 | 39673 | 39673 | 39570 | 39673 |
| **Panel B: Cereal Calories** | | | | | | | |
| % Households without toilet in PSU | 2.304*** | 1.103*** | 1.332*** | 1.197*** | 1.060*** | 0.565** | 0.438* |
| | (0.206) | (0.236) | (0.241) | (0.240) | (0.231) | (0.227) | (0.260) |
| Household open defecation | 36.811*** | 96.057*** | 68.131*** | 58.710*** | 31.183*** | 58.41*** | 95.22*** |
| | (10.389) | (10.437) | (10.727) | (10.393) | (9.858) | (10.19) | (10.52) |
| Log of MPCE | | 178.018*** | 116.779*** | 129.576*** | 166.333*** | 172.3*** | 191.7*** |
| | | (7.691) | (8.332) | (8.873) | (9.280) | (8.433) | (7.813) |
| Urban residence | | -266.414*** | -246.726*** | -217.579*** | -205.830*** | -204.5*** | -211.0*** |
| | | (15.467) | (15.500) | (15.640) | (15.072) | (14.43) | (15.96) |
| N (households) | 39673 | 39673 | 39673 | 39673 | 39673 | 39570 | 39673 |
| Demographics | | | X | X | X | X | X |
| Income sources | | | | X | X | X | X |
| Social groups | | | | | X | | |
| Education | | | | | X | | |
| Cereal prices | | | | | | X | |
| Average PSU (local area) MPCE | | | | | | | X |

Notes: Standard errors in parentheses are clustered by PSU (2473 clusters). The dependent variable is per capita calories consumption of either staple (cereals) or all foods. "Household open defecation" means that the household has 'no toilet/open fields'. MPCE stands for monthly per capita consumer expenditure and is measured in 2005 Rupees. "Demographics" includes household size and number of children in households. For "Income sources," the IHDS assigns households to one of 11 categories: Cultivation, Allied Agriculture, Agricultural Labor, Non-Agricultural Labor, Artisan, Petty Trade, Business, Salaried, Profession, Pension/Rent, and Others. "Social groups" classifies households into one of 8 groups: Brahmin, High Caste, OBC, Dalit, Adivasi, Muslim, Sikh/Jain, and Christian. "Education" consists of highest educational attainment of adult 21+ years and of at least one literate adult in household. Households in top and bottom 1% of per capita calories distribution are dropped from the estimating sample. ***p<0.01,**p<0.05,*p<0.10

**TABLE 4. Higher average calorie consumption in villages/urban sub-blocks with higher infant mortality (India Human Development Survey, 2005)**

|  | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A: Cereal Calories** | | | | | | |
| Infant mortality rate, PSU-level | 0.924*** | 0.799*** | 0.648*** | 0.278** | 0.554*** | 0.394** |
|  | (0.143) | (0.141) | (0.142) | (0.124) | (0.147) | (0.131) |
| % Households without toilet in PSU | | | | | 0.878*** | |
|  | | | | | (0.239) | |
| Household open defecation | | | | | 32.308*** | |
|  | | | | | (9.821) | |
| Log of MPCE | | 148.491*** | 156.392*** | 194.493*** | 166.729*** | 205.7*** |
|  | | (7.982) | (9.066) | (8.116) | (9.205) | (8.759) |
| Urban residence | | -331.257*** | -240.138*** | -178.206*** | -207.715*** | -179.5*** |
|  | | (13.077) | (13.643) | (11.055) | (15.054) | (12.45) |
| Price of wheat (Rs/kg) | | | | | | -26.56*** |
|  | | | | | | (2.552) |
| Price of rice (Rs/kg) | | | | | | -28.30*** |
|  | | | | | | (1.574) |
| N (households) | 39652 | 39652 | 39652 | 39652 | 39652 | 39652 |
| **Panel B: Total Calories** | | | | | | |
| Infant mortality rate, PSU-level | 0.768*** | 1.017*** | 0.832*** | 0.349*** | 0.711*** | 0.426** |
|  | (0.141) | (0.152) | (0.152) | (0.130) | (0.155) | (0.134) |
| % Households without toilet in PSU | | | | | 1.249*** | |
|  | | | | | (0.280) | |
| Household open defecation | | | | | 26.357** | |
|  | | | | | (11.217) | |
| Log of MPCE | | 448.985*** | 429.049*** | 443.311*** | 441.223*** | 468.9*** |
|  | | (9.659) | (10.955) | (9.741) | (11.104) | (10.66) |
| Urban residence | | -369.829*** | -265.138*** | -201.027*** | -223.038*** | -201.5*** |
|  | | (15.828) | (17.052) | (13.252) | (18.525) | (14.69) |
| Price of wheat (Rs/kg) | | | | | | -61.20*** |
|  | | | | | | (2.971) |
| Price of rice (Rs/kg) | | | | | | -18.90*** |
|  | | | | | | (1.957) |
| N (households) | 39652 | 39652 | 39652 | 39652 | 39652 | 39652 |
| Household controls | | | X | X | X | X |
| Extended household demography | | | | | | X |
| State fixed effects | | | | X | | |

Notes: Standard errors in parentheses are clustered by PSU (2470 clusters). The dependent variable is per capita calories consumption of either staple (cereals) or all foods. Infant mortality rates are computed from IHDS. "Household open defecation" means that the household has 'no toilet/open fields'. MPCE stands for monthly per capita consumer expenditure and is measured in 2005 Rupees. Household controls include household size, number of children in household, social group, education, and main income source. Extended household demography includes indicator controls for each number of persons, children, teens, married men, and married women. Households in top and bottom 1% of per capita calories distribution are dropped from the estimating sample. ***p<0.01,**p<0.05,*p<0.10

**Table 5. People who live near children with diarrhea eat more calories, IHDS.**

| | (1) | (2) | (3) |
|---|---|---|---|
| | per capita daily calorie consumption | | |
| Diarrhoea local fraction | 348.2*** | 366.8*** | 346.0*** |
| | (68.29) | (71.62) | (70.83) |
| Fever local fraction | 56.99 | 80.91 | 69.36 |
| | (60.48) | (60.89) | (58.72) |
| Cough local fraction | 8.104 | -18.30 | 10.33 |
| | (62.50) | (63.31) | (60.98) |
| Urban | -378.8*** | -343.0*** | -292.4*** |
| | (15.54) | (15.70) | (17.13) |
| Log of per capita household consumption | 435.5*** | 366.8*** | 432.7*** |
| | (8.998) | (9.222) | (8.618) |
| Log of PSU average per capita household consumption | | | -165.2*** |
| | | | (20.13) |
| | | | |
| Household structure controls | | X | X |
| | | | |
| N (households) | 39,654 | 39,654 | 39,654 |

Notes: Standard errors clustered by survey PSU.

Disease "local fractions" are the fraction of children under 5 in the survey PSU who were reported by their mother to have had that symptom within the last month.

***p<0.01, **p<0.05, *p<0.10

**TABLE 6. Lower average BMI among women who live in villages/urban sub-blocks with lower sanitation coverage (India Human Development Survey, 2005)**

|  | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Panel A: BMI regressed on sanitation and cereal calories |  |  |  |  |  |  |  |
| % Households without toilet in PSU | -0.02993*** | -0.02954*** | -0.01903*** | -0.01086*** | -0.01118*** | -0.00996*** | -0.00968*** |
|  | (0.00096) | (0.00097) | (0.00124) | (0.00179) | (0.00170) | (0.00162) | (0.00160) |
| Per capita calories consumption: Cereals |  | -0.00015** | -0.00032*** | -0.00029*** | -0.00024*** | -0.00002 | -0.00005 |
|  |  | (0.00007) | (0.00007) | (0.00007) | (0.00006) | (0.00007) | (0.00006) |
| Log of MPCE |  |  | 1.01412*** | 0.91590*** | 0.79735*** | 0.28203*** | 0.35188*** |
|  |  |  | (0.05844) | (0.05645) | (0.05613) | (0.06330) | (0.06407) |
| Urban residence |  |  | 0.44273*** | 0.47965*** | 0.39016 | 0.07605 | 0.09114 |
|  |  |  | (0.08661) | (0.08662) | (0.31245) | (0.30412) | (0.30087) |
| Household open defecation |  |  |  | -0.86832*** | -0.78528*** | -0.34951*** | -0.35241*** |
|  |  |  |  | (0.12580) | (0.11550) | (0.10703) | (0.10778) |
| Height, in centimeters |  |  |  |  |  |  | -0.07480*** |
|  |  |  |  |  |  |  | (0.00593) |
|  |  |  |  |  |  |  |  |
| N (PSU) | 2452 | 2452 | 2452 | 2452 | 2452 | 2452 | 2452 |
| N (adult women) | 27135 | 27135 | 27135 | 27135 | 27135 | 27135 | 27123 |
|  |  |  |  |  |  |  |  |
| Panel B: BMI regressed on sanitation and total calories |  |  |  |  |  |  |  |
| % Households without toilet in PSU | -0.02993*** | -0.03017*** | -0.01929*** | -0.01095*** | -0.01128*** | -0.00998*** | -0.00971*** |
|  | (0.00096) | (0.00096) | (0.00124) | (0.00180) | (0.00170) | (0.00162) | (0.00160) |
| Per capita calories consumption: Total |  | 0.00020*** | -0.00016*** | -0.00014** | -0.00011** | 0.00000 | -0.00001 |
|  |  | (0.00005) | (0.00006) | (0.00006) | (0.00005) | (0.00005) | (0.00005) |
| Log of MPCE |  |  | 1.03137*** | 0.92900*** | 0.80109*** | 0.27624*** | 0.34249*** |
|  |  |  | (0.06338) | (0.05979) | (0.05835) | (0.06461) | (0.06520) |
| Urban residence |  |  | 0.48228*** | 0.51623*** | 0.42054 | 0.08030 | 0.09875 |
|  |  |  | (0.08606) | (0.08622) | (0.31285) | (0.30401) | (0.30089) |
| Household open defecation |  |  |  | -0.88278*** | -0.79561*** | -0.34843*** | -0.35013*** |
|  |  |  |  | (0.12740) | (0.11659) | (0.10652) | (0.10730) |
| Height, in centimeters |  |  |  |  |  |  | -0.07470*** |
|  |  |  |  |  |  |  | (0.00592) |
|  |  |  |  |  |  |  |  |
| N (PSU) | 2452 | 2452 | 2452 | 2452 | 2452 | 2452 | 2452 |
| N (adult women) | 27135 | 27135 | 27135 | 27135 | 27135 | 27135 | 27123 |
|  |  |  |  |  |  |  |  |
| Income source |  |  |  |  | X | X | X |
| Household controls |  |  |  |  |  | X | X |

Notes: Standard errors in parentheses are clustered at the PSU level (2452 clusters). PSU stands for "primary sampling unit," which is typically a village or city sub-block. The dependent variable is BMI or body mass index. Sample primarily consists of married women between 18-49 years old. Household controls include main income source, household assets index, social group, household size, number of children in household, and any literate adult in household. ***$p<0.01$,**$p<0.05$,*$p<0.10$

**TABLE 7. Higher average calorie consumption in villages/urban sub-blocks with lower sanitation coverage, conditional on detailed demography-specific household employment information (National Sample Survey Round 38, 1983)**

| Dependent variable: | (1) Cereals | (2) Cereals | (3) Cereals | (4) Cereals | (5) Cereals | (6) Cereals | (7) All | (8) All |
|---|---|---|---|---|---|---|---|---|
| % Households without toilet in PSU | 5.431*** | 3.241*** | 2.405*** | 1.266*** | 1.380*** | 1.206*** | 1.639*** | 1.186*** |
| | (0.128) | (0.271) | (0.258) | (0.240) | (0.242) | (0.235) | (0.275) | (0.256) |
| Log of MPCE | | 403.525*** | 475.928*** | 555.460*** | 509.428*** | 521.768*** | 1,046.001*** | 1,005.566*** |
| | | (8.083) | (8.107) | (7.193) | (7.682) | (7.493) | (7.531) | (8.007) |
| Urban residence | | -341.163*** | -271.747*** | -167.560*** | -175.736*** | -149.592*** | -296.680*** | -165.460*** |
| | | (17.833) | (17.634) | (15.073) | (15.765) | (14.812) | (18.684) | (16.561) |
| | | | | | | | | |
| Household controls | | | X | X | X | X | X | X |
| State-region fixed effects | | | | X | X | X | X | X |
| Demography-specific employment controls | | | | X | | X | | X |
| Household primary occupation controls | | | | | X | X | | X |
| | | | | | | | | |
| N (PSUs) | 2991 | 2991 | 2991 | 2991 | 2991 | 2991 | 2991 | 2991 |
| N (households) | 108,330 | 108,330 | 108,330 | 108,330 | 108,330 | 108,330 | 108,098 | 108,098 |
| R-squared | 0.064 | 0.188 | 0.225 | 0.356 | 0.370 | 0.386 | 0.518 | 0.562 |

Notes: Standard errors in parentheses are clustered by PSU (2991 clusters). PSU stands for "primary sampling unit," which is typically a village or city sub-block. "Cereals" includes rice, wheat, coarse grains, cereal substitutes, and the products of these items. The dependent variable is per capita calories consumption of either cereals or all food groups and is adjusted for meals eaten away from home. Household controls include: scheduled tribe, scheduled caste, Muslim, other religion (non-Hindu and non-Muslim), mud flooring, open water source, and dirty cooking fuels. Demography-specific employment controls are the percent of household members who fall into sex-age-industry (based on principal industry division for individuals 15+ years old) category. Household occupation controls include indicators for 3-digit NCO-1968 codes (665 groups), indicator for agricultural industry, and indicators for crops cultivated. MPCE stands for "monthly per capita expenditures," which is measured in 1983 Rupees. There were 77 NSS state-regions at the time of survey; state-regions are contiguous districts grouped by geographical features, population densities, and cropping patterns. Households in top and bottom 1% of per capita calories distribution were trimmed before estimation. ***p<0.01, **p<0.05, *p<0.10

**TABLE 8: How much of the calorie decline can an improving disease environment account for?**

| Table | Column | Outcome | Slope | Percent Explained |
|-------|--------|---------|-------|-------------------|
| PANEL A: IMR gradient | | | | |
| 2 | 9 | all | 1.741 | 43% |
| 2 | 9* | all | 1.504 | 37% |
| 2 | 3 | cereal | 1.313 | 32% |
| 2 | 3* | cereal | 1.252 | 31% |
| 2 | 6 | cereal | 1.203 | 30% |
| 2 | 6* | cereal | 1.151 | 28% |
| 4 | 2 | all | 1.017 | 25% |
| 2 | 4* | cereal | 0.974 | 24% |
| 2 | 5* | cereal | 0.823 | 20% |
| 4 | 2 | cereal | 0.799 | 20% |
| 4 | 6 | all | 0.426 | 10% |
| 4 | 6 | cereal | 0.394 | 10% |
| *median percent* | | | | *27%* |
| PANEL B: Sanitation gradient | | | | |
| 6 | 3 | cereal | 1.694 | 27% |
| 6 | 7 | all | 1.639 | 26% |
| 3 | 5 | all | 1.486 | 23% |
| 6 | 6 | cereal | 1.206 | 19% |
| 6 | 8 | all | 1.186 | 19% |
| 3 | 5 | cereal | 1.060 | 17% |
| *median percent* | | | | *21%* |
| PANEL C: Decomposition of calorie change due to IMR, NSS panel | | | | |
| Oaxaca-Blinder, equal weight | | cereal | | 82% |
| Oaxaca-Blinder, pooled | | cereal | | 87% |
| Non-parametric reweighting | | cereal | | 77% |
| Oaxaca-Blinder, equal weight | | all | | 23% |
| Oaxaca-Blinder, pooled | | all | | 51% |
| Non-parametric reweighting | | all | | 44% |

Note: As in Table 2, decomposition analysis in Panel C omits the top and bottom 1 percent of households.

*Extensive demography controls included

**Appendix table 1. Districts with larger declines in IMR saw bigger decreases in calorie consumption -- district-level analysis (National Sample Survey, 1987/88 to 2004/05)**

| Outcome (calorie type): | Cereals | Cereals | All | All |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Infant mortality rate | 1.205** | 1.002 | 1.953** | 0.927 |
| | (0.608) | (0.741) | (0.836) | (0.904) |
| Round 61 (2004/05) | -181.433*** | -273.311*** | -80.970*** | -255.631*** |
| | (22.255) | (43.999) | (28.828) | (51.903) |
| Log of real MPCE, in Rs 1987/88 | | 294.443*** | | 742.571*** |
| | | (58.177) | | (63.260) |
| Urban | | -312.378* | | -256.024 |
| | | (183.257) | | (214.767) |
| F-statistic: IMR = 1 | 0.114 | 0.000 | 1.301 | 0.006 |
| P-value | 0.736 | 0.998 | 0.255 | 0.936 |
| N (districts) | 390 | 390 | 390 | 390 |
| N (district-years, i.e. observations) | 780 | 780 | 780 | 780 |

Notes: Heteroskedasticity-robust standard errors are shown in parentheses. Sample consists of districts; districts in the bottom and top 2% of calorie distribution were excluded from the sample. All regressions include district fixed effects. MPCE stands for "monthly per capita expenditures," which is adjusted for inflation using the CPIAL for rural households and CPIIW for urban households. Controls for share of district population that is Muslim, non-Hindu & non-Muslim, scheduled caste, scheduled tribe, and primary household occupation were added to regression models in columns (2) and (4). ***p<0.01,**p<0.05,*p<0.10