This paper is a draft submission to the

WIDER Development Conference

Human capital and growth

6-7 June 2016 Helsinki, Finland

Sanitation and child health in India

Britta Augsburg (IFS) and Paul Rodríguez-Lesmes (UCL)*

May 23, 2016

Abstract

Our study contributes to the understanding of key drivers of stunted growth, a factor widely recognized as major impediment to human capital development. Specifically, we examine the effects of sanitation coverage and usage on child height for age in a semi-urban setting in Northern India. Our study is the first to address the endogeneity of sanitation coverage exploiting variation in raw material construction prices. Estimating an IV model, we find that sanitation coverage plays a significant and positive role in height growth during the first years of life and that this causal relationship holds particularly for girls. Our findings suggest that a policy that aims to increase sanitation coverage in a context such as the one studied here, is not only effective in reducing child stunting but also implicitly targets girls.

JEL I12

Keywords: child health; sanitation coverage; open defecation; India.

*We thank the ESRC-DfID (grant number ES/J009253/1) as well as FINISH Society for financial assistance and Sanghmitra Gautam for allowing access to the price data. The paper further greatly benefitted from anonymous referee input. All errors that remain our own. Author affiliations and contacts: Augsburg (Institute for Fiscal Studies, britta_a@ifs.org.uk); Rodríguez-Lesmes (University College London, p.lesmes.11@ucl.ac.uk).

1 Introduction

The failure to reach linear growth potential early in life has been widely recognized as a major impediment to human capital development. There is increasing evidence that growth failure is correlated, likely in a causal way, with lower educational and labour market attainments as well as higher risks of health impediments such as diabetes, heart diseases and strokes (Spears and Lamba (2016), Adair et al. (2013), Victora et al. (2010), Behrman et al. (2009), Hoddinott et al. (2008, 2013), Maluccio et al. (2009), Grantham-McGregor et al. (2007)). Rates of stunting, the general term for a child being short for its age, have been reducing over recent years, but 159 million children around the world are still estimated to be affected, more than half of these living in Asia (de Onis et al. (2015)).

While a growing body of literature is contributing to our understanding of the consequences of stunting, knowledge is still limited with respect to the key drivers of low height for age. It is generally understood that inadequate diet and diseases are important immediate causes of stunting (Black et al. (2008), Smith and Haddad (2014), Bozzoli et al. (2009)). However, dealing with the endogeneity of these inputs remains a challenge in the literature (Deaton (2007)), and with that in informed policy decision-making. Understanding these key drivers of stunting and identifying interventions that tackly them remain therefore an important point on the development agenda.

In this study we focus on the role of diseases, by analysing the impact of an improved disease environment - specifically an increase in the use of sanitation technology - on the growth trajectory of children under the age of 5 years. We address the endogeneity of the disease environment using an instrumental variable approach. Diseases have been linked to stunting[1] (Checkley et al. (2008)) but have also shown direct associations with short (Nokes et al. (1992), Nokes et al. (1998), Walker et al. (2011)) and long-term effects on human capital (Almond and Currie (2011), Bozzoli et al. (2009)). Understanding the potential of improving the disease environment that children live in is hence of direct policy relevance.

The WHO identifies diarrhoea as the disease of primary concern: it is said to be the leading cause of child mortality and morbidity in the world, killing an estimated 760,000 children every year (WHO, 2013). Most of these diarrhoea cases are believed to be due to contamination of the environment. Eighty percent are seen to be linked to unsafe water, inadequate sanitation or insufficient hygiene, as estimated in a 2008 report by the WHO (Pruess-Uestuen et al., 2008). More recent thinking associates lack of sanitation

[1]The WHO describes stunted growth (low height-for-age) as “a process of failure to reach linear growth potential as a result of suboptimal health and/or nutritional conditions”.

additionally with a gut disorder called environmental enteropathy, which in turn has been linked to impaired growth (Lunn et al. (1991), Campbell et al. (2003), Lin et al. (2013)). In fact, environmental enteropathy is now seen by many as a much larger contributor to stunting than diarrhoea (Mbuya and Humphrey (2016)). In either case, an important focus of global stunting reduction efforts has therefore been effective and affordable interventions that aim to improve the disease environment by tackling access to safe drinking water, adequate sanitation and hygiene behaviour (Mbuya and Humphrey (2016)).

While intervention that address *simultaneously* water and sanitation environment have been shown to positively affect child health (WHO (2008); Duflo et al. (2015b); Pruess-Uestuen et al. (2008); Checkley et al. (2008); Merchant et al. (2003)), their relative effectiveness, and the role of improved household sanitation in particular has been proven harder to manifest.[2] While a number of studies have been able to rigorously show positive impacts of improved household sanitation (see for example Spears (2012); Kumar and Vollmer (2013); Pickering et al. (2015)), recent randomized controlled trials showed no health impacts (Clasen et al. (2014), Patil et al. (2014), Pattanayak et al. (2007))[3].

Our study similarly considers the impact of sanitation on child health. However, instead of focusing on individual household sanitation ownership, we concentrate on sanitation *coverage*, in the sense of the percentage of people owning a toilet in a community. The main motivation lies in the understanding that individual household sanitation is unlikely to live up to promises in improving health statuses when neighbours are still contaminating the environment, i.e. that externalities are at play.[4] The percentage of households in a community that own a toilet, rather than private ownership, is hence hypothesised to be the more relevant unit when trying to understand the potential of sanitation in improving child health.

A number of researchers have turned their attention to linking sanitation coverage to child health. Most relevant in the context of our study is the working paper of Spears (2012) and Hammer (2013). Exploiting the staggered introduction of India's Total Sanitation Campaign to conduct a difference-in-differences analysis, and the eligibility rules for a village sanitation prize to conduct a regression discontinuity analysis, Spears

[2] **?** in fact suggests that clean water and sanitation are substitues in the context of the Philippines, having large unintended consequences on sanitation uptake.

[3] Hypothesised reasons for this are manifold and mostly link to technological, financial and behavioural challenges. Such limited understanding and evidence of effectiveness is particularly problematic for an investment that faces significant challenges, including lack of appropriate technology, local capacity and most importantly, lack of financial resources - and with that is easily discouraged.

[4] This is another hypothesis why recent RCT trials have failed to demonstrate health impacts of improved sanitation, namely that the coverage increased achieved was not significant enough.

(2012) show that infant mortality decreased by 4 per 1,000 and children's height increased by 0.2 standard deviations at the mean program intensity.[5] Hammer (2013), concentrating on a special experimental effort in the same area of India, find through an RCT that the program was associated with a 0.3 to 0.4 standard deviation increase in children's height-for-age z -scores.

We contribute to this active and growing literature in three ways: For one, we explore the impact of sanitation coverage on child health in a (semi-)urban context by considering Indian households residing in slums and peripheral villages. Slum populations are an important group in this context since a distinctive characteristic of their environment is very crowded conditions, implying more important sanitation externality links, while at the same time experiencing on average worse access to sanitation. A sequence of ongoing work by the R.I.C.E. Institute suggests that it is such population *density* children grow-up in that matters most for sanitation exposure and hence impacts on health (see for example Hathi et al. (2014); Spears (2014); Vyas et al. (2014); Coffey (2013)). A second reason why this population is particularly relevant is its fast growth. **?** estimates that 40% of the world's urban expansion is taking place in slums. At the same time, cities are struggling to keep pace with necessary infrastructure investment, leading to a phenomena referred to as "urbanization of poverty" - partly driven by the externalities of inadequate sanitation.

The second contribution of our study is that it identifies the marginal effect of sanitation coverage on children's health by exploiting village level variation of sanitation investment prices, which - as an economic model we present highlights - determine the marginal cost of this investment and hence induce exogenous variation in the sanitation environment. We find that a ten percentage point increase in sanitation coverage translates into an approximately 0.7cm increase in height at age four.

Finally, our third contribution is our consideration of differential effects by gender of the child. Our findings suggest that girls benefit more from an improved sanitation environment than boys, an association that has also been shown in the context of rural

[5] Other work in progress includes Geruso and Spears (2014), which uses the fraction of Muslims in a village as an instrument; Gertler et al. (2014) also use an instrument in estimating the impact of open defecation rates on child health (measured by child height). They exploit random allocation of sanitation intervention in their data set. However, it is unlikely that the interventions impacted child health only through reducing open defecation rates, given that intervention activities included for example hygiene behaviour campaigns. Their suggested impact of a one standard deviation reduction in their constructed open defecation index would lead to an average increase of standard deviation in children's height. Finally, Andres et al. (2014), use a simple cross-sectional approach, not attempting to account for endogeneity in their variables of interest, finding that 'a 47 percent reduction in diarrhoea prevalence between children living in a household without access to improved sanitation in a village without coverage of improved sanitation and children living in a household with access to improved sanitation in a village with complete coverage'.

Ecuador (Fuller et al. (2016)). This finding implies that sanitation investments can be used as a strategy to implicityly target girls. Such strategies can be of particular importance in a country like India, where research has shown that boys receive higher parental investment (Barcellos et al. (2014)).

The rest of the paper proceeds as follows: We will discuss the methodology we apply in Section 3, followed by an exposition of the data and study context in Section 2. Our results are presented and discussed in Section 4. Section 5 concludes

2 Data and context

The context of our study are households residing in slums and peripheral villages of the city Gwalior. Gwalior is a historical and major city in the state of Madhya Pradesh, India, with an estimated slum population of one fourth of its citizens (Aggarwal and Kumar (2008)). This puts Gwalior above the country average of about 17% of urban households living in slums according to the 2011 slum census. We argue that this is an important population to study since on the one hand, they typically live in very crowded conditions, implying more important sanitation externality links, while at the same time experiencing on average worse access to sanitation than the already low national average. The 2008-09 National Sample Survey Organisation (NSSO, 2010) estimates that 81 per cent of slum-dwellers in India have inadequate access to sanitation, which compares to national urban sanitation coverage rates of 26% in 2011.

At the same time, Madhya Pradesh is amongst states experiencing the worst rates of underweight and stunting for children. A nationwide survey, the Rapid Survey on Children (RSoC), conducted in 2013-14 by the Ministry of Women and Child Development in cooperation with UNICEF, revealed that a staggering 44.7 percent of children under 5 years of age were stunted (18.5% severely stunted) in Madhya Pradesh compared to a national average of 38.7%.

The data we use in this study was collected with the intention of evaluating a sanitation intervention, which focused primarily on an increased uptake and use of private household sanitation. The evaluation design allocated 39 slums and 17 peripheral, semi-urban communities (henceforth we will refer to them jointly as communities or clusters) to either the sanitation intervention or a control group. The baseline (BL) survey was conducted between February and April 2010, and the follow-up (FU) survey between March and December 2013. In total, 1,982 households (HHs) were interviewed at BL, covering 11,032 individuals. These households were a representative sample of the community at that time. For the FU survey 2,020 HHs were interviewed, covering 12,360 individuals. 1,816 of these 2,020 HHs are in both BL and FU, the remaining were included as a replacement sample. The attrition of panel households is hence 8%.[6]

Our observations in this analysis are children that are 5 years or younger at baseline (359 children) and follow-up (605 children), providing us with a sample of 964 child observations. [7]

Table 1 provides information on key characteristics of our sample children. The information under 'BASIC' are characteristics for all children in the communities we analyse in this study. Statistics under 'MAIN' refer to our sample from the regression analysis. We lose about 10% of children in the regression analysis due to some missing characteristics, which can be different variables for different children. While missing data on anthropometrics is significantly correlated to some characteristics, overall, our reduced sample does not seem to statistically differ. This can be seen the close comparability of the two sets of characteristics, as indicated by the t-tests in the last column of the table. We are therefore not concerned that the loss of sample introduces important bias in the analysis.

Table 1: Descriptive Statistics Children

	BASIC		MAIN		
	Mean	SD	Mean	SD	TTEST
Age in months	34.5	18.2	34.8	18.2	-1.55
Female	48.1%		48.4%		-0.45
Weight-for-age z-score	-1.7	1.6	-1.7	1.6	0.05
Length/height-for-age z-score	-1.6	2.2	-1.6	2.2	-0.81
Weight-for-length/height z-score	-1.1	1.7	-1.1	1.7	0.36
BMI-for-age z-score	-0.9	1.8	-0.9	1.8	0.09
Total Children Round 1	359		332		
Total Children Round 2	605		532		
Total Children	964		864		

Notes: Own calculations based on FINISH sanitation household data for Gwalior.

BASIC: For which there is information about children and main woman age, height, gender, and that live in a village where price data was collected.

MAIN: Same sample as the main regressions.

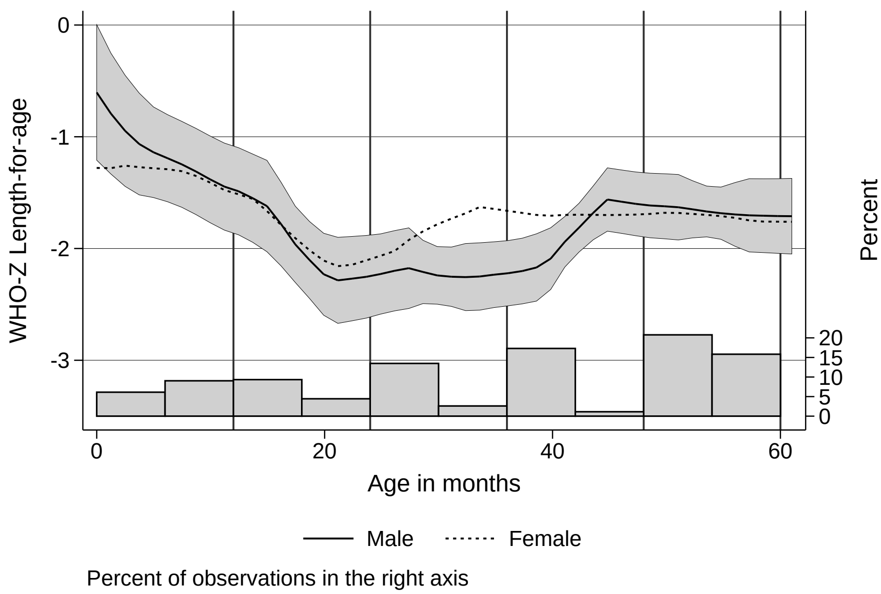
TTEST: t-statistic from a mean-comparison test between included and excluded observations.

[6] Details on the intervention, evaluation design and problems encountered in identifying impacts are outlined in the project's endline report available from the authors upon request.

[7] There are 1,915 children under the age of five in 56 communities. Once we consider only those for which there is price data, we have 1,342 observations in 43 communities. The number is reduced to 964 if we restrict it to valid measurements of date of birth, gender and main woman age and height. We note that this drop from 1,342 to 964 observations seems to be characterised by non-random item non-response in our anthropometrics data. Specifically, as can be seen in Appendix Table **??** we are less likely to have this data for younger children, which is likely to be related to the difficulty of measuring infants. We also see a significant negative correlation with the size of the household, but no correlation with for example incomeor religion religion. While this should be kept in mind, it is difficult to predict how it would affect our findings.

We can see from the Table that the average age of our sample children is 35 months, and slightly less than half are female. The average child lies below the reference population for both weight for age and height for age. Average height for age z-scores, our outcome variable, is -1.6, indicating that the average child is stunted with respect to the reference population. If our children were on track with respect to growth given their age, the average expected value would be zero. Figure 1 shows the WHO height for age z-score for our sample children. It can be seen that the children are already slightly short for their age just after birth and that the z-score reduces particularly in the first two years of life, after which children seem to catch up slightly again, staying still far from the standard population though. This trend is similar for boys (solid line) and for girls (dotted line).

Figure 1: Z-LEN by Gender



The primary survey data further includes detailed information on the households socio-economic status which will be important to account for in our analysis under the presumption that they are correlated with omitted inputs and will alleviate omitted variables bias. We show these in Table 2.

About one fourth of sample households are Muslim, remaining are Hindu and only for a very small percentage is the religion unknown[8]. In terms of the caste, almost 20% of sample households belong to the forward caste, almost 30% to scheduled tribes or schedule castes, remaining are from (minority) backward caste.

[8]We include this 'unknown' variable in our analysis so not to loose these observations while at the same time being able to account for the religion which has been shown an important determinant of sanitation behaviour (Geruso and Spears (2014)).

Our sample households have on average 6-7 household members of which half are male. Putting this number in context of the households' average annual income of US$2,000 (INR 70,000), it becomes clear that these households live much below the internationally used poverty line of US$1.25 per person per day. Not surprisingly then do we find that the total consumption expenditures exceed the household income.[9] Most households do however live in a dwelling of strong or semi-strong structure, reflecting that the study slums are registered.

Table 2: Descriptive Statistics Households

	BASIC		MAIN	
	Mean	SD	Mean	SD
Social background				
Religion: Muslim	22.3%		24.1%	
Caste: Forward caste	18.5%		18.7%	
Caste: Minority backward caste	5.2%		5.7%	
Caste: Scheduled caste or tribe	28.6%		28.8%	
HH Characteristics				
Number of HH members	6.6	2.5	6.5	2.4
Number of children under 5	1.5	0.7	1.5	0.7
Number of male HH members	3.3	1.6	3.2	1.6
Any household shock last 12 months	9.8%		8.2%	
Income†	70.1	47.6	69.0	47.1
Consumption Expenditures†	101.8	81.8	97.6	76.2
Type of dwelling: strong	59.2%		56.7%	
Main woman characteristics				
Education: no formal	56.3%		56.7%	
Education: 1-5 yrs	14.2%		13.9%	
Education: 6-8 yrs	16.2%		16.8%	
Education: 9 yrs +	13.2%		12.5%	
Age (Yrs)	31.5	10.1	31.4	10.1
Height (cm)	149.6	6.7	149.6	6.6
Sanitation and Hygiene				
Owns a toilet	48.8%		48.1%	
Uses a toilet	47.2%		47.0%	
Total Households	299		278	
Households Round 1	267		248	
Households Round 2	440		383	

Notes: Own calculations based on FINISH sanitation household data for Gwalior. † Monetary values are in Indian Rupees of 2013: R1 values where adjusted by a factor of 1.32. It was calculated based on national level figures for 2011, 2012 and 2013.

BASIC: For which there is information about children and main woman age, height, gender, and that live in a village where price data was collected.

MAIN: Same sample as the main regressions.

[9]This is partly driven by the fact that consumption expenditures include value of home produced and traded food.

Table 2 also shows characteristics of the main woman of the household, often the mother of the child's mother. She is on average 30 years old and does not have any education. Only 13% have completed 9 years of schooling or more. They are on average 1,50m tall.

Finally, we see at the bottom of the Table that almost half (48%) of households with children 5 years or younger own a toilet, which they also use. This toilet ownership percentage of sample households with children under the age of five years is comparable to the community level averages, which are displayed in Table 3. At the time of the first data collection round, on average 41% of community members owned a toilet, which increased to almost 60% almost three years later, these average to 50%.

Table 3: Communities Sanitation Coverage

	BASIC		MAIN	
	Mean	SD	Mean	SD
Round 1				
% of HHs that own a toilet	41.7	32.8	40.4	34.0
% of HHs their members use a toilet	44.5	35.3	43.3	36.7
Round 2				
% of HHs that own a toilet	59.0	27.7	59.1	28.7
% of HHs their members use a toilet	56.6	30.4	56.7	31.4
Both Rounds				
% of HHs that own a toilet	50.9	31.3	50.6	32.4
% of HHs their members use a toilet	51.0	33.1	50.6	34.3
Villages Round 1	38		33	
Villages Round 2	43		40	

Notes: Own calculations based on FINISH sanitation household data for Gwalior.

BASIC: For which there is information about children and main woman age, height, gender, and that live in a village where price data was collected.

MAIN: Same sample as the main regressions.

Table 3 also shows descriptive stats for our main variable of interest, which we will discuss in more detail in the next. We define it as the percentage of households in the village that child i resides in, including its own household, that *use* a sanitation system for defecation. In the survey, households were asked about the sanitation behaviour of groups of household members (boys, girls, male adults, female adults, male elderly and female elderly). Only if *all* of these groups were reported to use a toilet facility (their own, their neighbours' or a community toilet) is our indicator equal to one for this household. Usage of community/public toilets (or usage of neighbour's toilet) is very rare in our sample. Figure 2 shows a breakdown of usage rates by whether households own a toilet or not, split by survey round. It can be seen that usage rates are driven by high usage of privately owned toilets.

Figure 2: Sanitation usage and ownership - by location and round

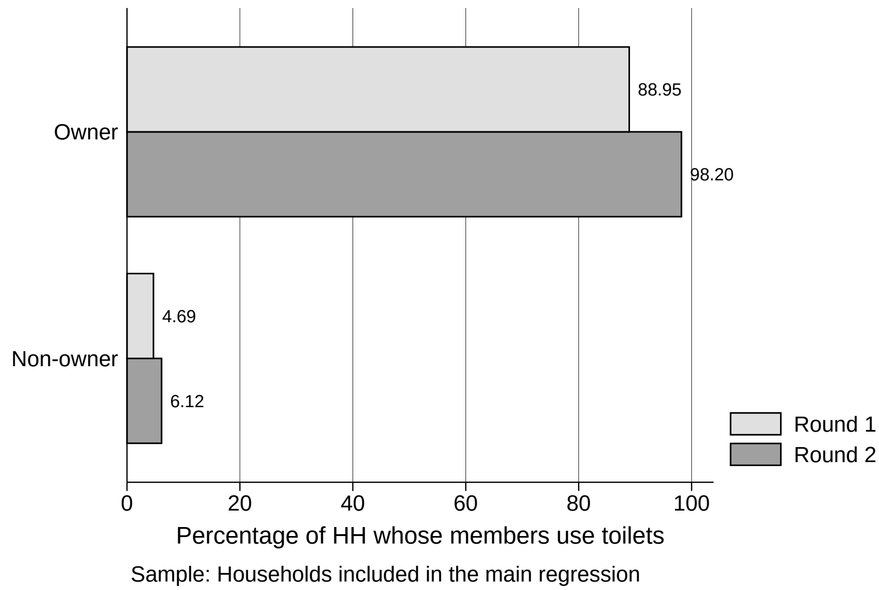


Figure 3 combines our community level sanitation environment measure and the average height z-score for our sample children. The bars at the bottom of this figure illustrate the variation in the sanitation environment variable. It can be seen that the sample spans communities where every household reports to use exclusively a sanitation facility for defecation and communities where no households does so. Within these extremes, a wide array of usage fractions are observed in our sample. The figure gives a first graphical indication that higher sanitation usage coverage is associated with lower stunting rates.

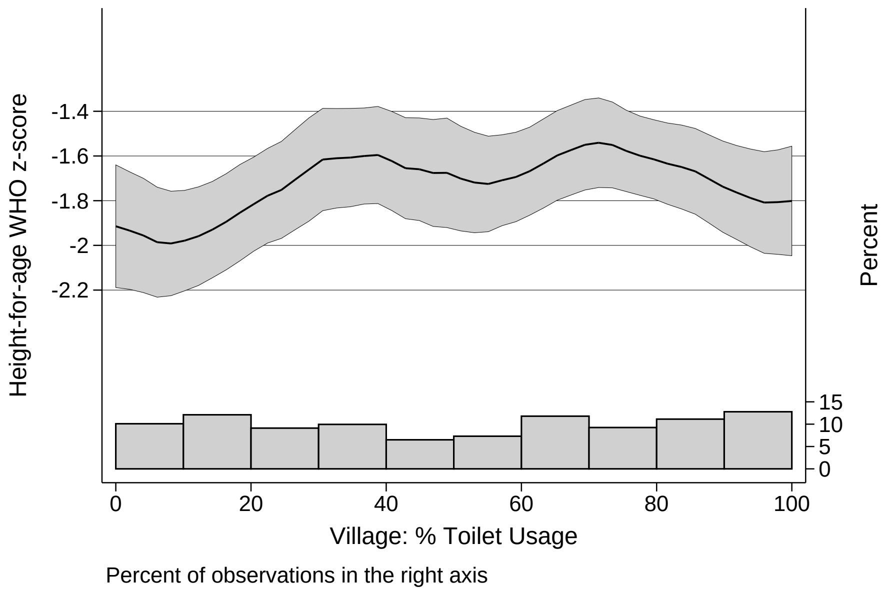
We will explore this further in our analysis below.

3 Methodology

The principal objective of this study is to understand the relationship between community level sanitation and child health, acknowledging that behavioural responses can induce endogeneity in our variable of interest.

Our empirical strategy is motivated by a model that combines insights from economists, demographers and epidemiologists. We take as a starting point Currie (2000)'s economic

Figure 3: Height of children and sanitation environment



model of determinants of child health and extend it, for one, by using insights from the Mosley-Chen framework (Mosley and Chen (1984)), which integrates approaches from both demographers and epidemiologists, and second, by building on recent advances in the production function literature. Details are provided in Appendix **??**.

3.1 Estimation specification

We take our theoretical model to the data by estimating the following regression specification:

$$Q_{i,v} = \alpha + \gamma ES_v + \delta_1 X^c_{i,v} + \delta_2 X^{hh}_{i,v} + \delta_3 X^v_{i,v} + \varepsilon^Q_{i,v} \tag{1}$$

where Q_{ivt} is the health of child i in village v. The sanitation environment of child i, ES_v is defined as the percentage of one-child households, i, in the village (including i) that use sanitation infrastructure for defecation, where $S_{i,v}$ is an indicator variable $= 1$ if all members of a randomly selected household in village v use the toilet they own or use a community toilet. The variable is zero otherwise. I_v indicates the total number of randomly selected households in village v. We get: $ES_v = \frac{1}{N_v}\sum_{i=1}^{I_v} S_{i,v}$. $X^c_{i,v}$ are relevant individual, i.e. child-level, characteristics, such as age andgender of the child. Household level variables, include the household composition, the education of the main woman in the household[10], income and shocks experienced; $X_{i,v}^{hh}$ are village level characteristics, include information on water and garbage disposal; $\varepsilon_{i,v}$ are shocks to health. The corresponding reduced form equation of the model is Equation 7 in Appendix A.2.

Since the surveys were desiged to track households and not children[11], we are unfortunately not able to use the data in a panel context and include fixed effects in our regression specification. Including such a fixed effect is often done to account for genetic endowment (see for example Puentes et al. (2014)). To rectify this, we proxy for health endowment of the child by controlling for it's mother's height.[12] This would primarily proxy for heritable endowment, which is seen as an important, unobserved determinant of child health. Medical papers suggest that 60-80 percent of height variation is determined by genetic factors (Ginsburg et al. (1998); Silventoinen (2003)). Ideally, we would like to also include the height of the child's father. Unfortunately, anthropometrics of male adult household members were not collected. It is however quite common to use only one parent's measure as a proxy for inherited endowment. This is also the case in the literature on early childhood development and education production functions, where for example the mother's mother's AFQT score is commonly used to proxy for genetic endowment of the child (Todd and Wolpin (2003)). Accounting for mother's height at the same time allows us to proxy for history of past inputs into child height, since we are not able to account for height for age of the child in the previous period.

In addition to emphasizing the importance of sanitation in child health production, the Mosley-Chen framework discussed the likely endogeneity of this input. For example, endogeneity might stem from households with a child that has a particularly weak immune system possibly being more likely to seek investment in infrastructure that keep the household's imminent environment free from contaminants, contributing to a negative correlation between demand for curative health inputs and good health. This is in contrast to the anticipated positive relationship of improvements in the imminent disease environment and health if such an improvement were randomly allocated to households of equally weak children.

As is the case for the individual ownership of sanitation infrastructure, also our broader definition of sanitation environment is likely to be endogenous. Take for example com-

[10]One could argue that a better indicator to include would be the education level of the child's mother specifically. Our data does not allow to identify this relationship within the household. Given the household composition, we can infer that in many cases the main woman is likely to be the mother. Where it is not, it is likely that the practices by the mother are influenced by the main woman in the household.

[11]In addition, given that three years passed between the two survey rounds, most children would not have fallen in the 0-5 year age category in both survey rounds.

[12]As mentioned before, for some children, this will not be the height of the mother, but the main woman in the household.

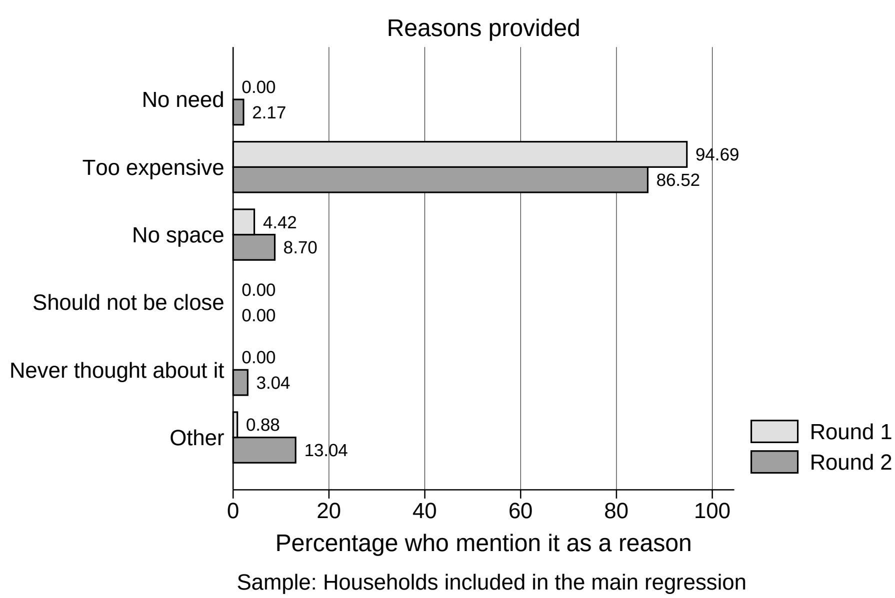
munities with very high population density and at the same time limited public (health) infrastructure, as is often the case in for example slums in developing countries. One can imagine that communities faced with such conditions which are likely to negatively impact health, to be more likely to make their own investments in infrastructure improving the disease environment.

To address this endogeneity of our main variable of interest, ES_v, we employ an instrumental variable approach, estimating the following first-stage regression:

$$ES_v = \mu_0 + \mu_1 X^c_{i,v} + \mu_2 X^{hh}_{i,v} + \mu_3 X^v_{i,v} + \mu_2 Z_v + \varepsilon^{ES}_{i,v} \quad (2)$$

Our choices of instrument, Z_v, is inspired by the production function literature. In this literature, input prices are typically acknowledged to affect investment choices as also outlined in the model (see Equation 9 in Appendix A.2), without entering the production function directly (Todd and Wolpin (2003), Puentes et al. (2014), Attanasio et al. (2015)). In line, we will argue below in Section 3.2 that prices for sanitation *raw* materials, which in our context exhibit sufficient geographic variation while at the same time being a significant predictor of sanitation uptake, are a suitable candidate.

Figure 4: Reported reasons for not owning a toilet



Given this instrumental variable approach, our parameter of interest, γ in Equation 1, is to be interpreted as a local average treatment effect. We are implicitly comparing the average level of child health in communities where dwellers are willing to build toilets but are restricted to do so by the level of raw material prices, to those dwellers for whom the restriction does not apply. Considering the large percentage of households in our study sample that report financial constraints to be the main barrier to sanitation uptake, we believe that this is a reasonable approach to follow. Figure 4 shows the reported reasons why households do not own a toilet for our study population. It can be seen that for almost all households that do not own a toilet the dominating reason for this is that this investment is too expensive, underlining the importance of prices in making the investment. Other studies have shown the importance of price and credit constraints in health purchasing decisions (Spears (2011); Dupas (2009); Cohen and Dupas (2010); Ashraf et al. (2010a)), including a recent study that demonstrates how availability of credit increases the willingness to pay for toilets (Ben Yishay et al. (2016)). We will discuss the validity of prices as an instrument in the next section.

3.2 Sanitation raw material prices as instruments

The key to our identification strategy is the sanitation raw material price we use as our instrument for sanitation coverage.

Input prices are generally acknowledged to affect investment choices without entering the production function directly (Heckman et al. (2007)) and are hence used, where feasible, to instrument endogenous variables in the context of production functions and beyond (see for example Attanasio et al. (2015); Puentes et al. (2014); Todd and Wolpin (2003)).

We have two types of input prices in our data: labour and raw materials.

Labour input prices are generally problematic since they might hide worker quality, which then enters the production function through the unobservable $\varepsilon^{Q}_{i,v}$ in Equation 1. As a result, $\varepsilon^{Q}_{i,v}$ is likely to be positively correlated with wages, invalidating the use of labour input prices as an instrument (Heckman et al. (2007)). We therefore do not present results here that include these prices[13] in the estimation. Findings are however comparable and available upon request.

Material input prices combine information on the price of four important components for toilet construction (cement, pipes, tiles and tin sheds) and the quantity needed for a pour-flush pit toilet, the predominant toilet type in the sample and, in fact, in the state more generally. We aggregate this information into a single input price. This price and quantity data was collected shortly after the baseline survey by contacting providers of raw materials for sanitation within each cluster of the study communities. Data is

[13]Specifically, we have available the approximate informal daily wage rate and the approximate time of building the standard design model supported by the Government of India.

available for 43 clusters in the study. More details on the collection are provided in (Gautam, 2015).

Table 4 gives average statistics for our available price variables. The average raw material prices for the listed toilet components amount to INR 8,300 (~US$ 178 at that time), ranging from about INR 5,500 to 10,800.[14]

Table 4: Toilet construction costs

	BASIC		MAIN	
	Mean	SD	Mean	SD
Cost of Raw Materials (1000s Rps)	8.3	1.7	8.2	1.7
Wages*Days required to build a toilet	0.3	0.1	0.3	0.1
Sanitation price in 1000Rps	8.6	1.7	8.6	1.7
Villages Round 1	38		36	
Villages Round 2	43		40	

Notes: Own calculations based on toilet prices from Gautam (2016).

BASIC: For which there is information about children and main woman age, height, gender, and that live in a village where price data was collected.

MAIN: Same sample as the main regressions.

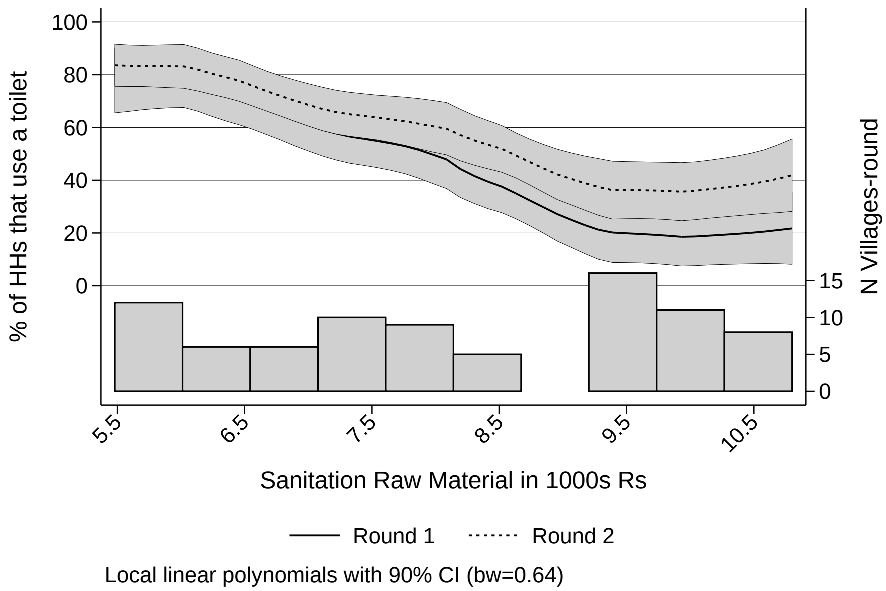
The necessary condition for this aggregate input price to be a valid instrument is that it is uncorrelated with $\varepsilon_{i,v}^{Q}$. Whether this is the case depends on the competitive nature of the input market that our households are operating in. If study households have power in input markets, input prices will be a function of the quantity of purchased inputs and the primary assumption to validate input prices as instruments breaks down. We argue that households are price takers with no power to influence the input prices. We believe this to hold for the following reasons: For one, the sanitation supply market in Madhya Pradesh is considered well developed and competitive in nature (Godfrey (2008)). Importantly, we do not consider input prices of materials specific for toilet construction but relevant for construction more generally. Demand for sanitation, which is growing but to date still relatively low in India, only makes up a small fraction of the construction market, hence unlikely to influence prices. It is further important to remember that we consider slum populations, which would typically build at small quantity and low costs, making it again unlikely to have impact on the overall market for these materials.

A typical limitation in the use of prices as instruments is limited variation, as one tends to think of input market prices as being fairly national in scope. This is not applicable in our context. The price variable displays econometrically helpful variation.

[14]We also show the estimated labour cost, which - when added to material costs - brings the cost average up to INR 8,600.

We display in Figure 5 the distribution of prices in relation to sanitation coverage in our communities, showing, as in previous figures, the distribution of observations across the x-axis in the bottom of the figure. The figure shows clearly the variation in input prices across study clusters. It also shows a clear downward trend in prices: The higher the price, the lower the coverage of used sanitation infrastructure.

Figure 5: Sanitation raw material prices and sanitation uptake



A natural question arises: What drives this variation in prices across a relatively small geographical area? Our analysis and discussion with experts in the field reveals that one of the main driver is access to the clusters. This is confirmed in Table 5 which shows significant correlations with variables proxying for access, particularly a dummy whether the location lies within the inner Gwalior area and a location index.

Table 5: Raw Materials Prices and Village Characteristics

OLS regression with Raw Materials Prices as a dependent variable	(1)	(2)	(3)	(4)	(5)
Inner Gwalior area	-1.411***	-1.423***	-1.033**	-1.020*	-0.600
	(0.459)	(0.511)	(0.449)	(0.574)	(0.529)
Village Scale and Location Index		-0.760***			-0.488**
		(0.252)			(0.234)
General Prices Index			-1.013***		-0.931***
			(0.162)		(0.161)
Water and Garbage disposal Index				-0.522	-0.321
				(0.352)	(0.274)
N Observations	78	69	74	75	68
N Villages	43	37	39	40	36
R Sqrd	0.148	0.297	0.459	0.207	0.556

Notes: Own calculations based on FINISH sanitation household data for Gwalior and raw materials prices from Gautam (2016). All specifications include a Round dummy. SE clustered at village level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

The fact that prices are higher in more central areas of the city is visualized in Figure 6 where withe round dots indicate sample communities with lowest raw material prices. These are almost exclusively located in the central part of the city (the area with a slightly darker shade of gray).

Figure 6: Community locations and prices of sanitation raw materials

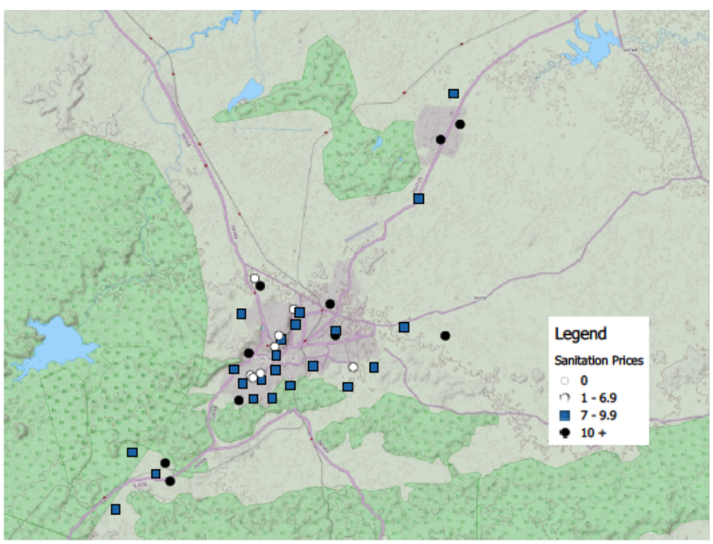


Table ?? in the Appendix provides a break-down of the location index. Correlations with individual components of the index are in line with the general significant correlation shown in Table 5. For example the variable whether the community is connected by local transport to the main bus stop in Gwalior has a significant and negative correlation with our price variable. The same holds for whether the cluster has its own shops and other variables proxying for better access.

The Table also shows that our instrument prices correlate with other prices in the village. This is not an unexpected finding, which we will pick-up in our robustness analysis.[15]

The correlation between our instrument and access becomes a concern when the input material price reflects access to other possible determinants and correlates of child health (particularly nutritional investments such as breastfeeding or dietary diversity) not acounted for in our analysis. We take this up in Section 4.1, where we conduct robustness checks on our findings, providing additional evidence that make us confident that our raw material prices are a suitable candidate to account for the endogeneity of sanitation coverage in our study area.

We will discuss in the next section how doing so affects our main question of interest: how an improvement in sanitation coverage affects children's height for age.

4 Results

Our main finding is shown in Table 6, which shows the coefficient of interest, γ in Equation 1.[16] Column (1) shows the OLS regression, which does not take into account the endogeneity of sanitation coverage. The regression results of this specification suggest that, while positive, the impact of an increased sanitation coverage on child height for age is small and not significantl. Once accounting for the endogeneity though through instrumenting with the price of raw materials, we see in column (2) that the coefficient becomes larger and is now significant at the ten percent level.

The fact that the OLS estimate is downward biased can be understood in the context of the example given previously:it indicates that households living in communities with worse conditions for health (such as very high population density coupled with limited (health) infrastructure) are more likely to make investments in private household sanitation. This behavioural response improves the wider sanitation and hence disease environment - leading to a negative, or lower, correlation between sanitation environment and health.

[15]The Table also includes also information on the correlation between prices and access to water and garbage disposal, other sanitation and hygiene related inputs. These we account for in our regression specification specifically.

[16]Full regression results, including the first stage, with information on all covariates are shown in Table 12 in the appendix.

Table 6: Avg Sanitation and height-for-age

	(1) OLS	(2) IV
Panel A: Second Stage		
Village % who uses a toilet	0.004	0.017**
	(0.005)	(0.008)
Panel B: First Stage		
Sanitation Raw Mat Price (1000 Rps)		−8.057***
		(2.244)
F-Stat		12.89
Obs	892	864
Clust	41	40
R2 Adj	0.11	0.10

Notes: Own calculations based on FINISH sanitation household data for Gwalior and raw materials prices from Gautam (2016). Controls: 3rd order polynomial on age, gender, mother education, quartiles of income, HH size, N of HH members who are males, any adverse shock last 2 years, slum and wave dummies, and a factor for quality of water and waste deposition. SE clustered at Village level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

The F-stat of the IV regression is 12.89, observably lower than in the 'pure' regression, where it is 17.13, due to accounting for a large set of covariates at the child, household and community level but still strong.[17]

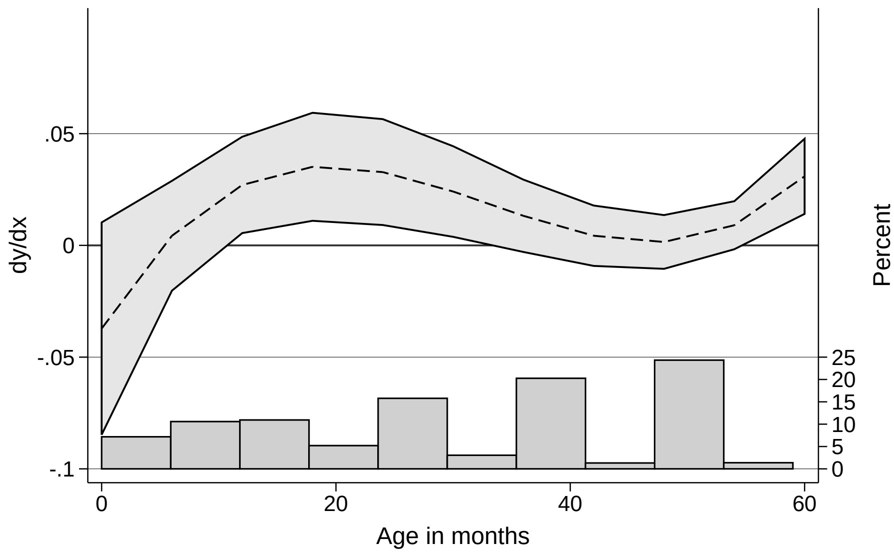
Figure 7 shows the marginal effects by the age of the child. We can see that the effects are positive and increasing in the approximate age range of six to 22 months. The fact that impacts start to materialize only when the child is around six months old is in line with the idea that a hygienic environment and breastfeeding can serve as substitutes due to antibodies in the milk (Van der Slice, Popkin and Briscoe 1994). It is typical in our study setting, for mothers to breastfeed almost exclusively for the first six months[18], so that we might indeed not expect an improvement in sanitation to have as much of an effect in this age range. Thereafter, and up to the age of two is on the other hand when most placidity in growth happens (Victoria et al. (2010)), providing hence the largest opportunity for impacts on child height for age .

The coefficient of 0.017 from the IV regression suggests that a 10% increase in sanitation usage coverage increases child height for age on average by 0.17 standard deviations of the z-score. To put this number in context, an increase of ten percentage point in sanitation coverage is translated into approximately 0.7 centimetres increase for a four year old child. This seems a sensible finding when compared to for example Richard

[17] It is typically said that, as a rule of thumb, the F-statistic of a joint test whether all excluded are irrelevant in the first-stage regression should be bigger than 10 (**?**).

[18] Around 47% reported to have given the baby some other type of liquid, like livestock milk, juice, or coffee.

Figure 7: Marginal effects by age



CI at 90% level based on SE clustered at Village Level. Marginal effects after 2SLS.

et al. (2013), who find - using data from seven cohort studies - that the cumulative effect on child's length from diarrhoea burden in the first two years of life to be 0.38 centimetres. This study has also been cited to suggest that impacts found by Hammer (2013) might be larger than biologically plausible. They suggest that an increase in toilet ownership of 8.2 percentage points leads to an increases in child height for age by 0.3-0.4 standard deviations or 1.3 centimetres when the child is four years of age. It is worth noting though that Richard et al. (2013) consider only diarrhoea and only the first two years of life.

4.1 Heterogeneous impacts by gender

In addition to looking at differential impacts by age of the child, we consider differential impacts by gender. The reasons for doing include potential differential exposure to the contaminated environment as well as differential immune capacity of male and femal children. We will go into more detail below.

Table 7 shows our results.[19] As in Table 6, we show both OLS (columns 1-3) and IV (columns 4-6) regressions results. We show for both estimation approaches first

[19] We now include in Equation 1 an interaction between toilet usage and the gender variable, and also interact prices with gender. Similarly, for differential impacts by age, presented previously in Figure 7 above, we interact both usage and prices with a third order polynomial of age. The rest of the specifications remain the same.

sub-sample regressions for males (columns 1 and 4) and females (columns 2 and 5). Columns 3 and 6 include an interaction of our sanitation environment variable and the gender of the child.

We find that the overall positive impacts of improved sanitation environments on child health are driven by impacts on girls. The impact on boys is *in*significant while that for girls is significant at one percent. Also the instruments exhibit greater power in this joint specification. The estimated coefficient for girls indicates an increase of approximately 1.05 centimetres for a four year old girl when sanitation coverage is increased by ten percentage points.

Table 7: Avg Sanitation and height-for-age

	(1)	(2) OLS	(3)	(4)	(5) IV	(6)
	MALE	FEMALE	BOTH	MALE	FEMALE	BOTH
Panel A: Second Stage						
Village % who uses a toilet	0.002	0.007		0.008	0.025***	
	(0.006)	(0.005)		(0.009)	(0.009)	
Village Avg * Boy			0.003			0.014
			(0.005)			(0.009)
Village Avg * Girl			0.004			0.021***
			(0.005)			(0.008)
Girl			0.017			−0.224
			(0.208)			(0.295)
Panel B: First Stage						
Sanitation Raw Mat Price (1000 Rps)				−8.252***	−8.045***	
				(2.115)	(2.309)	
F-Stat				15.22	12.13	18.65/ 12.93
Obs	459	433	892	446	418	864
Clust	40	40	41	39	39	40
R2 Adj	0.07	0.15	0.11	0.07	0.12	0.10
H0: $\beta_{Girls} - \beta_{Boys} = 0$			0.88			0.26

Notes: Own calculations based on FINISH sanitation household data for Gwalior and raw materials prices from Gautam (2016). Controls: 3rd order polynomial on age, gender, mother education, quartiles of income, HH size, N of HH members who are males, any adverse shock last 2 years, slum and wave dummies, and a factor for quality of water and waste deposition. SE clustered at Village level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

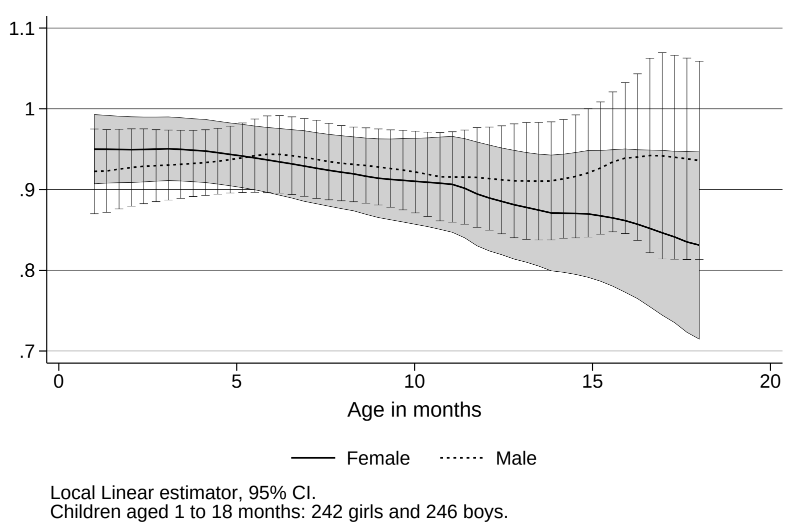
It is not immediately obvious why girls should benefit more from an improvement in the sanitation environment they live in. We discuss two possible reasons for this finding. These will have to be seen as suggestive as, unfortunately, data limitations do not allow us to test formally for their relevance,.

The first possible mechanism that could explain the differential impacts by gender relate to preferential investment. It is known that many Indian families have explicit preferences for sons over daughters (Pande and Astone, 2007). This male preference translates into differential investment: “[...] boys receive more childcare time than girls, they are breastfed longer and they get more vitamin supplementation” (Carvalho et al., 2013). There is further evidence that boys receive more nutrition (Das Gupta (1987)), more healthcare (Basu (1989), Ganatra and Hirve (1994)), are breastfed for longer (Jayachandran and Kuziemko, 2015), and are more likely to be vaccinated (Borooah, 2004) - all investments that are known to boost the immune system and increase health. Breastfeeding for example is shown to provide important inputs for the immune system, which some argue can act as a substitute for sanitation (VanDerslice et al., 1994).

Such differential investments in important health inputs. Favouring boys, could hence lead to girls responding more positively to improvements in the disease environment trough increased sanitation coverage.

Unfortunately, as we will discuss in more detail in the next section, our data is faces strong limitation in terms of measuring parental health investments, particularly nutritional inputs. We for example know only for children age 0-18 months whether they are currently breastfed and for children above the age of 18 months, what types of food items they ate the last day. And, this data does not support the possible mechanisms of differential parental health investment in a clear manner. When we for example consider breast-feeding information, we can see in Figure 8 that breastfeeding stops earlier for girls in our sample than it does for boys. However, this descriptive difference is not significant. Similarly, our data for other nutritional intake shows no evidence of a systematic gender difference[20].

Figure 8: Breastfeeding by age and gender



[20]More details in section A.5 in the Appendix.

The second possible mechanisms behind the differential impacts by gender found in our analysis relates to the possibility of girls being more directly affected than boys by the increase in sanitation usage coverage. Our data shows that if a toilet owned is not used by all household members, it is boys and men who are least likely to use the toilet. This suggests that girls frequent open defecation sites less than boys (possibly since their mums do not go anymore and hence do not take the girls along), and are hence exposed to a cleaner environment than boys (despite boys frequenting open defecating areas used by less people on average).

Robusness Checks

We consider three robustness checks to our analysis.

The first two checks relate back to our discussion of the validity of instruments in Section 3.2. One of these was a relatively minor point, where we showed that our raw material price index correlates with other community level prices, such as for food items. In line, when including index functions of food prices our estimate becomes less precisely estimated but stay otherwise similar (become rather larger), confirming that part of the variation in our instrument is related to other prices. This can be seen in Column (3) of Table 8, which displays our robusteness checks.

Importantly though, our results might be biased if the driving factor for the price variation is correlated with other child health inputs not accounted for in the analysis. Specifically, we argued in Section 3.2 that the price variation is driven by access. Column (2) in Table 8 repeats our analysis while accounting for the community location index which we showed before to correlate significantly with our price variable. It can be seen that our estimated coefficient remains equivalent in magnitude. It becomes slightly less precisely estimated, which might be driven by the reduced sample size (we do not have this location information available for three of the sample clusters).

The third robustness test considers similarly the relevance of omitted variables important in determining child health, namely nutrition. Unfortunately, although we have some information available, it does not allow us to make any strong conclusions. The surveys collected information on whether infants (18 months or younge) are breastfed and what types of food children older than 18 montha ate the day previous to the survey. Using this data comes with two main sacrifices: For one, including the information in our analysis reduces the sample to infants/older children, hence reducing our sample size significantly, and with that power. Splitting the sample into these two age groups however suggests that the impacts are driven by older children, as can be seen from columns (5) and (6) of Table8. Focusing more on these children, we can use thei nutrition information data to construct a dietary diversity index and include it in the regression to check for omitted variable bias (keeping in mind that this variable is likely to be endogenous itself!). The additional caveat we encounter is that we are faced with systematic non-reporting. We find that item non-response in children's consumption is related to household and child characteristics (age and gender of the child, age of the mother, and household size). When including it in our regression specification (column (7), we find that the impact of sanitation coverage on child health disappears. However, running our main specification (i.e. excluding the dietary diversity index) on only those children for which we have consumption data available (column (8)), we see that we similarly do not find impacts of sanitation coverage. This indicates that by including dietary diversity in the analysis, we are positively selecting those children into our analysis that are of better health and consequently likely to be less affected by the protective power of sanitation.

Table 8: Avg Sanitation and height-for-age Robustness Checks

	(1) BASE	(2) **MAIN**	(3) LOC	(4) PRI	(5) U18	(6) A18	(7) A18	(8) A18 †
Panel A: Second Stage								
Village % who uses a toilet	0.012**	0.017**	0.017*	0.035	0.022	0.017**	0.001	−0.000
	(0.006)	(0.008)	(0.010)	(0.039)	(0.024)	(0.007)	(0.008)	(0.007)
General Prices Index				−0.339				
				(0.450)				
Village Scale and Location Index			−0.049					
			(0.118)					
Dietary diversity measure							0.142***	
							(0.051)	
Panel B: First Stage								
F-Stat	17.13	12.89	10.20	1.28	8.51	14.97	17.00	15.94
Obs	964	864	813	820	200	664	472	472
Clust	43	40	37	39	37	40	40	40
R2 Adj	0.09	0.10	0.10	0.04	0.06	0.11	0.17	0.16

Notes: Own calculations based on FINISH sanitation household data for Gwalior and raw materials prices from Gautam (2016). SE clustered at Village level in parenthesis. **BASE:** Gender, age 3rd order polynomial, mother's height, age, adn dummies for Round 2 and for living near Gwalior centre. **MAIN:** The same as BASE but also mother's education, quartiles of HH income and its mean, quality of the dwelling, HH size, N of HH members who are males, any adverse shock last 2 years, religion and caste dummies, living near Gwalior centre and Round 2 dummies, and a factor for quality of water and waste deposition. **U18:** Age less than 18 months. **A18:** Age above 18 months. † Same sample as in column 6, the non-inclusion of dietary index is the sole difference. Significance: * 10%, ** 5%, *** 1%.

As an additional check we look at the correlation between our instrument and the dietary diveristy (as well as breastfeeding). We can do this for three samples: For one, all children for which dietary information as well as age and gender is available, second those for which we in additon have our covariates included in the main regression analysis, and third, those children for which we also have valid height for age measurements. These are displayed in Table 9. It can be seen that, correlations in all three cases suggest, if anything, that our raw material prices are positively correlated with dietary diversity, generating a downward bias on our estimates.

Table 9: Nutrition Investments and Raw Material Prices

	(1)	(2)	(3)
	Dietary Diversity		
	ALL	HH Cont	MAIN
Raw materials price index	0.053	0.101*	0.094
	(0.054)	(0.053)	(0.057)
Obs	627	520	463
Clust	43	40	40
R2 Adj	0.01	0.07	0.06

Sample: children aged 18 to 60 months.
ALL: All children for which there is information on age and gender. It also includes slum and wave dummies. **HH Cont:** These regressions include mother education, quartiles of income, HH size, N of HH memebers who are males, any adverse shock last 2 years and a factor for quality of water and waste deposition. **MAIN:** Children that on top of the previous controls, have a valid measure of height-for-age. SE clustered at Village level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

5 Conclusion

We make use of primary data collected as part of an evaluation exercise of a sanitation intervention to investigate the impact of improvements in the sanitation environment, defined as the fraction of households using private or community toilets, on child height for age, an indicator for health.

We do so in the context of slums and peripheral villages in a city in Northern India, Gwalior. This population is an important one to consider for two main reasons: India's slum population is growing rapidly while at the same time having no or only inadequate access to safe sanitation. High population density coupled with improper means of disposing faeces provides a breeding ground for preventable disease epidemics. Providing evidence on improvements in children's health that can be achieved by community-level sanitation improvements is hence of direct policy relevance.

Our results suggest that increases in sanitation usage rates significantly and positively affect children's height. This impact seems to be particularly relevant for girls. We suggest two possible mechanisms behind these impacts. Unfortunately we are only able to provide suggestive evidence for these mechanisms in our study population given data and sample size restrictions. However, independent of the drivers behind this finding, the results suggest that not only is investment in sanitation coverage worthwhile when children's health is one of the objectives, but increasing sanitation coverage seems to be at the same time a policy that implicitly targets girls.

References

Adair, L., Fall, C., Osmond, C., A.D., S., Martorell, R., Ramirez-Zea, M., Sachdev, H., Dahly, D., Bas, I., Norris, S., Micklesfield, L., Hallal, P., and Victora, C. (2013). Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: findings from five birth cohort studies. *Lance*, 382(9891):523–34. 1

Aggarwal, S. and Kumar, N. (2008). Level of poverty and employment pattern in slums: A case of gwalior in central india. *The Indian Journal of Labour Economics*, 51(2):315–30. 1

Almond, D. and Currie, J. (2011). *Human Capital Development Before Age 5*, volume pp.1315-1486. Elsevier. 1

Andres, L., Briceño, B., Chase, C., , and Echenique, J. (2014). Sanitation and externalities: Evidence from early childhood health in rural india. Technical report, World Bank. http://dx.doi.org/10.1596/1813-9450-6737. 5, A.2

Ashraf, N., Berry, J., and Shapiro, J. M. (2010a). Can higher prices stimulate product use? evidence from a field experiment in zambia. *American Economic Review*, 100:2383–2413. 3.1

Ashraf, N., Berry, J., and Shapiro, J. M. (2010b). Can higher prices stimulate product use? evidence from a field experiment in zambia. *American Economic Review*, 100(5):2383–2413.

Attanasio, O., Cattan, S., Fitzsimons, Meghir, C., and Rubio-Codina, M. (2015). Estimating the production function for human capital: Results from a randomized control trial in colombia. NBER Working Paper No. 20965. 3.1, 3.1

Augsburg, B., Caeyers, B., and Oteiza, F. (2015). The costs and benefits of investing in a toilet - views from indian and nigerian households and their policy implications. IFS Policy brief ISBN: 978-1-909463-94-3, Institute for Fiscal Studies.

Augsburg, B. and Rodriguez-Lesmes, P. (2015). Finish impact evaluation report.

Barcellos, S. H., Carvalho, L. S., and Lleras-Muney, A. (2014). Child gender and parental investment in india: Are boys and girls treated differently? *American Economic Journal: Applied Economics*, 6(1):157–189. 1

Basu, A. (1989). Is discrimination in food really necessary for explaining sex differentials in childhood mortality. *Population Studies*, 43:193–210. 4.1

Behrman, J., Calderon, M., Preston, S., Hoddinott, J., Martorell, R., and Stein, A. (2009). Nutritional supplementation of girls influences the growth of their children: Prospective study in guatemala. *American Journal of Clinical Nutrition*, 90(5):1372–9. 1

Ben Yishay, A., Fraker, A., Guiteras, R., Palloni, G., Sha, N., Shirrell, S., and Wang, p. (2016). Microcredit and willingness to pay for environmental quality - evidence from a randomized-controlled trial of finance for sanitation in rural cambodia. Technical report, Maryland Population Research Center Working Paper No. 2016-003. 3.1

Black, R. E., Allen, L. H., Bhutta, Z. A., Caulfield, L. E., de Onis, M., Ezzati, M., Mathers, C., and Rivera, J. (2008). Maternal and child undernutrition: Global and regional exposures and health consequences. *The Lancet*, 371(9608):243–60. 1

Borooah, V. K. (2004). Gender bias among children in india in their diet and immunisation against disease. *Social Science & Medicine*, 58(9):1719–1731. 4.1

Bozzoli, C., Deaton, A., and Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, 46(4):647–69. 1

Cameron, L. A., Shah, M., and Olivia, S. (2013). Impact evaluation of a large-scale rural sanitation project in indonesia. *World Bank policy research working paper*, (6360).

Cameron, Lisa; Shah, M. (2010). Scaling up rural sanitation: Findings from the impact evaluation baseline survey in indonesia. Technical report, World Bank. License: CC BY 3.0 IGO.

Campbell, D., Elia, M., and Lunn, P. (2003). Growth faltering in rural gambian infants is associated with impaired small intestinal barrier function, leading to endotoxemia and systemic inflammation. *Journal of Nutrition*, 133(5):1332–8. 1

Carvalho, L. S., Barcellos, S. H., and Lleras-Muney, A. (2013). Child gender and parental investments in india: Are boys and girls treated differently? *American Economic Journal: Applied Economics*. 4.1

Cebu, S. T. (1992). Achild health production funct estimated from longitudinal data. *Journal of Development Economics*, 38:323–351.

Chambers, R. and Bongartz, P. (2009). Beyond subsidies: triggering a revolution in rural sanitation.

Checkley, W., Buckley, G., Gilman, R., Assis, A., Guerrant, R., Morris, S., Molbak, K., Valentiner-Branth, P., Lanata, C., and Black, R. (2008). Multi-country analysis of the effects of diarrhoea on childhood stunting. *International Journal of Epidemiology*, 37(4):816–30. 1

Clasen, T., Boisson, S., Routray, P., Torondel, B., Bell, M., Cumming, O., Ensink, J., Freeman, M., Jenkins, M., Odagiri, M., et al. (2014). Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in odisha, india: a cluster-randomised trial. *The Lancet Global Health*. 1, 21

Coffey, D. (2013). Sanitation externalities, disease, and children's anemia. Working paper presented at annual meetings of population association of america, princeton university, R.I.C.E. 1

Coffey, D., Gupta, A., Hathi, P., Khurana, N., Spears, D., Srivastav, N., and Vyas, S. (2014). Revealed preference for open defecation. *Economic & Political Weekly*, 49(38):43.

Cohen, J. and Dupas, P. (2007). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Evidence from a Randomized Malaria Prevention Experiment (December 2007). Brookings Global Economy and Development Working Paper*, (11).

Cohen, J. and Dupas, P. (2010). Free distribution or cost-sharing? evidence from a randomized malaria prevention experiment. *Quarterly Journal of Economics*, 125(1):1–10. 3.1

Cunha, F. and Heckman, J. J. (2007). The technology of skill formation. *NBER Technical Report*.

Cunha, F., Heckman, J. J., and Schennach, S. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Currie, J. (2000). *Handbook of Health Economics*, volume 1, chapter 19. Child Health in developed countries, pages 1053–1090. Elsevier. 3, A.2, A.2

Das Gupta, M. (1987). Selective discrimination against female children in rural punjab, india. *Population and Development Review*, 13:377–400. 4.1

De Cao, E. (2013). The height production function from birth to maturity. University of Groningen Working Paper, 14018-EEF.

de Onis, M., Brown, D., Blossner, M., and Borghi, E. (2015). Levels and trends in child malnutrition. joint child malnutrition estimates. Technical report, UNICEF Geneva and WHO Washington DC and World Bank. 1

Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences*, 104(33):13232–13237. 1

Dickinson, K. and Pattanayak, S. (2012). Open sky latrines: do social effects influence technology adoption in the case of a (very) impure public good. Technical report, Duke University.

Duflo, E., Greenstone, M., Guiteras, R., and Clasen, T. (2015a). Toilets can work - short and medium run health impacts of addressing complementarities and externalities in water and sanitation. NBER Working Paper w21521. A.2

Duflo, E., Greenstone, M., Guiteras, R., and Clasen, T. (2015b). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation. Technical report, National Bureau of Economic Research. 1

Dupas, P. (2009). What matters (and what does not) in households' decision to invest in malaria prevention? *American Economic Review*, 99(2):224–230. 3.1

Dupas, P. (2011). Health behavior in developing countries. *Annu. Rev. Econ.*, 3(1):425–449.

Fuller, J., Villamor, E., Cevallos, W., Trostle, J., and Eisenberg, J. (2016). I get height with a little hep from my friends: herd protection from sanitation on child growth in rural ecuador. *International Journal of Epidemiology*, 1(10). 1

Ganatra, B. and Hirve, S. (1994). Male bias in health care utilization for under-fives in a rural community in western india. *World Health Organization Bulletin*, 72(1):101–4. 4.1

Gautam, S. (2015). Quantifying welfare effects in the presence of externalites: An ex-ante evaluation of a sanitation intervention. Unpublished. 3.1

Gertler, P., Shah, M., Cameron, L., Martinez, S., and Patil, S. (2014). How does health promotion work? evidence from the dirty business of eliminatingopen defecation. 5, A.2

Geruso, M. and Spears, D. (2014). Sanitation and health externalities: Resolving the muslim mortality paradox. 5, 8, A.2

Ginsburg, E., Livshits, G., Yakovenko, K., and Kobyliansky, E. (1998). Major gene control of human body height, weight and bmi in five ethnically different populations. *Annals of Human Genetics*, 62:307–322. 3.1

Godfrey, A. (2008). Situation assessment of the supply market for rural sanitation in himachal pradesh and madhya pradesh. Technical report, Water and Sanitation Porgram of the World Bank. 3.1

Grantham-McGregor, S., Cheung, Y., Glewwe, P., Richter, L., and Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lan*, 369(9555):60–70. 1

Hammer, Jeffrey; Spears, D. (2013). Village sanitation and children's human capital : Evidence from a randomized experiment by the maharashtra government. Technical report, World Bank. License: CC BY 3.0 Unported. 1, 4

Hathi, P., Haque, S., Pant, L., Coffey, D., and Spears, D. (2014). Place and child health - the interaction between density and sanitation in developing countries. Technical report, Policy Research Working Paper 7124 World Bank Group. 1

Heckman, J. J., Griliches, Z., Leamer, E., and Intriligator, M. (2007). *Handbook of Econometrics*. Elsevier. 3.1

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4):605–54.

Hoddinott, J., Behrman, J., Maluccio, J., Melgar, P., Quisumbing, A., Ramirez-Zea, M., Stein, A., Yount, K., and Martorell, R. (2013). Adult consequences of growth failure in early childhood. *The American Journal of Clinical Nutrition*, 98(5):1170–78. 1

Hoddinott, J. F., Maluccio, J. A., Behrman, J. R., Flores, R., and Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *The Lancet*, 371(9610):411–6. 1

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

Hutton, G., Haller, L., et al. (2004). *Evaluation of the costs and benefits of water and sanitation improvements at the global level*. Water, Sanitation, and Health, Protection of the Human Environment, World Health Organization.

Jayachandran, S. and Kuziemko, I. (2015). Why do mothers breastfeed girls less than boys? evidence and implications for child health in india. *Quarterly Journal of Economics*, 126(3):1485–1538. 4.1

Joint Monitoring Programme for Water Supply and Sanitation (2012). Progress on drinking water and sanitation: 2012 update. Technical report, WHO/UNICEF, Geneva, New York:.

Kumar, S. and Vollmer, S. (2012). Does access to improved sanitation reduce childhood diarrhoea in rural india? *Health Economics*, 22(4):410–27.

Kumar, S. and Vollmer, S. (2013). Does access to improved sanitation reduce childhood diarrhea in rural india? *Health Economics*, 22(4):410–427. 1

Leuven, E. and Sianesi, B. (2014). Psmatch2: Stata module to perform full mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. *Statistical Software Components*.

Lin, A., Arnold, B., Afreen, S., Goto, R., Nurul Huda, T., Haque, R., Raqib, R., Unicomb, L., Ahmed, T., Colford, J., and Luby, S. (2013). Household environmental conditions are associated with enteropathy and impaired growth in rural bangladesh. *American Journal of Tropical Medicine and Hygiene*, 89(1):130–37. 1

Lundeen, E., Stein, A., ., A. L., Behrman, J., Bhargava, S., Dearden, K., Gigante, D., Norris, S., Richter, L., Fall, C., Martorell, R., Singh, S. H., and Victora, C. (2014). Height-for-age z scores increase despite increasing height deficits among children in 5 developing countries. *American Journal of Clinical Nutrition*, ajcn.084368. First published online July 9, 2014. doi:10.3945/ajcn.114.084368.

Lunn, P., Northrop-Clewes, C., and Downes, R. (1991). Intestinal permeability, mucosal injury, and growth faltering in gambian infants. *Lancet*, 338(8772):907–10. 1

Maluccio, J. A., Hoddinott, J. F., Behrman, J. R., Quisumbing, A. R., Martorell, R., and Stein, A. D. (2009). The impact of nutrition during early childhood on education among guatemalan adults. *Economic Journal*, 119(537):734–63. 1

Martorell, R., Kahn, K. L., and Schroeder, D. G. (1994). Reversibility of stunting: Epidemiological findings in children from developing countries. *European Journal of Clinical Nutrition*, 48:S45–57.

Mbuya, M. and Humphrey, J. H. (2016). Preventing environmental enteric dysfunction through improved water, sanitation and hygiene: an opportunity for stunting reduction in developing countries. *Journal of Maternal Child Nutrition*, 12. Supplement 1, ahead of print. 1

Merchant, A., Jones, C., Kiure, A., Kupka, R., Fitzmaurice, G., Herrera, M., and Fawzi, W. (2003). Water and sanitation associated with improved child growth. *European Journal of Clinical Nutrition*, 57(12):1562–8. 1

Moore, S., Collinson, A., and Prentice, A. (2003). Birth season and environmental influences on patterns of thymic growth in rural gambian infants. *Acta Paediatrica*, 92(9):1014–20.

Mosley, W. and Chen, L. (1984). An analytical framework for the study of child survival in developing countries. *Bulletin of the World Health Organization*, 81(2):140–145. 3, A.2, A.2

Nokes, C., Grantham-McGregor, S., Sawyer, A., Cooper, E., Rpbinson, B., and Bundy, D. (1992). Moderate-to-heavy infection of trichuris trichiura affect cognitive function in jamaican school children. *Parasitology*, 104:539–547. 1

Nokes, C., van den Bosch, C., and Bundy, D. (1998). The effects of iron deficiency and anemia on mental and motor performance, educational achievement, and behavior in children: A report of the international nutritional anemia consultative group. Technical report, USAID: Washington, DC. 1

NSSO (2010). Some characteristics of urban sslum 2008-09. Report 534, National Sample Survey Office, National Statistical Organization, Ministry of Statistics and Programme Implementation, Government of India. 1

Pande, R. P. and Astone, N. M. (2007). Explaining son preference in rural india: the independent role of structural versus individual factors. *Population Research and Policy Review*, 26(1):1–29. 4.1

Patil, R., Arnold, B., Salvatore, A., Briceno, B., Ganguly, S., Colford Jr, J., and P.J., G. (2014). The effect of india's total sanitation campaign on defecation behaviors and child health in rural madhya pradesh: A cluster randomized controlled trial. *PLoS Med*, 11(8). http://dx.doi.org/10.1371/journal.pmed.1001709. 1

Pattanayak, S. K., Dickinson, K. L., Yang, J.-C., Patil, S. R., Praharaj, P., and Poulos, C. (2007). Promoting latrine use: Midline findings from a randomized evaluation of a community mobilization campaign in bhadrak, orissa. Working Paper 2, Research Triangle Institute. 1

Pattanayak, S. K., Yang, J.-C., Dickinson, K. L., Poulos, C., Patil, S. R., Mallick, R. K., Blitsteinb, J. L., and Praharajf, P. (2009). Shame or subsidy revisited: social mobilization for sanitation in orissa, india. *Bulletin of the World Health Organization*, 87:580–587.

Pickering, A., Djebbari, H., Lopez, C., Coulibaly, M., and Alzua, M. (2015). Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural mali: a cluster-randomised controlled trial. *The Lancet*, 3(11):701–711. 1

Pruess-Uestuen, A., Bos, R., Gore, F., Bartram, J., et al. (2008). *Safer water, better health: costs, benefits and sustainability of interventions to protect and promote health.* World Health Organization. 1

Puentes, E., Wang, F., Behrman, J., Cunha, F., Hoddinott, J., Maluccio, J., Adair, L., Borja, J., Martorell, R., and Stei, A. (2014). Early life height and weight production functions with endogenous energy and protein inputs. 3.1, 3.1, 3.1, A.5

Richard, S., Black, R., Gilman, R., Guerrant, R., Kang, G., Lanata, C., Mølbak, K., Rasmussen, Z., Sack, R., Valentiner-Branth, P., and Checkley, W. (2013). Diarrhea in early childhood: Short-term association with weight and long-term association with length. *American Journal of Epidemiolog*, 178(7). 4

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rosenzweig, M. and Schulz, T. (1983). Estimating a household production function: Heterogeneity, the demand for health input, and their effects on birthweight. *Journal of Political Economy*, 92:723–746.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Sanitation, C.-L. T. (2015). India. Last accessed 17th of August 2015.

Shah, N. B., Shirrell, S., Fraker, A., Wang, P., and Wang, E. (2013). Understanding willingness to pay for sanitary latrines in rural cambodia: Findings from four field experiments of ide cambodia's sanitation marketing program. Technical report, IDinsight.

Shakya, H. B., Christakis, N. A., and Fowler, J. H. (2014). Social network predictors of latrine ownership. *Social Science & Medicine*.

Silventoinen, K. (2003). Determinants of variation in adult body height. *Journal of Biosocial Science*, 35(2):263–285. 3.1

Sinroja, R. (2013). Slum sanitation in india - is there a case for private toilet provision? Blog. Last accessed 10th of September 2014.

Smith, L. and Haddad, L. (2014). Reducing child undernutrition: Past drivers and priorities for the post-mdg era. Institute of Development Studies Working Paper. Volume 2014 No. 441. 1

Spears, D. (2011). Economic decision-making in poverty depletes behavioral control. *BE Journal of Economic Analysis and Policy*, 11(1). 3.1

Spears, D. (2012). Effects of rural sanitation on infant mortality and human capital: Evidence from india's total sanitation campaign. Princeton. 1

Spears, D. (2014). Increasing average exposure to open defecation in india, 2001-2011. Working paper, R.I.C.E. 1

Spears, D. and Lamba, S. (2011). Effects of early-life exposure to rural sanitation on childhood cognitive skills: Evidence from india's total sanitation campaign. Princeton.

Spears, D. and Lamba, S. (2016). Effects of early-life exposure to rural sanitation on childhood cognitive skills: Evidence from india's total sanitation campaign. *Journal of Human Resources*. 1

Stopnitzky, Y. (2012). The Bargaining Power of Missing Women: Evidence from a Sanitation Campaign in India. MPRA Paper 37841, University Library of Munich, Germany.

Tarozzi, A., Mahajan, A., Blackburn, B., Kopf, D., Krishnan, L., and Yoong, J. (2014). Micro-loans, insecticide-treated bednets, and malaria: Evidence from a randomized controlled trial in orissa, india. *American Economic Review*, 104(7):1909–41.

Todd, P. and Wolpin, K. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):F3–F33. 3.1, 3.1, 3.1

Trémolet, S. and Ravi, K. (2013). Evaluating the potential of microfinance for sanitation in india. Research report, Sanitation and Hygiene Applied Research for Equity.

VanDerslice, J., Popkin, B., and Briscoe, J. (1994). Drinking-water quality, sanitation, and breast-feeding: their interactive effects on infant health. *Bulletin of the World health organization*, 72(4):589. 4.1

Victora, C., de Onis, M., Hallal, P., Bloessner, M., and Shrimpton, R. (2010). Worldwide timing of growth faltering: Revisiting implications for interventions. *Pediatrics*, 125(3):e473–e480. 1

Victoria, C., de Onis, M., Hallal, P., Bloessner, M., and Shrimpton, R. (2010). Worldwide timing of growth faltering: Revisiting implications for interventions. *Pediatrics*, 125(3):e473–480. 4

Vyas, S., Gupta, A., Medeazza, G., and Spears, D. (2014). Open defecation, population density, and child height in madhya pradesh, india: An ecological analysis of new data on 22,000 children. Working paper, R.I.C.E. 1

Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., Meeks Gardner, J. M., Powell, C. A., Rahman, A., and Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, 378(9799):1325–38. 1

WaterAid-India (2011). India country strategy 2011-2016.

WHO, W. (2008). Poor sanitation threatens public health. 1

World Health Organization (2013). Fact sheet n 330: Diarrhoeal disease. 1

A Appendix

A.1 Determinants of height and weight item non-response

A.2 A simple economic model for determinants of child health

The principal objective is to understand the relationship between community level sanitation and child health, acknowledging that behavioural responses can induce endogeneity in our variable of interest.

Mosley and Chen (1984) suggested a useful framework for us to built on. Their framework, which integrates approaches from demographers and epidemiologists, identifies a set of exogenous and endogenous determinants of child health and survival, acknowledging the role of household and community sanitation in determining child health and survival. Factors identified as exogenous include individual and household characteristics such as maternal education, income and family composition, institutional factors such as community infrastructure, ecological factors, such as rainfall and cultural factors, such as traditions and norms. Factors identified as endogenous are referred to as proximate determinants and include breastfeeding and household sanitation ownership. Combining insights from this framework with those from recent advances in the understanding of human capital production functions is useful in guiding the choice of variables to include in the estimation, in understanding which variables are likely to be endogenous and whether important determinants have been omitted. We extend Currie (2000) economic model of the determinants of child health in such direction. In this unitary household model, parents maximize the following objective function:

$$\sum_{t=1}^{T} E_t \beta^t U_{ivt} + B(A_{iv,T+1}) \tag{3}$$

In this household i in village v, inhabited by I_v households, parents are altruistic and get utility from their children's health status and the bequest, B, they leave to them. Period-specific utility is given by:

$$U_{ivt} = U(Q_{ivt}, S_{ivt}, C_{ivt}, L_{ivt}; X_{ivt}, u_{1iv}, \varepsilon_{1ivt}) \tag{4}$$

where, Q_{ivt} is child health, C_{ivt} is other consumption, L_{ivt} is leisure; and taste for them might differ according to some observed (X_{ivt}) and unobserved characteristics (u_{1iv}) and shocks (ε_{1ivt}). On top of this, households get utility from having access to a sanitation facility S_{ivt}. Reasons for this direct benefit of sanitation might include comfort, social-status, security, as well as health considerations of the adults.

In the original model by Currie (2000), the evolution of child health is shaped by parental physical, G_{ivt}, and time investments, V_{ivt}. Their productivity depends on observed (Z_{ivt}) and unobserved (u_{2iv}) characteristics as well as unobserved shocks (ε_{2ivt}). We extend the model to include an additional element: namely, what we term, 'environmental sanitation', or $ES_{v,t}$. We define this term as the percentage of one-child households, i, in the village (including i) that use sanitation infrastructure for defecation, where $S_{i,v,t}$ is an indicator variable = 1 if all members of a randomly selected household in village v use the toilet they own or use a community toilet. The variable is zero otherwise. I_v indicates the total number of randomly selected households in village v. We get: $ES_{v,t} = ES_{v,t} = \frac{1}{N_v} \sum_{i=1}^{I_v} S_{i,v,t}$.

This definition is driven by our interest in the role of infrastructure that isolates human waste, faeces, from the environment, i.e. sewage, community toilets and private household toilets. More specifically

Table 10: Determinants of Missing Antrophometrics Data

	(1)	(2)	(3)
Type of dwelling: strong	-0.0448*	-0.0265	-0.0295
	(-1.71)	(-1.05)	(-1.14)
Age in months	0.00309***	0.00242***	0.00242***
	(4.79)	(3.89)	(3.82)
Number of HH members	-0.0145***	-0.0141***	-0.0145***
	(-3.08)	(-3.03)	(-2.64)
=1 if Muslim	-0.0157	-0.0223	-0.0205
	(-0.47)	(-0.69)	(-0.62)
=1 if forward caste	0.0423	0.0398	0.0399
	(1.11)	(1.09)	(1.06)
=1 if minority backward caste	-0.0707	-0.0307	-0.0293
	(-1.33)	(-0.59)	(-0.56)
=1 if scheduled caste or tribe	0.00883	0.0130	0.0168
	(0.31)	(0.47)	(0.60)
Baseline Observation	-0.258***	-0.312***	-0.311***
	(-9.56)	(-12.12)	(-11.82)
Inner Gwalior area	-0.0238	-0.0385	-0.0363
	(-0.97)	(-1.63)	(-1.51)
Female		0.0189	0.0208
		(0.83)	(0.90)
Income: 40.15 - 60.00 K Rup			-0.00747
			(-0.24)
Income: 60.25 - 90.00 K Rup			-0.0308
			(-0.86)
Income: 90.30 - 280.00 K Rup			0.00640
			(0.17)
Constant	0.866***	0.934***	0.939***
	(17.04)	(18.43)	(17.12)
Observations	1338	1281	1249

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

we are interested in the *usage* of such facilities[21], primarily private household sanitation, but also the less common usage of community toilets and neighbours' toilets.

The Mosley and Chen (1984) framework defines sanitation ownership as a proximate determinant of child health, acknowledging its importance in providing a hygienic environment as well as the fact that it is likely to be endogenous. For example, endogeneity might stem from households with a child that has a particularly weak immune system possibly being more likely to seek investment in infrastructure that keep the household's imminent environment free from contaminants, contributing to a negative correlation between demand for curative health inputs and good health. This is in contrast to the anticipated positive relationship of improvements in the imminent disease environment and health if such an improvement were randomly allocated to households of equally weak children.

In our definition of disease environment we go beyond the imminent disease environment of the household, acknowledging that toilet ownership and usage provides a direct benefit as well as an external benefit, which is believed to be substantial (Duflo et al. (2015a); Gertler et al. (2014); Geruso and Spears (2014); Andres et al. (2014)). Using a toilet reduces own contact with faeces in addition to other private benefits a toilet might provide (time saving, privacy, etc). It further reduces the rate of open defecation, what is believed to be a major cause for parasite infections and diarrhoea, particularly observed in children under five years of age.

It is hence not just one's own toilet usage behaviour that determines health, but also the behaviour of neighbours and community members. As is the case for the individual ownership of sanitation infrastructure and discussed by Mosley-Chen, also this broader definition of sanitation environment is likely to be endogenous. Take for example communities with very high population density and at the same time limited public (health) infrastructure, as often the case in for example slums in developing countries. One can imagine that communities faced with such conditions which are likely to negatively impact health, to be more likely to make their own investments in infrastructure improving the disease environment. We will therefore need to deal with the likely endogeneity of $ES_{v,t}$.

Including this variable in Currie (2000)'s model of child health, we get a health production function which is a function of sanitation coverage. In other words, one of the relevant determinants of child health is determined at the village level. Depending on $f(\cdot)$, individuals might control or not this input. As a result, the health production function takes the following structure:

$$Q_{ivt} = f(Q_{iv,t-1}, G_{ivt}, ES_{vt}; Z_{ivt}, u_{2iv}, \varepsilon_{2ivt}) \tag{5}$$

The rest of the model follows Currie's structure. Parents get income from working H_{ivt} hours (where available time is normalised to unity), which reduces the amount of time available for leisure as well as investments in the child's health. Physical resources are distributed among savings, child-investments, a one-off sanitation investment T_{ivt}, and consumption. Relative to the standarised prices of other consumption, prices of child investments P^G_{vt} and toilet construction P^T_{vt} determine the marginal cost of both investments. Notice that once a household builds a toilet, its sanitation environment is assumed to improve permanently in the following period, through the personal ownership as well as the externality effect. This reflects the fact that gains from sanitation might not be immediate. Such resources grow with income Y which can come either from work at a wage w, from capital rent at a rate r, or from other source I_{ivt}. The related equations are:

[21] One of the reasons put forward for non-impacts on health in for example the study by Clasen et al. (2014) is that the constructed toilets were not used.

$$
\begin{aligned}
C_{ivt} =& Y_{ivt} - P^G_{vt} G_{ivt} - P^T_{vt} T_{ivt} - (A_{t+1} - A_t) \\
Y_{ivt} =& I_{ivt} + w_{vt} H_{ivt} + r A_t \\
1 =& L_{ivt} + V_{ivt} + H_{ivt} \\
S_{ivt} =& \max(S_{iv,t-1}, T_{iv,t-1})
\end{aligned}
$$

The model can be solved, and as in the original setup, to yield Frish demand functions. Within these, λ denotes the marginal utility of wealth and M corresponds to a vector of moments of the distribution of future observed and unobserved variables $\{X_{iv\tau}, Z_{iv\tau}, P_{iv\tau}, \varepsilon_{2iv\tau}, \varepsilon_{1iv\tau}, S^{-i}_{iv\tau}\}^T_{\tau=t+1}$. Here, S^{-i}_{ivt} is a vector which incorporates the sanitation status of all other households in village v, and P_{vt} is a vector of prices (including wage) at the village level for a given period, t. [22]

The Frish demand functions are of the following form:

$$C_{ivt}, H_{ivt}, T_{ivt}, G_{ivt} \text{ and } V_{ivt} = F(\beta, r, \lambda_{ivt}, X_{ivt}, Z_{ivt}, S_{ivt}, P_{vt}, ES_{vt}, u_{1iv}, u_{2iv}, \varepsilon_{2ivt}, \varepsilon_{1ivt}, M_{ivt}).$$

Given these, we can substitute both physical and time inputs into the health production function, Equation 5. If we also substitute for λ_{ivt} using the budget constraint, and assuming that M_{ivt} and A_{ivt} are functions of realizations of current, and past exogenous variables $J_{ivt} = \{X_{iv\tau}, Z_{iv\tau}, P_{v\tau}, I_{iv\tau}, \varepsilon_{2iv\tau}, \varepsilon_{1iv\tau}, S^{-i}_{iv\tau}\}^{t-1}_{\tau=1}$ and A_{iv0}, we get:

$$Q_{ivt} = f^{'}(Q_{iv,t-1}, A_{iv0}, \beta, r, I_{ivt}, , X_{ivt}, Z_{ivt}, S_{ivt}, P_{vt}, ES_{vt}, u_{1iv}, u_{2iv}, \varepsilon_{2ivt}, \varepsilon_{1ivt}, J_{ivt}) \tag{6}$$

and iterating over Q results in Equation 7. This reduced form equation of the production function makes it clear that there is a link between sanitation prices and health. Such link arises due to the reduction on the marginal cost of building a toilet, which increases demand for such good.

$$Q_{ivt} = f^{'}(Q_{iv0}, A_{iv0}, \beta, r, I_{ivt}, , X_{ivt}, Z_{ivt}, S_{ivt}, P_{vt}, ES_{vt}, u_{1iv}, u_{2iv}, \varepsilon_{2ivt}, \varepsilon_{1ivt}, J_{ivt}) \tag{7}$$

A reduced form expression for toilet ownership can also be derived:

$$T_{ivt} = T(A_{iv0}, \beta, r, I_{ivt}, , X_{ivt}, Z_{ivt}, S_{ivt}, P_{vt}, ES_{vt}, u_{1iv}, u_{2iv}, \varepsilon_{2ivt}, \varepsilon_{1ivt}, J_{ivt}) \tag{8}$$

As a result of the above, environmental sanitation at the village level is determined by a full set of present and past states $\theta_{vt} = \{Q_{iv0}, A_{iv0}, \{X_{iv\tau}, Z_{iv\tau}, I_{iv\tau}, \varepsilon_{2iv\tau}, \varepsilon_{1iv\tau}, S^{-i}_{iv\tau}\}^{t-1}_{\tau=1}\}^{I_v}_{i=1}$, which includes village level characteristics, and, importantly for our sub-sequent analysis, the village-specific vector prices.

$$ES_{iv,t+1} = f^*(\beta, r, S_{1vt}...S_{I_v vt}, \theta_{vt}, P_{vt}) \tag{9}$$

The model shows us that both the health production function as well as the demand for toilet ownership are influenced by unobserved idiosyncratic persistent and transitory shocks, initial conditions, and by the history of exogenous variables which might only be partially unobserved. Our goal is to identify

[22] Notice that if household i has an important weight in determining ES_{vt}, $S^{-i}_{iv\tau}$ might be a function of $\{X_{iv\tau}, Z_{iv\tau}, P_{iv\tau}, \varepsilon_{2iv\tau}, \varepsilon_{1iv\tau}\}^T_{\tau=t+1}$. Given this, household's i best response implies that the demands should include moments for all future variables of all individuals in the village $\{\{X_{\iota v\tau}, Z_{\iota v\tau}, P_{\iota v\tau}, \varepsilon_{2\iota v\tau}, \varepsilon_{1\iota v\tau}\}^T_{\tau=t+1}\}^{I_v}_{\iota=1}$. Here, for simplicity, we assume that this household has virtually no power in determining everyone else's adoption decision and that $S^{-i}_{iv,t+1}$ can be forecasted with some village characteristics.

$E[\partial Q_{ivt}/\partial ES_{ivt}]$, and given the presence of confounders, we will identify such marginal effect by exploit village level variation of P_{vt}, which induces exogenous variation on ES_{vt}. Notice that an additional channel is still open: the functional form of $U(\cdot)$ might imply that the demand for physical investments might be directly affected by the price of sanitation, for instance, with a CES specification. Such effects are expected to operate in an opposite direction to ES, as lower prices of raw materials will induce less physical investments, reducing Q. If that is the case, our estimates would be provide a bound of the impact of environmental sanitation. Another issue is if ES and the other inputs are substitutes or complements in the production function, which will imply different allocation of the inputs given the exogenous variation on ES. In the most extreme scenario, all the impact on health would be driven by agents that invest more on their children under the believe that the productivity of such investments is going to increase. Such questions on the functional form are beyond the scope of this paper.

In order to provide an estimate of such impacts, given the limitations of the data, we will impose some restrictions. First, $P_{vt} = P_{vt-1}$, as we do not have variation in time of such vector. Second, we will assume that the relationship between environmental sanitation and prices is as good as linear, as well as between child health and environmental sanitation. These strong assumptions restrict the analysis and avoid potentially key elements as non-linearity between ES and Q. Nevertheless, they allow us to get an idea of the strength of the link between both variables.

A.3 Location index

Table 11: Components of the Indexes (Round 1)

Variable	Mean (1)	Std Dev (2)	Correl. with Price Raw Materials (3)	Correl. with its Index
Village Scale and Location Index				
Were new dwellings built in this village in the last 12 months?	0.075	0.267	−0.297***	0.587***
Are autos available to drive to this bus stop?	0.875	0.335	−0.323***	0.223***
Village has kirana/general market shop?	0.951	0.218	−0.269***	0.251***
Village has wine shop?	0.400	0.496	−0.289***	0.796***
Village has tailoring shop?	0.750	0.439	−0.396***	0.582***
Village has fair price shop?	0.500	0.506	−0.320***	0.786***
Village has paan shop?	0.475	0.506	−0.316***	0.670***
Village has mahila mandal?	0.150	0.362	0.324***	0.115***
Village has community centre?	0.175	0.385	−0.272***	0.627***
Village has library?	0.050	0.221	−0.302***	0.421***
Village has panchayat office?	0.250	0.439	0.315***	0.186***
Village has fair price shop?	0.350	0.483	−0.305***	0.578***
Village has playground?	0.350	0.483	0.292***	0.357***
General Prices Index				
Price of 1kg sugar from market	40.200	3.757	−0.358***	0.483***
Price of 1l edible oils	55.050	6.664	0.333***	−0.465***
Price of 1kg onions	14.925	4.015	−0.380***	0.427***
Price of 1kg chicken	101.250	30.900	−0.508***	0.899***
Price of 1 tea	30.175	61.096	−0.522***	0.877***
Water and Garbage disposal Index				
Throw kitchen garbage away in waste baskets/trucks pick it up?	0.293	0.461	−0.328***	0.887***
Does the village community get water for cooking and drinking from hand pump?	0.537	0.505	0.433***	−0.249***
Does the village get water for cooking/drinking from household service connectio	0.390	0.494	−0.317***	0.945***

The first column is the average of the variable. The second corresponds to the correlation with raw material prices. The last one presents the correlation with the specific index.
Significance level: * 10%, ** 5%, *** 1%

A.4 Full regression results - main specification

Table 12: Avg Sanitation and height-for-age (Detailed)

Sample: Children aged 0 to 5.			
	(1) OLS	(2) IV-FS	(3) IV-SS
Panel A: Second Stage			
Village % who uses a toilet	0.004		0.017**
	(0.005)		(0.008)
Cost of Raw Materials (1000s Rps)		−8.057***	
		(2.244)	
Girl	0.047	0.828	0.089
	(0.128)	(1.052)	(0.127)
Age in months	−0.204***	0.358	−0.212***
	(0.055)	(0.335)	(0.055)
Age in months ^{2} (100s)	0.611***	−1.425	0.639***
	(0.194)	(1.281)	(0.194)
Age in months ^{3} (10000s)	−0.519***	1.526	−0.548***
	(0.192)	(1.320)	(0.191)
Mother Height (cm)	0.045***	−0.258*	0.048***
	(0.012)	(0.129)	(0.012)
Mother age	0.007	0.049	0.007
	(0.008)	(0.072)	(0.008)
Inner Gwalior area	−0.399**	12.779	−0.682**
	(0.149)	(7.633)	(0.272)
Baseline Observation	−0.971***	−13.052***	−0.866***
	(0.275)	(2.969)	(0.292)
Mother Education: 6-8 yrs	0.062	2.527	−0.010
	(0.269)	(2.435)	(0.279)
Mother Education: 9 yrs +	0.047	0.570	−0.031
	(0.257)	(2.098)	(0.264)
HH self-reported Income	−0.000	0.002	−0.000
	(0.003)	(0.046)	(0.003)
Total consumption expenditures of hh in last year	0.001	0.015*	0.001
	(0.001)	(0.009)	(0.001)
Income: 40.15 - 60.00 K Rup	−0.052	−2.647	0.005
	(0.181)	(2.483)	(0.188)
Income: 60.25 - 90.00 K Rup	0.219	−1.635	0.176
	(0.252)	(3.757)	(0.250)
Income: 90.30 - 280.00 K Rup	0.508	−3.286	0.502
	(0.397)	(6.954)	(0.415)
Type of dwelling: strong	−0.106	5.922	−0.215
	(0.252)	(4.798)	(0.279)
Type of dwelling: semi-strong	−0.282	−0.525	−0.266
	(0.218)	(3.930)	(0.227)
Number of HH members	−0.107**	0.072	−0.128***
	(0.048)	(0.372)	(0.049)
Number of male HH members	0.027	0.182	0.047
	(0.052)	(0.768)	(0.054)
Any household shock last 12 months	−0.201	−4.208	−0.099
	(0.337)	(3.421)	(0.343)
=1 if Muslim	0.146	−5.519	0.107
	(0.214)	(3.817)	(0.237)

Continued on next page

Table 12: (Continued)

Sample: Children aged 0 to 5.			
	(1)	(2)	(2)
	OLS	IV-FS	IV-SS
=1 if unkown religion	1.183	7.797	0.939
	(1.970)	(4.761)	(1.803)
=1 if forward caste	−0.229	5.643*	−0.312
	(0.274)	(2.900)	(0.309)
=1 if minority backward caste	−0.037	−1.853	−0.009
	(0.419)	(2.972)	(0.450)
=1 if scheduled caste or tribe	−0.238	1.366	−0.329
	(0.248)	(2.417)	(0.252)
=1 if unkown Caste	0.439	1.378	0.389
	(0.431)	(2.848)	(0.426)
Water and Garbage disposal Index	−0.162	16.853***	−0.447**
	(0.168)	(3.907)	(0.183)
Panel B: First Stage			
F-Stat			12.89
Obs	892	864	864
Clust	41	40	40
R2 Adj	0.11	0.73	0.10

SE clustered at Village level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

A.5 Gender differences on nutrition

We can compare the nutritional inputs for children above the age of 1.5 years by gender. For these set of children we have information on the types of foods they consumed in the day previous to the survey. This is the same information we use to construct the food diversity index, used in our robustness analysis. The available data does not provide information on quantities consumed, only a yes/no indicator on consumption of different food items.

For none of these food items do we find significant differences by the gender of the child, as shown in Table 13, except for food items rich in proteins consumed. This is an important food category within the context of this study. Puentes et al. (2014) for example show that "in contexts with substantial child malnourishment increases in protein-rich food intake in the first two years of life can have important effects on growth." However, our data suggests that it is girls that consum a higher amount of proteins than boys. This would suggest that differential nutrition investment is not at play in our setting.[23]

[23] We need to recall here that data on nutrition consumed by children above the age of two years is missing systematically, discussed earlier. The suggestive statements we make here need to be interpreted in this light and might not be valid for the representative household in the community.

Table 13: Nutritional inputs by gender

	Whole Sample	N	Female	Male	P-Value
Starchy staples	96.7	1262	96.8	96.6	0.429
Legumes	38.9	1229	37.9	39.9	0.349
Dairy (excluding breast milk)	87.0	1244	86.0	88.0	0.463
Meat, fish, egg	33.9	1237	37.0	30.9	0.014
Viamin A rich fruit or vegetables	44.5	1252	43.6	45.3	0.353
Other frutis or vegetables	42.5	1239	41.6	43.4	0.250
Foods made with oil, fats or butter	84.6	1246	84.7	84.5	0.515
Dietary diversity score of 0 to 2	8.0	1194	7.3	8.6	0.611
Dietary diversity score of 3 to 4	51.1	1194	52.9	49.3	0.131
Dietary diversity score of 5 to 7	41.0	1194	39.8	42.1	0.222

P-val corresponds to a t-test of difference of means between males and females clustering at village level. Round 2 data.