# Differential Access to Improved Sanitation in Nigeria

Trevor J. Bakker

Dissertation Supervisor: Imran Rasul

Dissertation submitted in part-fulfilment of the

MSc in Economics, University College London

September 2012

## Declaration

I, Trevor Jon Bakker, hereby declare that the work presented in this dissertation is my own original work. Where information has been derived from other sources, I confirm that this has been clearly and fully identified and acknowledged. No part of this dissertation contains material previously submitted to the examiners of this or any other university, or any material previously submitted for any other assessment.

Signature: [REDACTED]

Date: September 14, 2012

## Classification

This piece of research is primarily (please tick as appropriate):

- ☒ an empirical/econometric study
- □ the examination of a theoretical problem
- □ a critical analysis of a policy issue
- □ an analytical survey of empirical and/or theoretical literature

## Dissertation Library

"I give permission for this dissertation to be made available to future students by way of the Dissertation Library, should the Department of Economics choose to use it in this way."

*Yes*

## Abstract

Target 7.C of the Millennium Development Goals is to "Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation." Although most of the world is on track to meet this goal for safe drinking water, sanitation lags far behind. An estimated 2.5 billion people worldwide are unable to safely disposal of human waste. Children are especially likely to come into contact with faecal matter and are particularly vulnerable to diarrhoea. For children who survive to adulthood, childhood illness has been linked to lower earnings and lower levels of human capital. Access to improved sanitation technologies and related levels of child diarrhoea vary immensely with family income, and I investigate which varieties of sanitation might reduce diarrhoea using 2007 data from the Nigeria Multiple Indicator Cluster Survey. Logit specifications confirm benefits found in previous meta-analyses to improved sanitation in general and to flush toilets in particular, but conditional logit specifications that account for cluster fixed effects find no such benefit, suggesting that omitted variable bias may plague much of the existing literature. A gender differential is uncovered in which boys experience a lower probability of diarrhoea when shared sanitation rather than private facilities are available to them, while girls do not. Further investigation into the source of this differential and into the robustness of other data sets to cluster fixed effects is advised.



## Acknowledgements

I would like to thank my advisor, Imran Rasul, for invaluable advice in determining the scope of this project and for corrective guidance in framing the economic consequences of differential access to improved sanitation. I would also like to thank the US-UK Fulbright Commission and University College London, whose joint Fulbright-UCL Postgraduate Award made it possible for me to undertake the MSc Economics and write this dissertation. I also thank the Abdul Latif Jameel Poverty Action Lab (J-PAL) and Abhijit Banerjee, Rohini Pande, and Michael Walton, the principal investigators for whom I worked as a research associate from 2010–11. My encounter with the deprivation of Delhi slum areas was the germ of curiosity that developed into this inquiry of my own.

# Table of Contents

# List of Figures and Tables

# 1. Introduction

Over 2.5 billion people lack the ability to safely dispose of human waste. The consequences of this affront to human dignity are myriad, from exposure to faecal matter through open defecation to contamination of drinking water, but their combined burden falls most heavily on children. Diarrhoea, defined as the three or more loose or watery stools per day or blood in stool, kills approximately 1.5 million children each year, mainly through severe dehydration (WHO 2009). Those who survive will likely fall ill again, usually experiencing at least some degree of stunting and malnutrition (United Nations 2010, 61).

Children under age five are less able to understand and less likely to consistently apply good hygiene practices like hand washing, and their immune systems are less wellGequipped to fight off rotavirus and other common viral or bacterial causes of diarrhoea, so decreasing exposure to faecal matter is essential. Ensuring their access to improved sanitation regardless of family wealth decreases this exposure, and the Millennium Development Goals seek to halve the number of people who lack basic sanitation. The MDG sanitation goal is unlikely to be met by 2015, however, underlining the importance of continued attention to the problem (WHO/UNICEF JMP 2012, 21).

Across the developing world, a health gradient exists in which poorer families are less likely to have access to improved sanitation and poorer children are more likely to experience diarrhoea. Income-health gradients are self-reinforcing since children who suffer from illness and poor nutrition in childhood as predicted by low family income tend to enter adulthood with lower levels of health and human capital, reducing their earnings potential and thereby transmitting poor health to their own children (Case, Lubotsky and Paxson 2002, 1331-1332). My aim is not to prove the disease-earnings link, as this is well established, but instead to gain a clearer picture of how this link manifests itself through differential access to improved sanitation.

This paper takes Günther and Fink's 2010 meta-analysis of the U.S. Agency for International Development's Demographic and Health Surveys (DHS) as its primary reference, attempting to replicate its findings of reduced likelihood of diarrhoea for children with access to improved sanitation. I use data from the Multiple Indicator Cluster Surveys (MICS) by United Nations Children's Fund (UNICEF), focusing on the case of

Nigeria as the most populous MICS-3 country and one with high levels of child diarrhoea. A stated aim of MICS-3 is comparability with the Demographic and Health Survey, and the Nigeria team was sensitive to this in formulating the survey (UNICEF and National Bureau of Statistics 2007, 2). Thus, the findings of my analysis may assess the plausibility of Günther and Fink's findings and the reliability of their methodology.

Both logit and conditional logit specifications are used to test three classification systems for basic sanitation with and without interactions between explanatory variables for a total of twelve main specifications. Confirming previous findings, the logit models reveal 21.8 percent lower odds of child diarrhoea associations with improved sanitation and 30.2 percent lower odds associated with flush toilets. However, a major problem I identify is omitted variable bias present in specifications that do not account for the clustered sampling used in MICS and DHS by including cluster fixed effects. I account for the complex survey design, including stratification and multiple stages of random sampling, to give more accurate reporting of standard errors, and this influences the significance of some coefficients of interest.

At the same time, I extend the analysis beyond Günther and Fink, investigating whether shared facilities are related to the outcome of diarrhoea. Conventional wisdom holds that shared facilities are likely not as well maintained, but I find an 18.0 to 22.8 percent decrease in the odds of child diarrhoea associated with shared facilities for boys. This result holds with and without cluster fixed effects, but it never materialises for girls, suggesting a curious gender differential. While I am unable to explain the source of this differential from MICS-3 data, I offer some tentative ideas as to its origin that should help guide future investigations.

Motivating this analysis is a belief that if aspects of sanitation are found to be associated with lower incidence of diarrhoea, they should be treated as investments to be made in children's well-being. Cutting the transmission of poor health between generations requires suppressing the channels like diarrhoea that stunt child development and lower life prospects for health and income.

# 2. Literature Review

## 2.1. The Income-Health Gradient

Economists have a longstanding concern for social inequality and its economic determinants, and a recurring finding in the economics of development and health is that differences in income manifest themselves as differences in health. Known as the income-health gradient, this finding is particularly pervasive because of its self-reinforcing nature: poorer children experience more frequent and more severe illness throughout childhood, develop chronic conditions, accumulate less human capital in the form of education, and enter adulthood with lower earnings potential, setting the stage for transmission of poor health to the next generation. Case, Lubotsky and Paxson 2002 offer compelling evidence that “adverse health effects of lower income accumulate over children’s lives” and result in “lower levels of human capital accumulation” as children miss school (1308 and 1331).

One way of explaining the pervasiveness of this problem is to see education as a form of general human capital that enables a person to maintain good health and recover from illness. Those with better education can more readily make use of advances in technology and apply knowledge about best practices in sanitation (Cutler, Deaton and Lleras-Muney 2006, 115). On the other hand, the problem is about more than knowledge, as health foundations developed in childhood influence adulthood earnings. Chronic illness keeps people out of work and causes them to work less productively, taller people command higher wages on average, people with a higher body mass tend to be more productive and earn more through this channel, etc. (Strauss and Thomas 1998, 767). Income-health gradients dog the poor from birth to death.

These gradients exist in both richer and poorer parts of the world, but only in the latter do they continue to manifest themselves in terms of mortality and morbidity from diarrhoea. Diarrhoea is easily treatable through oral rehydration therapy at the cost of a few pence per treatment, but poorer people tend to live further from hospitals and in areas with fewer community health resources (WHO 2009). Among the poorest, an urban/rural gradient is well-documented, including among other effects lower access to basic sanitation in the rural areas. Worldwide, an estimated 949 million people practice open defecation in rural areas compared to 105 million in urban areas (WHO/UNICEF JMP

2012, 20), and adults engaging in this practice also suffer elevated rates of diarrhoea. Maternal diarrhoea can hinder foetal growth, causing poor children to suffer the ill effects of the income-health gradient before leaving the womb.

Since the social benefit of sanitation far outstrips the private benefit, there is reason to suspect that households undersupply good sanitation relative to their ability (Günther and Fink 2010, 36). This creates scope for public provision of toilet facilities, a strategy employed in many urban areas, where space is limited and the extremely poor are often clustered. Shared toilets are now used by 11 percent of the world's population and continue to rise in both absolute numbers and relative proportions (WHO/UNICEF JMP 2012, 21). Nonetheless, the Joint Monitoring Programme (JMP) does not consider shared facilities to be improved on the grounds that they are less hygienic, and I investigate whether the evidence from child diarrhoea supports this concern.

Economists have also observed a gap between stated willingness to pay and actual purchases of technologies that improve health. Mobarak et al. (2012) find that women in rural Bangladesh do not prioritise purchasing improved cookstoves because they do not perceive indoor air pollution as a significant health hazard. Much as these women have cooked with their mothers and inhaled smoke from burning solid fuels from a young age, many children have engaged in open defecation for as long as they can remember. Without perceive a causal relationship between their practice of open defecation and health consequences, they may be unwilling to expend scarce financial resources to purchase improved private toilets. Without in any way diminishing the usefulness of improved engineering, there is clear scope for contribution from economists in understanding why currently available technologies have not disseminated and which may deliver the biggest benefit in reducing child morbidity. The problem of adoption goes beyond the purview of this paper, but it emphasises the need for proven benefits backed by compelling evidence for specific forms of sanitation if the poor are to adopt them.

A final possible gradient in sanitation is between the sexes. Oster and Thornton's (2011) work on female sanitary pads and menstruation suggests a need to consider how boys and girls may be differently affected by access to washrooms and toilet technology. To that end, I consider the possibility that boys and girls may have different responses to private and shared sanitation facilities and explicitly incorporate this in my research design. Since breastfeeding is associated with decreased child exposure to faecal matter through water and food, I also consider findings from India showing that breastfeeding is lowest for daughters and children without brothers (Jayachandran and Kuziemko 2011). I do not expect to find such pronounced son preference in Nigeria, but my investigation considers the possibility that boys and girls may be affected differently by breastfeeding.

### 2.2. Sanitation Benefits and the Millennium Development Goals

Three meta-analyses examine the possible benefits to sanitation in terms of reduced diarrhoea. Together they reveal that fewer than a dozen studies from the last few decades are both focused the relationship between sanitation and diarrhoea and of good quality, with far more examining the water-diarrhoea connection. Esrey et al. (1991) find a 22 percent reduction in morbidity from improvements in sanitation over eleven studies, noting that infants of illiterate mothers and non-breastfed infants are more likely to benefit from the improvements (212-213). Reviewing four relevant studies and finding two applicable, Fewtrell et al. (2005) calculate a pooled relative risk of 0.68, or a 32 percent reduction in odds of diarrhoea, which is relatively comparable to odds given the low probability of the even of diarrhoea, about 10 percent in a two week sample (44). A potential concern is that publication bias could be driving estimated effects upward by excluding studies with lower effects from meta-analysis. These results also highlight the paucity of research devoted to the contribution of sanitation to diarrhoea and the scope for finer-grained research. Günther and Fink (2010) venture into this void with their meta-analysis of DHS, combining 172 surveys to find that latrines and flush toilets lower the odds of diarrhoea by 7 and 13 percent, respectively (20-21).

Target 7.C of the Millennium Development Goals is to “Halve, by 2015, the proportion of the population without sustainable access to safe drinking water and basic sanitation.” Although most of the world is on track to meet this goal with regard to safe drinking water, basic sanitation lags far behind, with only 67 percent of the world projected to be using improved sanitation by 2015, falling short of the goal of 75 percent (WHO/UNICEF JMP 2012, 15). Basic sanitation refers to the safe disposal of human waste, which an estimated 2.5 billion people throughout the world lack (United Nations 2010, 60). Throughout this analysis I refer to “improved sanitation,” a term defined by the JMP and detailed in section 3.2, but I also consider the possibility of sanitation benefits from unimproved “basic” technologies like pit latrines.

## 2.3. Case Selection

Although the *indicators* in MICS are designed to be comparable across countries for purposes of monitoring the Millennium Development Goals, the samples are not generally intended to be combined. Indeed, the pooling of data from various multistage designs for several dozen countries raises econometric concerns, and many meta-studies perform regression analysis on these disparate data as though they were gathered in a single survey. I choose to avoid the problems inherent in meta-analysis in favour of a detailed examination of MICS-3 data from a single country.

Nigeria is a compelling case given the size of the country and the magnitude of the sanitation problems it faces. The 2006 Census counted a total population of 140,431,790 (National Bureau of Statistics 2010, 18), which is projected to have grown to 170 million as of July 2012, ranking seventh in the world (Central Intelligence Agency 2012). As the most populous of 53 countries represented in the MICS-3 surveys, and one with a sizeable share of urban households (roughly one quarter) in the sample, Nigeria is a ripe case for analysis. Of the twelve countries where more than 35 percent of the urban residents use shared facilities, eleven are in Sub-Saharan Africa (WHO/UNICEF JMP 2010, 23), yet relatively greater attention has been given in the literature to South Asia. As the most populous country for which MICS-3 is available and as a representative of Sub-Saharan Africa, Nigeria is highly relevant to the sanitation discussion.

Moreover, investigation of sanitation data is relevant for Nigeria. As UNICEF explains, "Diarrhoea is the second biggest killer of children in Nigeria, responsible for about 17 percent of child deaths every year, the majority (88 percent globally reported) of which is attributed to inadequate water, sanitation and hygiene" (2010, 1). Though my analysis is centred on the sanitation leg of this stool, water technology is also included and coefficients for its role as an explanatory variable are included. Hygiene data for vegetable washing, food storage, and cooking is beyond the scope of this analysis and likely better collected through direct observation than reporting through interviews. With this brief introduction to the case of Nigeria and a picture of its critical sanitation needs in mind, I now examine the MICS-3 survey and preliminary findings about the relationship of diarrhoea prevalence and sanitation technology to income.

# 3. Data

### 3.1. Survey Methodology

All data is taken from the third round of the Multiple Indicator Cluster Survey (hereafter "MICS-3"), in particular from the batch of data collected in Nigeria in 2007. MICS-3 is a standard set of questions designed by the United Nations Children's Fund (UNICEF) to measure *multiple indicators* related to mother and child wellbeing by surveying *clusters* of households in countries throughout the world. Encompassing all 36 of Nigeria's states and the Federal Capital Territory of Abuja for a total of 37 strata, this batch from MICS-3 is designed to give estimates that, among other things, can monitor progress at the state and national levels toward fulfilling the Millennium Development Goals. Nigeria MICS-3 consists of three core questionnaires that cover 28,603 households, 27,093 women, and 17,093 children under age five. The child data, merged with relevant household and mother variables, is the basis of this analysis.

MICS-3 is particularly useful for its common core of sanitation questions:

- WS7: "What kind of toilet facility do members of your household usually use?"
- WS8: "Do you share this facility with other households?"
- WS9: "How many households in total use this toilet facility?"

Knowing what sanitation is available to children and whom they share it with may identify pressing needs and provide insight into how sanitation construction could best meet them. MICS-3 data also benefits from relatively high response rates: 94.0 percent among households, 85.3 percent among women, and 91.0 percent among children.

All fieldwork was concluded within the span of a month from March 14 to April 12, 2007 (UNICEF and National Bureau of Statistics 2007, xvii). An advantage of this approach, which required more surveyors to work simultaneously than a survey of the same sample over a longer period, is that it reduces the scope for exogenous shocks to affect some interview subjects but not others. The longer the period of surveying, the higher the likelihood, for example, that local water might become contaminated and cause an outbreak of diarrhoea that would only affect the responses of those surveyed afterward. The shorter the period, the more safely we can think of this data an instantaneous cross-section of the population.

In anticipation of the 1991 census, the Nigerian government for the first time delineated its entire territory into 209,505 enumeration areas, which were designed to be "complete and distinct with no overlapping settlements or housing units" (National Population Commission 1998, 10). These enumeration areas, updated to reflect the creation of six new Nigerian states in 1996, were relied upon by UNICEF and the National Bureau of Statistics for sampling in MICS-3 (2007, 3). In each state, 30 enumeration ares were randomly selected for 1,110 total, a full list of households was drawn up, and 25 per enumeration area were randomly selected for survey. All children in the household under age five were surveyed. As a result of this design, the survey oversampled people from less populous states.[1] Fortunately, the survey data includes weights to account for both overrepresentation of smaller states and nonresponse, allowing these problems to be addressed in my specifications.

### 3.2. Prevalence of Child Diarrhoea

Following the convention in medical literature, MICS-3 states, "Diarrhoea is determined as perceived by mother or caretaker, or as three or more loose or watery stools per day, or blood in stool" (UNICEF and National Bureau of Statistics 2007, 255). MICS-3 data records whether a child has experienced diarrhoea in the two weeks prior to the survey.

Overall, 10.6 percent of children surveyed reported diarrhoea (Table 4, Appendix). As Figure 1 illustrates, the income-health gradient is pervasive in Nigeria. Diarrhoea is decreasing in wealth quintiles, with each successive quintile representing a noticeably lower proportion of children with diarrhoea the higher one moves on the scale. Close to 13 percent of children surveyed in the lowest wealth quintile reported diarrhoea compared to only 7.5 percent of those in the highest quintile. Poorer households may be unable to afford private sanitation facilities and may live closer to unimproved water sources, raising their risk of childhood diarrhoea. Figure 2 shows that diarrhoea incidence rises sharply in a child's first year of life, peaking around 17 months old at 18 percent and declining to around 5 percent in later childhood.

---

[1] "In the end, 30 EAs were selected into the sample as PSU from each state in spite of the huge differentials in state populations. The most potent argument in favour of this disproportionate allocation is that the state as the second tier of governance is the most critical to national development; there is also this political fact about equality of states" (UNICEF and National Bureau of Statistics 2007, 59).

**Figure 1: Proportion of Children with Diarrhoea by Wealth Quintile**

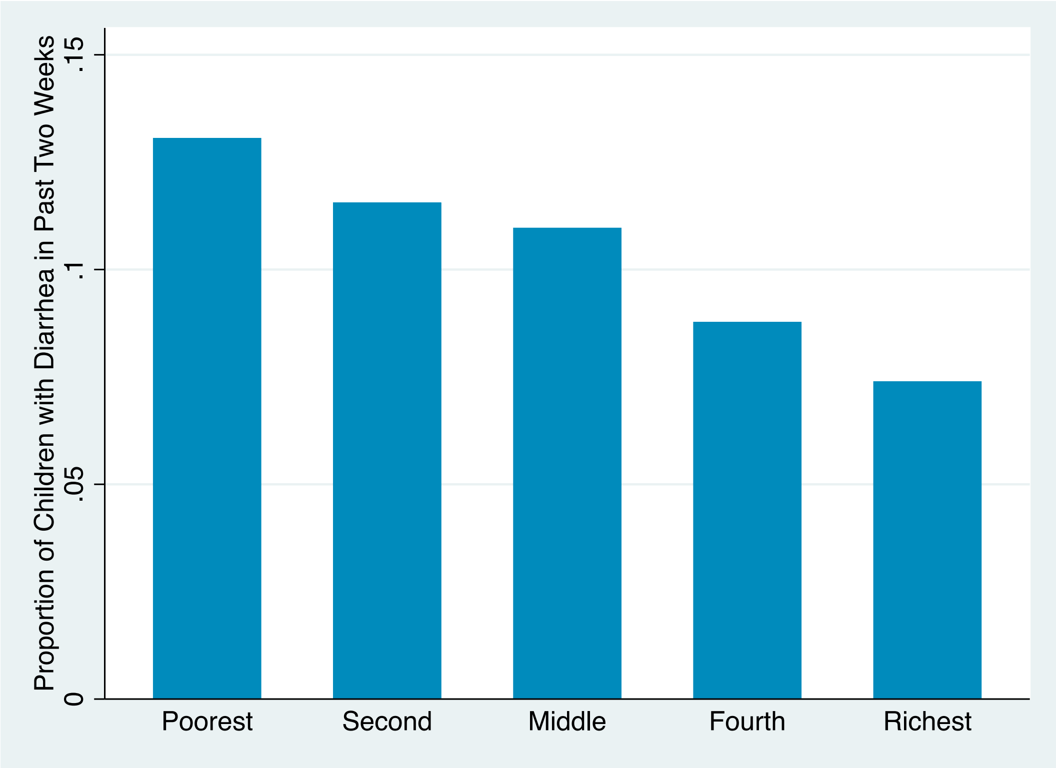
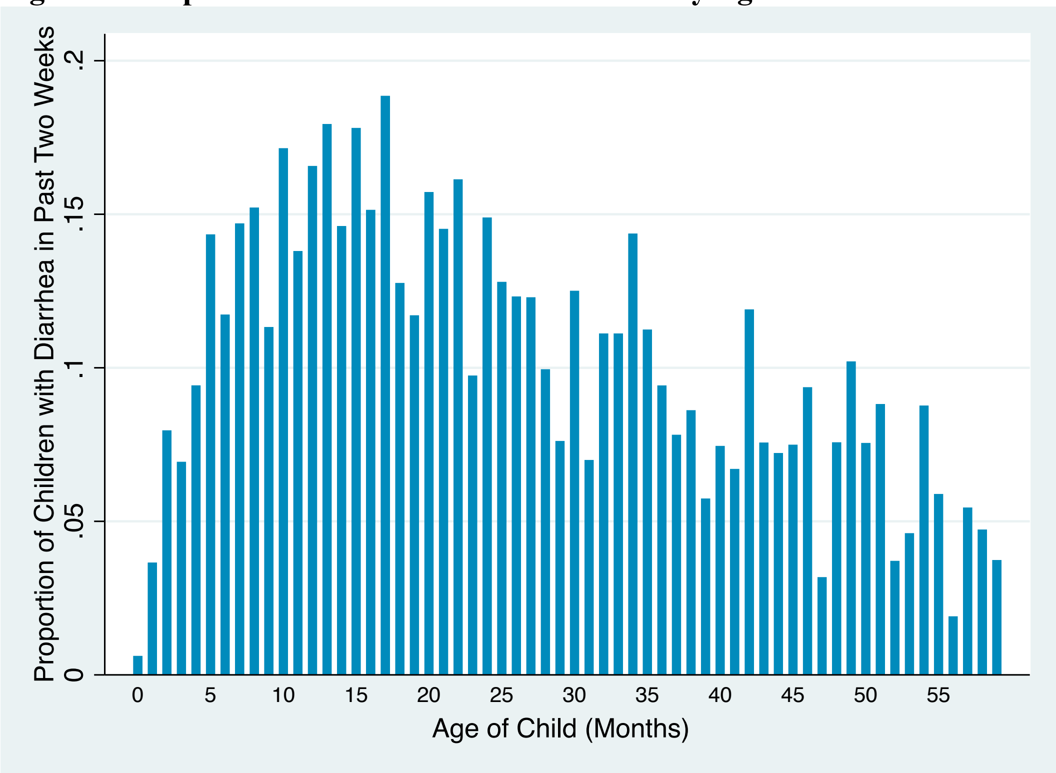



**Figure 2: Proportion of Children with Diarrhoea by Age**



### 3.3. Sanitation Classification and Prevalence

Following Günther and Fink 2010, I employ three classifications of sanitation variables:

1) **JMP Basic** – This definition follows that of the WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP) exactly, separating sanitation facilities into a simple improved/unimproved dichotomy. In the special case of flush to an unknown place, the assumption made by JMP is that the respondent may not know that the toilet is connected to a septic tank or sewer. In general, excreta are contained somewhere unexposed if the toilet is improved, reducing the likelihood of contact with drinking water.

2) **Technology Tiered** – Toilets are divided by underlying technology in three tiers: flush toilets, latrines, and open defecation. Note that "flush" indicates movement of waste to a sewer system or septic tank, "latrine" includes the case of flushed waste that ends up in a latrine, and "open defecation" includes the rare case of retention of faeces in a

**Table 1: Classification for Sanitation Explanatory Variables**

| Question WS7: "What kind of toilet facility do members of your household usually use?" | (1) JMP Basic | (2) Technology Tiered | (3) JMP Tiered |
|---|---|---|---|
| Flush to piped sewer system | improved | flush | improved |
| Flush to septic tank | improved | flush | improved |
| Flush to pit (latrine) | improved | latrine | improved |
| Flush to somewhere else | unimproved | latrine | unimproved |
| Flush to unknown place/not sure/DK | improved | excluded | excluded |
| Ventilated Improved Pit latrine (VIP) | improved | latrine | improved |
| Pit latrine with slab | improved | latrine | improved |
| Pit latrine without slab/open pit | unimproved | latrine | unimproved |
| Composting toilet | improved | latrine | improved |
| Hanging toilet/hanging latrine | unimproved | latrine | unimproved |
| Bucket | unimproved | open defecate | open defecate |
| No facilities or bush or field | unimproved | open defecate | open defecate |

bucket for regular removal. The special case of “flush to unknown place” is considered ambiguous since the respondent does not know the destination of waste, and it therefore cannot be consistently classified. Rather than imposing an assumption, the case is excluded since it encompasses only eight observations.

3) **JMP Tiered** – This is the same as the first definition (JMP Basic) except that two adjustments are made for comparability with Technology Tiered. First, a category of open defecation is added to distinguish unimproved sanitation without a toilet from unimproved toilets. Second, the special case of “flush to unknown case” is excluded.

Responses of “other” are dropped across all three classification systems. Although shared toilets are ordinarily not considered improved, I leave the shared/unshared distinction aside for purposes of classification and consider shared status alongside classification status rather than mediated through it. The means for each classification category are summarised in Table 1. Overall, 7.6 percent have flush technology and 62.3 percent have latrines; under JMP definitions these equal 37.2 percent improved and 32.7 percent unimproved. 30.1 percent of Nigerians sampled engage in open defecation.

Regardless of the classification system used, there are clear differentials in access to improved technology that are strongly related to household wealth. Richer Nigerian households possess better sanitation technologies and are far less likely to practice open defecation than their poorer counterparts. Figures 3 shows JMP Tiered sanitation by wealth quintile, and it also stands in for JMP Basic if the red and yellow bars are combined into a single lower tier. It reveals that over 80 percent of the richest quintile can access improved sanitation compared to a mere 16 percent of the poorest. Figure 4 shows that almost no one in the bottom three wealth quintiles can access flush sanitation, few in the fourth quintile can, but almost 40 percent of the richest can. Across all quintiles, the majority of people use latrines, so understanding whether latrines are a significant improvement over open defecation will clarify whether improved technology is needed for reducing incidence of diarrhoea among children.

**Figure 3: JMP Tiered Sanitation Classification by Wealth Quintile**

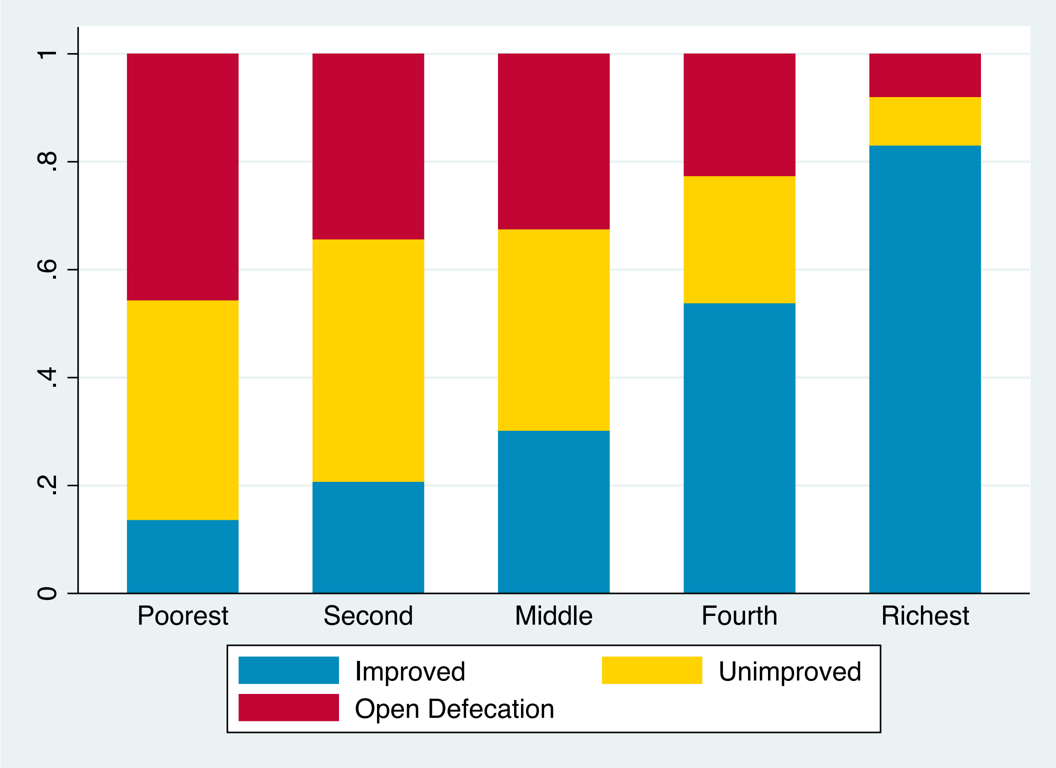
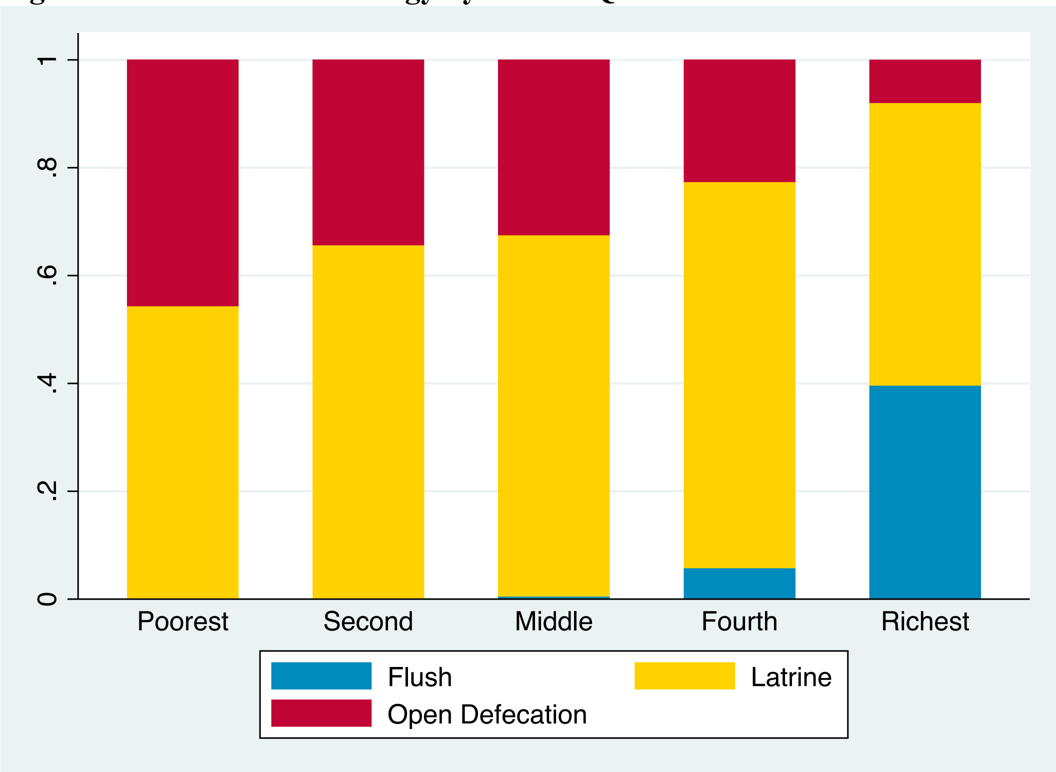



**Figure 4: Sanitation Technology by Wealth Quintile**



# 4. Models

### 4.1. Binary Response Variable

All models seek to explain the outcome of diarrhoea in children under age five in the two weeks preceding survey. I call this outcome $Diarrhoea_{icj}$, where throughout this paper the subscripts $i$, $c$, and $j$ indicate the household, cluster, and state, respectively. Since this is a binary response variable, linear models are inappropriate. A linear probability model, for example, predicts probabilities of diarrhoea that fall outside the interpretable range (i.e., 0–1) for various combinations of inputs.

Instead, I use a logit model, which involves assuming that the error term follows a logistic distribution. Both logit and probit contain response probabilities in the range 0–1, and both produce very similar estimates in each of my specifications. I opt to report logit results in keeping with Günther and Fink 2010, and this is the better choice for introducing fixed effects since the conditional logit is substantially easier to compute than conditional probit (usually called multinomial probit).

### 4.2. Key Explanatory Variables

As described above, I have classified sanitation technology using three different definitions: a basic one involving a single binary variable and two tiered ones involving two binary variables each. MICS-3 uses well-defined categories to assign a type to each household's sanitation technology, but inherent in the process of categorization is information loss: not all pit latrines with slabs are the same, for example, and in any case there is no check by the surveyor that all toilets described by respondents as pit latrines actually meet the JMP definition. So long as sanitation technologies are not systematic miscategorised, there is no reason for concern.

Using sanitation classifications that combine multiple sanitation types corrects for miscategorisation among the types combined together, and it also emphasises that the purpose of this analysis is not to make detailed comparisons between, for example, ventilated and non-ventilated latrines (Günther and Fink 2010, 11). While the use of classifications is appropriate, however, I cannot be sure ex ante that any one classification system is necessarily better than another, and this is why I have repeated all four basic models for each of three different classification systems.

For example, if a particular grouping placed highly statistically significant sanitation technologies in the same classification as nonsignificant technologies, it would tend to show an unduly insignificant effect for the former or an unduly significant effect for the latter. Testing various classifications of sanitation technology could alleviate this problem. If we believe that latrines are a substantial improvement over open defecation in terms of reducing a child's likelihood of suffering from diarrhoea, then we would predict a significant negative coefficient on $Latrine_{icj}$ and prefer the tiered classification systems. By contrast, if open defecation is relatively similar to other unimproved sanitation types in its relationship to child diarrhoea, we would prefer the JMP Basic classification.

Besides the classifications used by Günther and Fink, I include a variable to indicate that the respondent household shares its toilet with other households. Table 4 (Appendix) shows that 53 percent of sampled households share a toilet, and this may be expected to affect the outcome of child diarrhoea. On the one hand, shared facilities are a common good to which no particular household may feel a responsibility to contribute, potentially bringing users into contact with a greater concentration of faecal matter. On the other hand, shared facilities may be relatively better maintained than comparable private facilities, especially if public funds are allocated to their maintenance. Note that the case of "no facilities or bush," which accounts for nearly all of open defecation in the tiered categories, is considered "shared"; thus using a binary variable for shared status after controlling for sanitation type captures the effect of using higher tiers of sanitation, which are more evenly split between shared and unshared.

Taking the analysis of shared sanitation one step further, MICS-3 includes data about how many households a given household shares with. Figure 5 shows a sharp decline between one household (unshared) and two households, followed by a gradual decline in frequency with each additional household shared with until the data spikes upward again at ten or more households. The reason for this spike, unfortunately, is top-coding imposed by the MICS-3 questionnaire.

There are good reasons to question the accuracy of a respondent's answer when the number of shared households becomes large, but this measurement error would arguably be preferable to the loss of information from top-coding. The use of a top-coded variable without regard for its censored nature has been shown in general to introduce expansion bias, causing its importance in the regression to be overstated (Rigobon and

Stoker 2007, 1442). Indeed, in variants of the actual specification I tested (and have not reproduced here), the coefficient on this censored count of sharing households is highly statistically significant. Since top-coding involves censoring the upper tail of the underlying distribution, an important source of variation is been lost (Ibid., 1448). Dummy variables to indicate top-coded data do not solve the problem since "there is zero semiparametric information in the censored observations" (Ibid., 1464), and although it is theoretically possible to rectify the problem with additional data, this would go beyond the scope of the sanitation data collected in MICS-3.

**Figure 5: Number of Households with whom Respondent Family Shares Toilet**

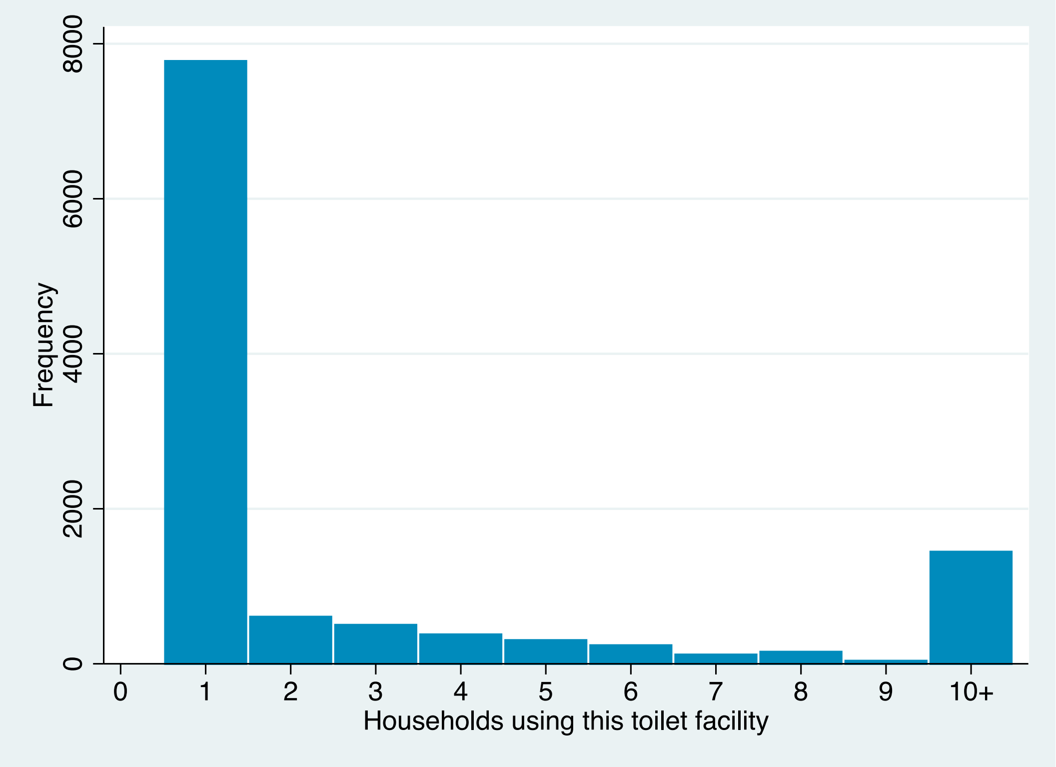



Figure 5 illustrates that nearly half of the households do not share facilities, and this provides a convenient breaking point. Even if shared facilities were more frequently used by 2–9 households, allowing us to disregard a small amount censored beyond 10, an indicator for shared status would still be useful. After all, we would still expect there to be a structural break between one and two households in terms of exposure to faecal matter since this is the threshold between a private and a public (or at least non-private) good, allowing us to assess the relative maintenance of each. Thus, I maintain the dummy for shared facilities rather than replacing it with a count of how many households share them.

Finally, to test whether parental sanitation behaviour might affect childhood diarrhoea, I preliminarily developed a score for how the mother or caretaker dealt with the child's waste based on the question, "The last time (child name) passed stools, what was done to dispose of the stools?" However, the design of MICS-3 stipulates that this question is only asked of children under three years old, which means that roughly 40 percent of the sample of children is excluded immediately. Since neither this score nor a household average of scores across different children in the household are statistically significant in logit specifications and since the inclusion of either or both significantly reduces the amount of data available for estimating parameters, driving up the standard errors, I omit them.

### 4.3. Control Variables

Control variables are variously included for their plausible correlation with possible omitted variables as well as for their direct relationship to the outcome of interest, child diarrhoea. One set of controls included for both reasons relates to drinking water. A household's water supply type may facilitate or inhibit the spread of diarrhoea, and it is known to be correlated with water quality, an unobserved variable in this paper. Using the same three classification standards, I categorise a household's water supply as (1) improved or unimproved under JMP Basic, (2) pump, pipe or surface water under Technology Tiered, and (3) improved, unimproved or surface water under JMP Tiered, the latter case being the omitted reference category for each classification.

Breastfeeding is widely believed to impart additional immunities to a child, so I follow Günther and Fink 2010 in including a dummy variable to indicate that the child is currently breastfeeding at the time of the survey. Likewise, I add a dummy variable to indicate that the child is female. In even-numbered specifications, I add interaction terms between female and breastfeeding as well as female and shared sanitation to allow for unequal effects on girls and boys from being currently breastfed and having access to a shared rather than an unshared toilet. In the case of breastfeeding, this interaction is motivated by research showing that mothers tend to stop breastfeeding girls and boys at different times, in particular that girls are breastfed for less time if parents have son preference and seek to have another child (Jayachandran and Kuziemko 2011, 1527); in the case of shared sanitation this is motivated by the possibility that parents might be more likely to let their sons use shared sanitation than their daughters or that men's and women's facilities may be maintained at different standards.

Mother education is included in the form of an ordered categorical variable indicating that the child's mother (or primary caretaker) has received no education (=1), some primary education (=2) or some secondary education (=3). This variable serves as a proxy for unobserved variables like the family's class, which may be related to the quality of the drinking water it can access and other omitted variables that may very considerably between households within an enumeration area. We might also suspect that it is positively correlated with the mother's knowledge about good sanitation practices.

All children in the MICS-3 survey are under five years old, and I use the child age variable, which varies from zero to 59 months with a mean of 30 in the sample (Table 4, Appendix). There is a degree of measurement error involved since some children's dates of birth are unknown, leaving their whole-number ages in years to be converted into months (12, 24, 36, 48, etc.) and thus shifted downward from the unobserved true values. I include squared and cubed terms versions of age along with the linear term to represent age as a cubic function within the logit model. Thought it is also highly statistically significant, I omit the quartic term in age for fear of over-fitting the data with a nonlinear model. Whereas stepping up from a quadratic to a cubic model of age noticeably affects some of the other coefficients in the model, stepping up from a cubic to a quartic makes little difference.

Additionally, I construct a dummy variable indicating vaccination against measles as a proxy for parental investment in child health. This variable combines information from the child's immunisation card (if available) with the mother or caretaker's recollection of vaccination and with further information from three nationwide vaccination campaigns conducted between September 2006 and January 2007. Measles vaccination is a good choice because it is widely available, and my use of it aligns with Günther and Fink.

To these control variables I add household characteristics, namely household size and indicators of wealth. Household size is the only control variable other than child age that is not binary, and it ranges from two to thirty-six household members. To indicate household wealth, Günther and Fink suggest considering a bundle of assets: electricity, radio, TV, bicycle and fridge as well as the material used to construct the house of residence (2010, 17). I depart slightly from their suggestion and omit construction material since this encompasses three different variables – one each for the floor, roof, and walls – and a wide range of answers for each. Since MICS-3 also includes wealth scores calculated by a more complex process for each household, I am able to substitute this for the asset bundle, and I find very little difference. As a result, I do not report tested specifications that use wealth scores instead of assets.

Furthermore, I include a dummy for the month in which a child was interviewed (indicating April against a base case of March) to account for the possibility of diarrhoeal outbreaks during the course of surveying that only affects households surveyed later. This could also detect an outbreak that was present at the beginning of the survey period but faded before the end.

Finally, I add a dummy variable indicating that a child is located in an urban enumeration area as opposed to the omitted case of a rural one. Just under one in four children sampled live in urban areas, according to Table 4 (Appendix), and the inclusion of this dummy controls for omitted variables common to all urban areas. Urban areas may be substantially different in their population density, for instance, leading to more frequent outbreaks of diarrhoea.

Since the optional MICS birth history module was not used in Nigeria, I cannot include the variables used by Günther and Fink for whether the child is the first born and for the length of time to the preceding birth if not. I also do not include mother age and marital status for lack of certainty that the person responding is actually the child's mother, as this person may be a caretaker.[2]

### 4.4. Clusters and Fixed Effects

As its name reminds us, MICS-3 is a clustered survey, meaning that households are sampled from within a limited number of clusters (enumeration areas) selected from each strata (state or National Capital Territory). As a result, we have reason to believe that a large number of children who come from the same cluster may share certain

[2] Recall that what I have referred to as "mother education" is also mother/caretaker education. I include this variable because I am more confident that this data is useful regardless of whether the respondent is a mother, as it proxies for social standing in the community and knowledge about sanitation practices that affect child exposure to faecal matter. By contrast, mother age and marital status are likely unrelated to caretaker age and marital status in the case where a child's mother and caretaker are different people.

idiosyncrasies not captured by the data. In particular, we must be wary of omitted variables that are constant among individuals within a group (Chamberlain 1980, 225).

Omitted variables that are constant at the family level may include sanitation practices like hand washing and food storage, diet and nutrition, and genetic factors such as immunological deficiency. Omitted variables constant at the group level may include shared environmental conditions, including common exposure to local outbreaks of diarrhoea. Without their inclusion, all of these factors could cause omitted variable bias, altering the coefficients reported in the logit regressions.

A conditional logit approach has the advantage over logit of being able to condition the probability of child diarrhoea on fixed effects. Unlike the random effects model, it does not impose that the effects $c_j$ are random, i.e $E(\boldsymbol{X}_{icj}c_j) = 0$, which would assume away the omitted variable bias we suspect (Chamberlain 1980, 233). In order to proceed, however, we need to decide which set of fixed effects is a more important control. Comparing my two potential sets of clusters, I find that household cluster size has a mean of 1.79 and a standard deviation of 0.90 in the sample. The size of household clusters shown in Figure 6 is highly skewed right, with each cluster size of one child larger occurring less than half as frequently as the preceding cluster size. Of 11,159 households clusters, over 7,000 contain just one child.

By contrast, the size of enumeration area clusters shown in Figure 7 forms a relatively more normal distribution with a large tail for outlier areas where more than thirty to forty children under five have been surveyed. Enumeration area clusters have a mean size of 14.93 and a standard deviation of 8.56 in the sample. Given the choice between household or enumeration area fixed effects, the latter are clearly more compelling to account for given the structure of the sampling. Household clusters only include children under five and are therefore very small, so the bias of omitted household factors on my estimates of interest is less pronounced.

The fixed effects model assumes that, conditioning on the cluster fixed effects, individuals within the cluster are independent of one another. While it is true that the *households* chosen from within each enumeration area are independent of one another, *children* are the unit of analysis here. Since all children in the household are surveyed, my two-stage sample design technically violates independence since children in the same

**Figure 6: Household Cluster Size**

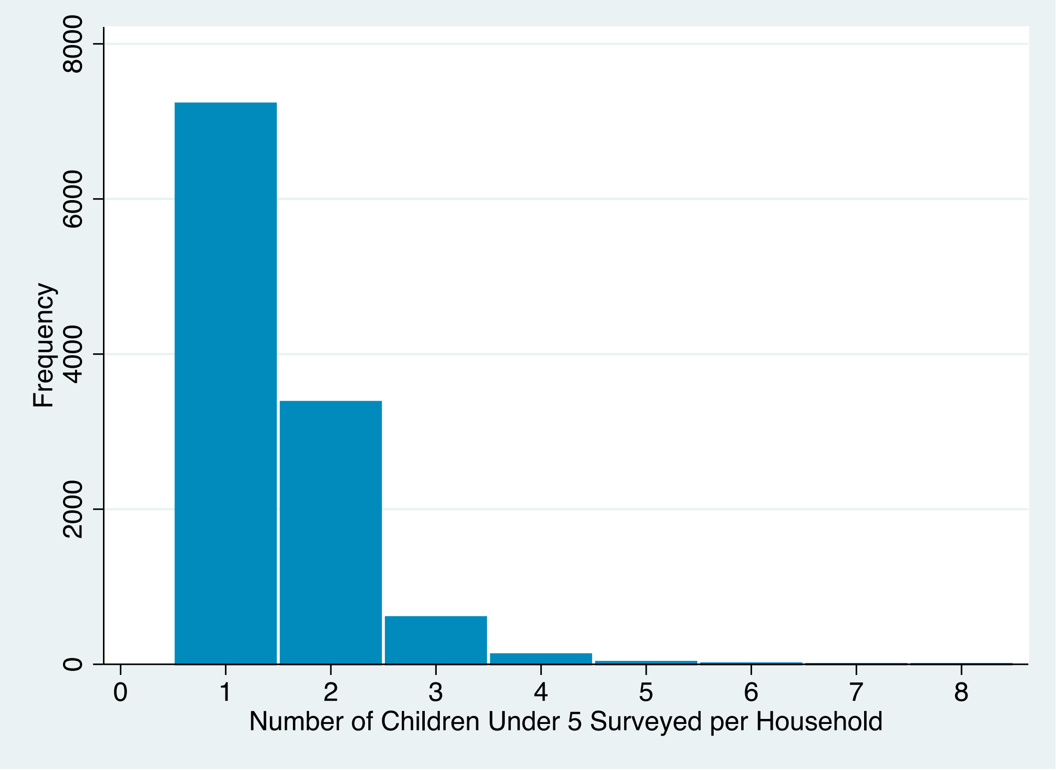
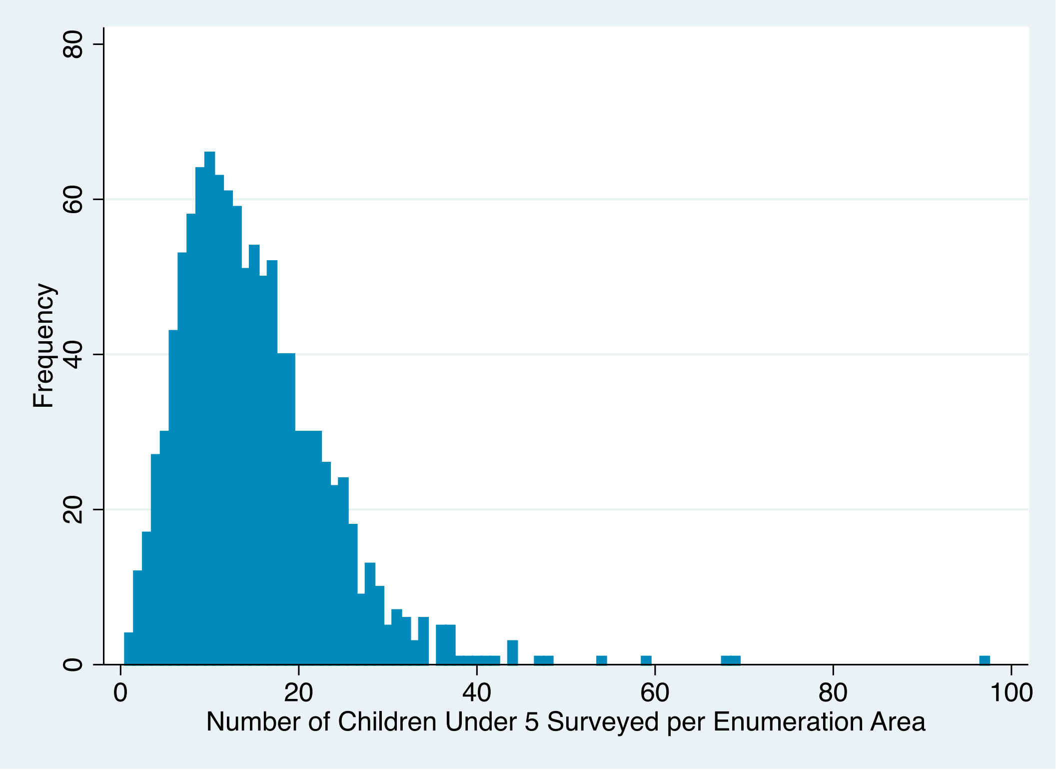



**Figure 7: Enumeration Area Cluster Size**




household are not independent of one another, even conditioning on the enumeration area fixed effect, since they share unobserved household characteristics. I could correct for this by further adding household fixed effects, but this would be computationally difficult and would massively reduce my sample size since households without variation in the dependent variable would be excluded.

### 4.5. Specifications

Given the three classifications of my explanatory variables, two versions of the controls variables (with and without interactions), and two regressions types (logit and conditional logit for enumeration area fixed effects), there are twelve specifications reported. The first four take the following form:

$$(1\text{–}2)\text{: } P(\mathit{Diarrhoea}_{icj}=1) = \frac{\exp(\alpha_1+\beta_1 \mathit{ImpS}_{icj}+\gamma_1 \mathit{ImpW}_{icj}+\boldsymbol{\delta_1 X}_{icj}+u_{icj})}{1+\exp(\alpha_1+\beta_1 \mathit{ImpS}_{icj}+\gamma_1 \mathit{ImpW}_{icj}+\boldsymbol{\delta_1 X}_{icj}+e_{icj})}$$

$$(3\text{–}4)\text{: } P(\mathit{Diarrhoea}_{icj}=1|c_j) = \frac{\exp(\alpha_1+\beta_1 \mathit{ImpS}_{icj}+\gamma_1 \mathit{ImpW}_{icj}+\boldsymbol{\delta_1 X}_{icj}+c_j+e_{icj})}{1+\exp(\alpha_1+\beta_1 \mathit{ImpS}_{icj}+\gamma_1 \mathit{ImpW}_{icj}+\boldsymbol{\delta_1 X}_{icj}+c_j+e_{icj})}$$

Each even-numbered specification is identical to the preceding odd-numbered specification save for the addition of explanatory interaction terms between $\mathit{Female}_{icj}$ and $\mathit{Shared}_{icj}$ and between $\mathit{Female}_{icj}$ and $\mathit{Breastfed}_{icj}$.

Specifications (5–8) and (9–12) are as above but with $\beta_1 \mathit{ImpS}_{icj}+\gamma_1 \mathit{ImpW}_{icj}$ replaced by $\beta_{21}\mathit{Flush}_{icj}+\beta_{22}\mathit{Latrine}_{icj}+\gamma_{21}\mathit{Pump}_{icj}+\gamma_{22}\mathit{Piped}_{icj}$ and $\beta_{31}\mathit{ImpS}_{icj}+\beta_{32}\mathit{UnimpS}_{icj}+\gamma_{31}\mathrm{ImpW}_{icj}+\gamma_{32}\mathit{UnimpW}_{icj}$ , respectively. Thus the first four specifications use JMP Basic definitions, the second four use Technology Tiered definitions, and the final four use JMP Tiered definitions.

Conditional logit specifications (3–4, 7–8, and 11–12) include fixed effects for each of the clusters $c_j$, leaving a reduced error term $e_{icj}=u_{icj}-c_j$. The dummy variable $\mathit{Urban}_j$ is dropped for all conditional logit specifications since households are classified as urban or rural at the cluster level, leaving no variation within clusters.

All models explicitly account for the stratified two-stage design in which states are the strata, enumeration areas are the primary sampling units, and households are the secondary sampling units. Since the sample is far from self-weighting at either the national or state level (UNICEF and National Bureau of Statistics 2007, 58), probability weights are calculated and incorporated in the sample design of all models. At the heart of the weights is the sampling fraction, which itself is the product of an enumeration area's probability of being sampled in its state and a household's probability of being sampled in its enumeration area. All children under five in a sampled household are included in the sample design, but the level of nonresponse (totalling 9.0 percent for the entire country) is accounted for in the weights as the ratio of completed child questionnaires to eligible children. To form the weights, the response rate and sampling fraction are multiplied, the inverse of this product is taken, and the results are normalised so that the sum of the sample units interviewed equals the total sample size for Nigeria (Ibid., 60-61).

Since thirty enumeration areas were drawn from each state regardless of the state's size, I might like to introduce a finite population correction of $\sqrt{\frac{N-n_i}{N-1}}$ for each strata, where $N$ is the population size (number of enumeration areas) and $n_i$ is the population of enumeration area $i$. Performing a finite population correction on the first stage was not possible, however, due to the unavailability of data on the total number of enumeration areas per state.[3] Consequently, the models must treat enumeration areas as if they were drawn with replacement from an infinite population. This assumptions is reasonable because the 30 selected areas very likely account for less than 10 percent of the total in every state (and less than 5 percent in all but one) under even the most generous assumptions.[4] The result of this assumption is that the sampling procedure in the second stage and any weights or finite population corrections at that stage need not be explicitly accounted for in the model.

All standard errors are robust in both logit and conditional logit specifications, and in the latter case all standard errors are clustered and account for possible correlation of other covariates with the cluster fixed effects, following Wooldridge (2003, 134-135). I now proceed to report and interpret the results of these specifications.

---

[3] The report by UNICEF and the National Bureau of Statistics (2007) that accompanies the Nigeria MICS-3 data makes no mention of how many enumeration areas each state contains. Extensive searching through electronic resources and library archival material from the 1991 census that demarcated the enumeration areas did not turn up the required listing, and attempts to contact Nigerian statistical officers responsible for MICS-3 went unanswered.

[4] At 205,299 people, the National Capital Territory of Abuja is the smallest of the 37 strata and less than a third of the size of the next largest state. Enumerations areas during the time of the 1991 Census were designed to have an estimated population of 200-250 persons in rural areas and 400-650 persons in urban areas (National Population Commission 1998, 9). Under the most generous interpretation that all enumeration areas in Abuja are urban and that the maximum 650 people per enumeration area are present, we would expect at least 316 enumeration areas, meaning that the thirty chosen were less than 10 percent of the total. This amounts to a finite population correction of 0.953 for Abuja and corrections closer to the infinite population assumption of 1 for all other states.

# 5. Results and Analysis

Table 2 shows the results of my twelve specifications with coefficients expressed as log odds ratios. In short, the major findings are that different sanitation classification schemes affect the degree of significance and size of the average marginal effect of their estimators, that there are substantial differences between results in logit and conditional logit models, and that shared sanitation appears to have a different effect on boys than on girls. Some of these results are counterintuitive, and I address each of these points in turn.

**5.1. Sanitation Coefficients in the Logit Models**

First, consider the findings from the six logit models. Here classification largely determines the degree of significance accorded to sanitation explanatory variables. Under the JMP Basic classification, the coefficient $\beta_1$ on improved sanitation is highly statistically significant (i.e., significant at the 1 percent level), meaning that a child under five whose household has an improved toilet has a lower probability of diarrhoea than one without, ceteris paribus. In both specifications (1) and (2), the average marginal effect of switching from unimproved to improved sanitation is a 2.1 percentage point decrease in the probability of a child having experienced diarrhoea in the past two weeks.[5]

Comparing with specifications (9) and (10), which use the JMP Tiered classification, I find very similar coefficients on all control variables. Here improved sanitation is defined to exclude a handful of ambiguous responses ("flush to unknown place") but the category is otherwise unchanged. The estimated coefficient $\beta_{31}$ is significant at the 10 percent level and provides an average marginal effect of 1.9 percentage point decrease in the probability of child diarrhoea for switching from open defecation to improved sanitation. This improves to a 2.1 percentage point decrease when switching from unimproved sanitation, though unimproved ($\beta_{32}$) is not statistically different from open defecation. That the lower two tiers are indistinguishable from one another suggests that JMP's basic dichotomy may be effective without modification.

[5] Given that most of the variables in my specifications are binary, I prefer the average marginal effect (AME), which computes the marginal effect in each case of the sample, to the marginal effect at the means (MEM), which computes the marginal effect in the case where all values are taken to be their means. In any event, the estimates are very similar, e.g., -.0208 for AME and -.0198 for MEM in (1).

**Table 2: Logit and Conditional Logit Regressions of Diarrhoea on Sanitation**

| | Dependent variable: Diarrhoea | | | | | |
|---|---|---|---|---|---|---|
| | logit | logit | clogit | clogit | logit | logit |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 1 Imp. San. | -0.245*** | -0.246*** | -0.00199 | -0.00325 | | |
| 1 Imp. Water | -0.124 | -0.126 | -0.0895 | -0.0959 | | |
| 2 Flush San. | | | | | -0.360* | -0.359* |
| 2 Latrine San. | | | | | -0.0829 | -0.0831 |
| 2 Pump Water | | | | | -0.0851 | -0.0836 |
| 2 Pipe Water | | | | | -0.0405 | -0.0421 |
| 3 Imp. San. | | | | | | |
| 3 Unimp. San. | | | | | | |
| 3 Imp. Water | | | | | | |
| 3 Unimp. Water | | | | | | |
| Female | -0.126** | -0.205** | -0.125* | -0.273** | -0.127** | -0.198** |
| Shared | -0.0978 | -0.199** | -0.109 | -0.259** | -0.102 | -0.198* |
| Female*Shared | | 0.216* | | 0.325** | | 0.204* |
| Breastfed | 0.341*** | 0.380*** | 0.15 | 0.176 | 0.344*** | 0.385*** |
| Female*Breastfed | | -0.0823 | | -0.054 | | -0.0857 |
| Mother Edu | -0.0622 | -0.0627 | -0.0212 | -0.0227 | -0.0741 | -0.0747 |
| Child Age | 0.162*** | 0.162*** | 0.174*** | 0.175*** | 0.162*** | 0.163*** |
| [Child Age]$^2$ | -0.00583*** | -0.00584*** | -0.00652*** | -0.00657*** | -0.00583*** | -0.00583*** |
| [Child Age]$^3$ | 5.60e-05*** | 5.61e-05*** | 6.37e-05*** | 6.42e-05*** | 5.59e-05*** | 5.59e-05*** |
| Vaccination | -0.067 | -0.0662 | -0.0973 | -0.0924 | -0.0792 | -0.0782 |
| Household Size | 0.00683 | 0.00691 | -0.0119 | -0.0115 | 0.00665 | 0.00671 |
| Electricity | 0.0213 | 0.0226 | -0.136 | -0.134 | -0.0346 | -0.0338 |
| Radio | 0.0808 | 0.0809 | -0.0542 | -0.0578 | 0.0934 | 0.0935 |
| TV | -0.0771 | -0.0777 | -0.0764 | -0.0781 | -0.0989 | -0.0997 |
| Bicycle | -0.039 | -0.0407 | 0.03 | 0.0304 | -0.0306 | -0.032 |
| Fridge | -0.157 | -0.159 | -0.107 | -0.107 | -0.129 | -0.132 |
| Survey Month | -0.0946 | -0.0937 | 0.43 | 0.43 | -0.0848 | -0.0839 |
| Urban | -0.114 | -0.115 | | | -0.135 | -0.136 |
| Constant | -2.909*** | -2.872*** | | | -2.884*** | -2.852*** |
| Observations | 15,330 | 15,330 | 10,346 | 10,346 | 15,330 | 15,330 |
| Model df | 18 | 20 | 17 | 19 | 20 | 22 |
| F | 10.14 | 9.158 | 8.62 | 7.788 | 9.45 | 8.637 |

*** p<0.01, ** p<0.05, * p<0.1

**Table 2: Logit and Conditional Logit Regressions of Diarrhoea on Sanitation (cont.)**

| | Dependent variable: Diarrhoea | | | | | |
|---|---|---|---|---|---|---|
| | clogit | clogit | logit | logit | clogit | clogit |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| 1 Imp. San. | | | | | | |
| 1 Imp. Water | | | | | | |
| 2 Flush San. | -0.0517 | -0.0455 | | | | |
| 2 Latrine San. | 0.0678 | 0.0689 | | | | |
| 2 Pump Water | 0.0861 | 0.082 | | | | |
| 2 Pipe Water | 0.134 | 0.123 | | | | |
| 3 Imp. San. | | | -0.229* | -0.229* | 0.0647 | 0.065 |
| 3 Unimp. San. | | | 0.0226 | 0.023 | 0.0816 | 0.0835 |
| 3 Imp. Water | | | -0.105 | -0.105 | 0.0557 | 0.0515 |
| 3 Unimp. Water | | | 0.0359 | 0.0397 | 0.224 | 0.228 |
| Female | -0.124* | -0.267** | -0.126** | -0.206** | -0.123* | -0.270** |
| Shared | -0.0908 | -0.237* | -0.0827 | -0.184 | -0.0805 | -0.231* |
| Female*Shared | | 0.318** | | 0.217* | | 0.326** |
| Breastfed | 0.155 | 0.183 | 0.343*** | 0.383*** | 0.156 | 0.185 |
| Female*Breastfed | | -0.0582 | | -0.0828 | | -0.0601 |
| Mother Edu | -0.0211 | -0.0231 | -0.0589 | -0.059 | -0.0229 | -0.0247 |
| Child Age | 0.175*** | 0.176*** | 0.162*** | 0.162*** | 0.174*** | 0.175*** |
| [Child Age]$^2$ | -0.00656*** | -0.00661*** | -0.00582*** | -0.00583*** | -0.00651*** | -0.00655*** |
| [Child Age]$^3$ | 6.41e-05*** | 6.46e-05*** | 5.60e-05*** | 5.61e-05*** | 6.35e-05*** | 6.40e-05*** |
| Vaccination | -0.101 | -0.0965 | -0.0629 | -0.0617 | -0.096 | -0.0911 |
| Household Size | -0.0116 | -0.0112 | 0.00698 | 0.00707 | -0.0115 | -0.0111 |
| Electricity | -0.149 | -0.146 | 0.0186 | 0.0198 | -0.144 | -0.142 |
| Radio | -0.0604 | -0.0643 | 0.0801 | 0.0802 | -0.0592 | -0.063 |
| TV | -0.0781 | -0.0801 | -0.0772 | -0.0778 | -0.0784 | -0.0803 |
| Bicycle | 0.0294 | 0.0298 | -0.0394 | -0.041 | 0.0312 | 0.0317 |
| Fridge | -0.0931 | -0.0938 | -0.156 | -0.159 | -0.1 | -0.1 |
| Survey Month | 0.441 | 0.442 | -0.0939 | -0.093 | 0.443 | 0.444 |
| Urban | | | -0.118 | -0.12 | | |
| Constant | | | -2.957*** | -2.924*** | | |
| Observations | 10,346 | 10,346 | 15,330 | 15,330 | 10,346 | 10,346 |
| Model df | 19 | 21 | 20 | 22 | 19 | 21 |
| F | 7.808 | 7.128 | 9.135 | 8.333 | 7.726 | 7.064 |

*** p<0.01, ** p<0.05, * p<0.1

The Technology Tiered classification system (5–6) also produces a significant result at the 10 percent level for its top tier ($\beta_{21}$), with a 3.1 percentage point decrease in probability of diarrhoea associated with flush technology. This is a 2.4 percentage point decrease relative to latrine sanitation, though once again this intermediate tier ($\beta_{22}$) is not statistically distinguishable from open defecation, the lowest tier and reference category. It is worth noting that there may not be enough variation in the Nigeria MICS-3 data to fully distinguish the contribution of flush technology from that of its covariates. Only 1,232 children in the sample (7.60 percent) lived in households with a flush toilet, and these households were disproportionately found in the top wealth quintile (see Figure 4). Fewer than 2 percent of children with access to flush toilets were classified in the bottom three wealth quintiles. The narrower category comes with proportionately greater standard error for $\beta_{21}$ and lower significance compared to JMP Basic, but by including only the two sanitation types whose users have the lowest incidence of diarrhoea, the average marginal effect associated with $Flush_{icj}$ is greater.[6]

The inclusion of interaction variables does not affect sanitation coefficients or their statistical significance under any classification system, suggesting that their addition does not control for omitted variable bias to the β coefficients on sanitation. If average marginal effects of -1.9, -2.1, or -3.1 percentage points for improved or flush sanitation sound small, consider that only 10.6 percent of the sample had experienced diarrhoea in two weeks before being surveyed (Table 4, Appendix). Exponentiating the log-odds ratios to obtain odds ratios and subtracting from 100 percent, my results equate to 21.8 to 30.2 percent decreases in the odds of a child having experienced diarrhoea in the last two weeks. This is in line with meta-analyses by Esrey et al. (1991) and Fewtrell et al. (2005) and higher than the 13 percent reduction in odds in the recent DHS meta-analysis, which the authors admit is lower than expected, possibly because of attenuation bias (Günther and Fink 2010, 20-21). One difference is my inability to find a significant effect for latrine sanitation, though my finding of 8.0 percent reduction in odds would be close to Günther and Fink's finding of a 7 percent reduction in odds had it been significant.

---

[6] In parallel logit specifications for (5) and (6) with the same variables that use heteroskedasticity-robust and clustered standard errors (at the enumeration area level) but do not account for the complex survey design, the coefficients on flush rise to statistically significance at the 5 percent level. I decline to consider results with erroneously low standard errors but note this to caution the reader against making imbalanced comparisons of these findings with analyses of data obtained by cluster sampling that, intentionally or not, treat their data as if it were derived from a simple random sample.

### 5.2 Comparison with the Conditional Logit Models

Logit and conditional logit specifications exhibit many similarities in their estimated coefficients for control variables. Across all specifications, $Female_{icj}$ is statistically significant at the 5 or 10 percent levels and associated with lower probability of a child experiencing diarrhoea. Likewise, $Age_{icj}$ is highly statistically significant (1 percent level) across all specifications, and it enters into the logit function as a cubic expression that peaks around 17–22 months, declines to a local minimum at about 50 months, and begins climbing again until the data stops at 59 months. Although the trough around the local minimum is exaggerated, this pattern corresponds to the prevalence of diarrhoea across different ages shown in Figure 2.

The inclusion of fixed effects leads to some substantially different results, however. In the logit models, $Breastfed_{icj}$ is highly statistically significant and associated with a higher probability of having experienced diarrhoea. This result is concerning because it is well established that breastfeeding causes children to consume less contaminated food and water, reducing the risk of diarrhoea, and that breastfeeding is associated with lower rates of infant mortality from diarrhoea (Jayachandran and Kuziemko 2011, 1517). This effect is erased in the conditional logit models, and it suggests that enumeration areas with higher levels of breastfeeding also experienced higher levels of diarrhoea. The coefficients had been biased upward and are at least partly corrected by the inclusion of fixed effects.[7]

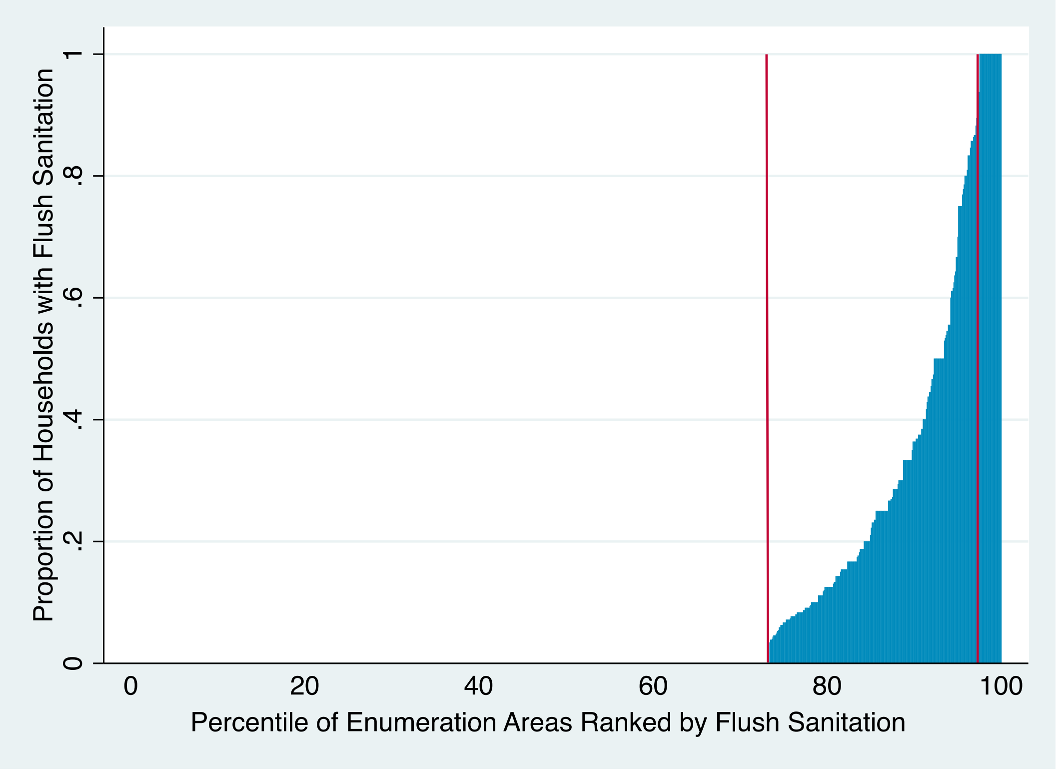
Under conditional logit, none of the sanitation classifications produce statistically significant coefficients, meaning that sanitation type and child diarrhoea are unrelated after controlling for the enumeration area in which a child lives. In light of the findings for $Breastfed_{icj}$, it is likely that the benefits to sanitation were overstated, too. However, it is possible that there is too little variation within clusters to trust these conditional logit results. For the vast majority of children – every child in 841 of the 1107 enumeration areas – the enumeration area perfectly predicts whether a flush toilet is used by the

---

[7] Water quality, for example, is unobserved, varies by enumeration area, and is likely related to child diarrhoea. Including variables for the type of water system (e.g., pump vs. pipe) cannot fully control for the quality of water delivered by that system. Women may choose to breastfeed longer in areas where water is poor to protect their children, who still receive higher exposure to faecal matter through solid food prepared with poor water. A more detailed investigation is needed to better understand the determinants of a mother’s breastfeeding status.

household. This leaves only 266 clusters (3,522 children from a sample of over 16,000) to explain how differential access to flush toilets would affect child diarrhoea. Figure 8 shows that over 76 percent of enumeration areas do not have any flush toilets, 1.7 percent all have only flush toilets, and less than 22 percent (the band between two red bars) have some variation. Regressors that vary little across observations within a cluster or panel are generally estimated imprecisely by fixed effects models.

**Figure 8: Concentration of Flush Technology in Few Enumeration Areas**




Another drawback evident from the bottom of Table 2 is that conditional logit models use only 67.5 percent of the observations used in the logit model. This is because all observations from enumeration areas that lack variation in the dependent variable (i.e., where all children in the cluster have one of either $Diarrhoea_{icj}$=1 or $Diarrhoea_{icj}$=0) are dropped since the outcome would be perfectly predicted by the enumeration area fixed effect.[8] The smaller sample size resulting from this dropping of data tends to drive up the standard errors in my conditional logit specifications, possibly reducing the significance of explanatory variables of interest. Sanitation coefficients near zero rather than large

[8] The vast majority are enumeration areas with no cases of diarrhoea among surveyed children. The only other dropped case came from an enumeration area with only a single child surveyed and that child reporting diarrhoea.

standard errors render the conditional logit results insignificant, however, and I am more concerned about possible bias introduced by excluding close to one-third of the data. This points to the usefulness in future data collection of documenting more instances of infrequent events like diarrhoea, possibly by checking on children multiple times over a longer period of surveying and recording each instance of diarrhoea.

Despite these drawbacks, cluster fixed effects cannot be ignored without overstating the effect of sanitation. While the logit specifications are useful baselines for comparison with other findings, they are omitting cluster fixed effects that alter the estimated sanitation coefficients. The fact that none of the reviewed literature tests specifications with fixed effects is reason to question the reliability of current findings. Günther and Fink are unable to implement even regional (sub-national) clusters in their logit models, instead comparing versions of a linear probability model with and without regional fixed effects as a check on their estimates (2010, 19). I repeat this exercise in Table 5 (Appendix) with the MICS-3 data to illustrate that when controlling for fixed effects for the 37 states of Nigeria, few of my sanitation coefficients change dramatically, as they have in Table 2 with the introduction of enumeration area clusters.[9] Region fixed effects are a poor substitute for cluster fixed effects, and failure to use the latter would seem to call into question the validity of highly significant findings prevailing today.

### 5.3 Shared Facilities Coefficients and the Gender Differential

Unlike the coefficients on sanitation categories, the binary indicator of shared sanitation provides a set of consistent results across logit and conditional logit models and across different sanitation classifications schemes. Challenging the assumption by the WHO that shared facilities are inferior to equivalent private facilities and would increase the probability of a child experiencing diarrhoea, my estimations with interactions find that shared facilities generally *reduce* this probability for boys and have a statistically significantly different effect for girls, one that is not distinguishable from zero. The negative coefficients on $Shared_{icj}$ in the even-numbered specifications of Table 2 provide evidence that boys experience a reduced likelihood of diarrhoea when their household

[9] All sanitation coefficients are downgraded in significance in Table 5 when state fixed effects are added, but all keep the same sign and more than half of their original magnitude. With higher standard errors from a larger sample or relaxed assumptions about the survey design, they would likely be significant. By contrast, the enumeration area cluster fixed effects remove significance from the results.

shares its toilets, an 18.0 to 22.8 percent reduction in odds of diarrhoea when one's facilities are shared rather than unshared, ceteris paribus. Note that this comparison effectively excludes open defecation, which is always indicated as shared, meaning it compares private flush to shared latrines, private "improved" to shared "improved", etc.

Furthermore, Table 2 gives a coefficient for the interaction term $Female_{icj}$ $*Shared_{icj}$ that is significant at the 5 or 10 percent level in every specification, indicating that girls are affected differently by shared sanitation. To assess this effect, I test the hypothesis that this coefficient plus that on $Shared_{icj}$ sum to zero. Table 3 shows that I cannot reject this hypothesis, meaning there is no statistically significant effect on girls of having shared facilities. Estimates of the coefficient on $Shared_{icj}$ in the odd-numbered specifications, i.e., *without* interaction terms, are consistently negative but statistically insignificant, suggesting that the different effects of shared sanitation on boys and girls counteract one another in the combined sample.

**Table 3: Adjusted Wald Test of Significance of Shared Facilities for Girls**

| | Test $H_0$: $\delta^{Shared} + \delta^{Female*Shared}=0$ vs $H_A$: $\delta^{Shared} + \delta^{Female*Shared}\neq 0$ | | | | | |
|---|---|---|---|---|---|---|
| Spec. | (2) | (4) | (6) | (8) | (10) | (12) |
| F | 0.03 | 0.27 | 0.04 | 0.29 | 0 | 0.42 |
| Prob > F | 0.8523 | 0.606 | 0.8364 | 0.5879 | 0.9933 | 0.5182 |

*** p<0.01, ** p<0.05, * p<0.1

What could explain this curious result, which survives even the cluster fixed effects that render my sanitation coefficients insignificant? Any satisfactory explanation of the gender differential must account for the young age of the children. One possibility is that shared facilities are generally better maintained, but young girls do not get to use them. Facilities shared by a few families might be maintained better to avoid upsetting other families, and public facilities shared by the community might be better maintained because someone is contracted to perform this service and because soap and running water are provided to remove faecal matter. Especially if toilets are public, norms may prevent girls from using toilets unaccompanied, meaning in effect that they practice open defecation despite the reported availability of a toilet facility. This framing relies on the assumption that girls enjoy a natural advantage of lower rates of diarrhoea (represented by the negative coefficient on $Female_{icj}$) and that boys "catch up" to them in terms of lower diarrhoea incidence only when they have differential access to shared facilities.

Without measures of toilet usage by gender or toilet cleanliness by shared status, I cannot test these hypotheses. Clues may be more readily gleaned from qualitative investigation into how young boys and girls use private and shared facilities differently. There is also heterogeneity in how many families share facilities that is not fully explored due to top-coding, as explained in section 4.2. Nonetheless, the recognition of this gender differential for shared toilets and diarrhoea incidence is important in that it reframes the discussion about whether sanitation improvements should come through private or public facilities as a conversation about which gender's needs are to be prioritised.

# 6. Conclusion

To understand the operation of the income-health gradient in Nigeria, I have examined the differential effect of sanitation technologies on diarrhoea incidence in children under five using logit and conditional logit models. Benefits found in the logit models include a reduction in odds of diarrhoea by roughly 22 percent for improved sanitation under JMP Basic or Tiered classification, by 30.2 percent for flush toilets, and by 18.0 to 22.8 percent across all categories for boys using a shared toilet. Under conditional logit specifications that account for omitted variables that are constant at the enumeration area level, the benefits to shared facilities remain while the sanitation type benefits disappear. While there is reason to believe that limited intracluster variation in diarrhoea incidence and flush toilet availability may contribute to non-significance in this model, the coefficients are so reduced in magnitude that omitted variable bias is surely present without the fixed effects.

My findings therefore give reason for concern about the present state of analysis of data on child diarrhoea and sanitation, which fails to account for cluster fixed effects. Care must be taken to account for the stratified, clustered nature of survey designs, to use population weights, and to control for omitted variables. MICS and DHS both use cluster sampling, and greater econometric sophistication is needed if this wealth of observational data is to shed further light on the relationship between sanitation and diarrhoea.

While limited to the case of Nigeria, my findings also suggests a greater role for shared facilities in efforts to achieve the Millennium Development Goals. As many developing countries experience rapid urbanisation, JMP has begun to acknowledge the role that shared sanitation is having in cities (WHO/UNICEF JMP 2012, 21). Even if the health benefit to boys of shared facilities does not withstand future investigation, the absence of demonstrable benefits to unshared facilities relative to shared would favor construction of more cost-effective public facilities where there is no improved sanitation.

A major limitation of this investigation is that it relies upon observational data. Every effort has been made to control for variables such as wealth (through assets), investment in child health (through vaccination and breastfeeding), sanitation practices (through mother education), family structure (through household size), child characteristics (through age, gender, and gender's interactions with shared facilities and breastfeeding), and geography (through urban/rural location and cluster fixed effects). Inevitably there are still omitted variables that could plausibly be biasing my estimates of the coefficients of interest; the question is whether I have adequately controlled for them. If there are variations within the cluster, say between an affluent neighbourhood and a slum in an urban area or between two villages in a rural area, these would not be controlled for by my fixed effects.

I am not able to measure the quality of water that each child drinks, for example, even though this is almost certainly a determinant of diarrhoea and is likely to be correlated with the type of sanitation the child uses.[10] Some variation in water quality is very likely caused by sanitation type and therefore endogenous to the model, as open defecation potentially exposes oneself and one's community to faecal matter and viruses or bacteria that are associated with low water quality. On the other hand, at least some variation is surely exogenous, for example from contamination of water upstream by livestock. Depending on the direction of correlation with the explanatory variables, this omitted variable would bias my reported coefficients upward or downward. Sanitation type may be virtually irrelevant where water quality is exogenously poor, whereas it may play a far larger role in determining diarrhoea incidence where water quality is good. By omitting this variable, therefore, my specifications could be masking the true nature of sanitation's contribution to well-being.

An ambitious mode of future investigation would be to conduct a randomised controlled trial in which a particular model of improved toilet is provided to a randomly selected treatment group. In a sufficiently large sample, the treatment and control groups would be indistinguishable except for the fact that one has been offered the toilet at no cost, and this would solve the omitted variables problem described above. Researchers affiliated with the Abdul Latif Jameel Poverty Action Lab are already conducting such an evaluation in the slums of Orissa, India, to determine whether community-governed and privately run public facilities have better success in improving sanitation outcomes like childhood diarrhoea (Barnhardt, Chevalier and Mobarak).

[10] This problem is acknowledged by JMP as the motivation for classifying water sources as improved: "One major information gap is the safety of drinking water supplies. […] cost-effective, periodic and standardised water quality testing was not possible on a global scale when the MDG target was formulated." However, relevant data may emerge in the near future: "The main international household surveys – MICS and DHS – are piloting the inclusion of a water-quality module that will include testing for the presence of *E. coli*. This is made feasible in part by the availability of new, rapid, low-cost water quality testing kits." (WHO/UNICEF JMP 2012, 5)

Another possibility for a randomised controlled trial would be to offer individual facilities installed in each house to people in some communities and shared facilities with regular cleaning provided in others. Although the short-term effect of each type on child diarrhoea might become evident very quickly, the two types of sanitation might have different effects over time, for example, as private toilets break down and are not repaired while public facilities continue to be maintained. Working with particular models of toilet procured at a given price, the researcher could produce reliable cost estimates that simply are not possible with the observational data from MICS-3.

An immediate starting point for extending this analysis, however, is to repeat it with other countries to see which if any results hold. Of particular interest in the MICS-3 round of data collection is Bangladesh, another very populous country with a dire sanitation record. Care should be given to make sure there is sufficient variation in the diarrhoea outcome, as surveys in which the vast majority of children have not experienced the condition in the past two weeks can tell us little about diarrhoea's relationship to basic sanitation. The low frequency of diarrhoea in the data, about 10.6 percent of the sample, is a limitation that should motivate future research to examine incidence of diarrhoea over a longer period. Even if a country faces a serious problem of morbidity from diarrhoea, this may not be reflected in the results if surveying is conducted in a season with relatively few outbreaks.

The landscape of sanitation technologies is rapidly changing, and the sanitation categories may quickly begin to appear too crude as new varieties of toilet spring up and receive widespread adoption. The Bill and Melinda Gates Foundation's campaign to "Reinvent the Toilet" aims to make basic sanitation less expensive and more sustainable and useful to its users by engineering a technological breakthrough. Already it is funding the testing of a prototype that breaks waste into component parts, generates hydrogen for fuel, generates chlorine to disinfect the toilet, and turns waste into safe fertiliser (*Economist* 2012). In the future, piped water and sewage may no longer be considered the ideal sanitation system. Indeed, a finer analysis to extend this work on MICS-3 with additional data might consider the negative externality of piped waste that is untreated to those in other communities who come into contact with it, as sewers may not necessarily be directing wastewater to a treatment facility.

The problem of differential access to improved sanitation is urgent, and economists should not wait for a technological breakthrough to solve it. Our econometric methods and disciplinary insights into the nature of income differentials, public goods, and economic development have so much to contribute to understanding and overcoming this problem, and I very much hope the present analysis of the Nigerian case has successfully questioned common assumptions about sanitation and diarrhoea and framed the next steps for investigation.

# References


Barnhardt, Sharon, Judy Chevalier and Mushfiq Mobarak. "Communal Sanitation Solutions for Urban Slums in Orissa, India." http://www.povertyactionlab.org/evaluation/communal-sanitation-solutions-urban-slums-orissa-india (accessed August 30, 2012).

Case, Ann, Darren Lubotsky and Christina Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review* 92(5): 1308-1334.

Central Intelligence Agency. 2012. The World Factbook. https://www.cia.gov/library/publications/the-world-factbook/geos/ni.html (accessed August 6, 2012).

Chamberlain, Gary. 1980. Analysis of Covariance with Qualitative Data. *Review of Economic Studies* 47: 225-238.

Cutler, David, Angus Deaton and Adriana Lleras-Muney. 2006. "The Determinants of Mortality." *Journal of Economic Perspectives* 20(3): 97-120.

*Economist*. 2012. "Flushed with Pride." August 25. http://www.economist.com/blogs/babbage/2012/08/sanitation (accessed September 10, 2012).

Esrey, S., J. Potash, L. Roberts and C. Shiff. 1991. "Effects of Improved Water Supply and Sanitation on Ascariasis, Diarrhoea, Dracunculiasis, Hookworm Infection, Schistosomiasis, and Trachoma." *Bulletin of the World Health Organization*, 69 (5): 609-621.

Fewtrell, Lorna, Rachel B. Kaufmann, David Kay, Wayne Enanoria, Laurence Haller, and John M Colford Jr. 2005. "Water, Sanitation, and Hygiene Interventions to Reduce Diarrhoea in Less Developed Countries: A Systematic Review and Meta-Analysis." *Lancet Infectious Diseases*, 5: 42-52.

Günther, Isabel, and Günther Fink. 2010. "Water, Sanitation and Children's Health: Evidence from 172 DHS Surveys." World Bank Policy Research Working Paper 5275.

Jayachandran, Seema, and Ilyana Kuziemko. 2011. "Why Do Mothers Breastfeed Girls Less than Boys? Evidence and Implications for Child Health in India." *Quarterly Journal of Economics* 126, 1485–1538.

Mobarak, Ahmed Mushfiq, Puneet Dwivedi, Robert Bailis, Lynn Hildemann, and Grant Miller. 2012. "The Low Demand for New Cookstove Technologies." *Proceedings of the National Academy of Sciences* 109(27): 10815-20.

National Bureau of Statistics (Nigeria). 2010. *Annual Abstract of Statistics, 2010*. Abuja.

National Population Commission (Nigeria). 1998. *1991 Population Census of the Federal Republic of Nigeria: Analytical Report at the National Level*. Abuja.

Oster, Emily, and Rebecca Thornton. 2011. "Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation." *American Economic Journal:* Applied Economics 3 (January): 91–100.

Rigobon, Roberto and Thomas M. Stoker. 2007. "Estimation with Censored Regressors: Basic Issues." *International Economic Review* 48(4): 1441-1467.

Strauss, John, and Duncan Thomas. 1998. "Health, Nutrition, and Economic Development." *Journal of Economic Literature* 36(2): 766-817.

United Nations. 2010. *The Millennium Development Goals Report 2010*. New York. http://www.un.org/millenniumgoals/pdf/MDG Report 2010 En r15 -low res 20100615 -.pdf#page=60 (accessed February 23, 2012).

United Nations Children's Fund (UNICEF). 2010. “Water, Sanitation and Hygiene in Nigeria.” April 2010. http://www.unicef.org/nigeria/ ng_media_WASH_fact_sheet_Apr_2010.pdf (accessed June 9, 2012).

UNICEF and National Bureau of Statistics (Nigeria). 2007. *Nigeria Multiple Indicator Cluster Survey 2007*. New York: UNICEF.

World Health Organization (WHO). 2009. “Diarrhoeal Disease.” http://www.who.int/ mediacentre/factsheets/fs330/en/index.html (accessed June 16, 2012).

WHO/UNICEF Joint Monitoring Programme (JMP) for Water Supply and Sanitation. 2010. *Progress on Sanitation and Drinking-water: 2010 Update*. Geneva: WHO Press.

——. 2012. *Progress on Sanitation and Drinking-water: 2012 Update*. Geneva: WHO Press.

——. “Types of Drinking-water Sources and Sanitation.” http://www.wssinfo.org/ definitions-methods/watsan-categories/ (accessed June 24, 2012).

Wooldridge, Jeffrey M. 2003. “Cluster-Sample Methods in Applied Econometrics.” *American Economic Review* 93(2):133-138.

# Appendix

**Table 4: Sample Summary Statistics**

| Variable | Mean | Std. Dev. | Min. | Max. | Obs. |
|---|---|---|---|---|---|
| Diarrhoea | 0.1062802 | 0.3082054 | 0 | 1 | 16146 |
| 1 Imp. San. | 0.3721791 | 0.4834007 | 0 | 1 | 16218 |
| 1 Imp. Water | 0.4553582 | 0.4980185 | 0 | 1 | 16218 |
| 2 Latrine San. | 0.6232572 | 0.4845845 | 0 | 1 | 16210 |
| 2 Flush San. | 0.0760025 | 0.2650102 | 0 | 1 | 16210 |
| 2 Pump Water | 0.6115427 | 0.4874145 | 0 | 1 | 16218 |
| 2 Pipe Water | 0.1085214 | 0.3110474 | 0 | 1 | 16218 |
| 3 Imp. San. | 0.3718692 | 0.4833186 | 0 | 1 | 16210 |
| 3 Unimp. San. | 0.3273905 | 0.4692756 | 0 | 1 | 16210 |
| 3 Imp. Water | 0.4553582 | 0.4980185 | 0 | 1 | 16218 |
| 3 Unimp. Water | 0.2935627 | 0.455408 | 0 | 1 | 16218 |
| Female | 0.4906568 | 0.4999281 | 0 | 1 | 16215 |
| Shared | 0.5321248 | 0.4989823 | 0 | 1 | 16218 |
| Breastfed | 0.3439032 | 0.475024 | 0 | 1 | 16115 |
| Mother Education | 1.722916 | 0.8341583 | 1 | 3 | 15775 |
| Child Age (Months) | 27.97089 | 16.52545 | 0 | 59 | 16215 |
| Vaccine | 0.4552886 | 0.4980125 | 0 | 1 | 15902 |
| Household Size | 6.736219 | 3.254527 | 2 | 36 | 16218 |
| Electricity | 0.4088667 | 0.4916397 | 0 | 1 | 16218 |
| Radio | 0.7507707 | 0.4325801 | 0 | 1 | 16218 |
| TV | 0.2983105 | 0.4575306 | 0 | 1 | 16218 |
| Bicycle | 0.3630534 | 0.4808949 | 0 | 1 | 16218 |
| Fridge | 0.1289925 | 0.3352019 | 0 | 1 | 16218 |
| Survey Month | 0.2336109 | 0.4231405 | 0 | 1 | 16215 |
| Urban | 0.2478111 | 0.4317548 | 0 | 1 | 16218 |

**Table 5: Linear Probability Model as Robustness Check**

| | Dependent variable: Diarrhoea | | | | | |
|---|---|---|---|---|---|---|
| | LPM | LPM | LPM | LPM | LPM | LPM |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 1 Imp. San. | -0.0204*** | | | -0.0134* | | |
| 1 Imp. Water | -0.011 | | | -0.00589 | | |
| 2 Flush San. | | -0.0224 | | | -0.0124 | |
| 2 Latrine San. | | -0.00631 | | | -0.00884 | |
| 2 Pump Water | | -0.00714 | | | -0.00755 | |
| 2 Pipe Water | | -0.00373 | | | -0.000968 | |
| 3 Imp. San. | | | -0.0177* | | | -0.0155 |
| 3 Unimp. San. | | | 0.00404 | | | -0.00296 |
| 3 Imp. Water | | | -0.00854 | | | -0.0052 |
| 3 Unimp. Water | | | 0.00483 | | | 0.0011 |
| Female | -0.0182** | -0.0176* | -0.0160* | -0.00806 | -0.00887 | -0.0091 |
| Shared | 0.0191* | 0.0183* | 0.0192* | 0.0194* | 0.0191* | 0.0195* |
| Female*Shared | -0.00542 | -0.00658 | -0.00499 | 0.00183 | 0.00136 | 0.00186 |
| Breastfed | 0.0143*** | 0.0144*** | 0.0143*** | 0.0137*** | 0.0138*** | 0.0137*** |
| Female*Breastfed | -5.12e-04*** | -5.13e-04*** | -5.11e-04*** | -4.97e-04*** | -4.97e-04*** | -4.97e-04*** |
| Mother Edu | 4.94e-06*** | 4.94e-06*** | 4.94e-06*** | 4.81e-06*** | 4.81e-06*** | 4.81e-06*** |
| Child Age | -0.0173** | -0.0168** | -0.0173** | -0.0172** | -0.0170** | -0.0172** |
| [Child Age]$^2$ | 0.0431*** | 0.0436*** | 0.0433*** | 0.0368*** | 0.0369*** | 0.0369*** |
| [Child Age]$^3$ | -0.0113 | -0.0115 | -0.0113 | -0.00908 | -0.00891 | -0.0091 |
| Vaccination | -0.00563 | -0.00668 | -0.00505 | 0.00556 | 0.00494 | 0.00555 |
| Household Size | 0.000518 | 0.000523 | 0.000538 | -0.000847 | -0.000869 | -0.000855 |
| Electricity | 0.00128 | -0.0037 | 0.000811 | 0.0104 | 0.00862 | 0.0107 |
| Radio | 0.00736 | 0.00841 | 0.00728 | 0.00456 | 0.00515 | 0.00457 |
| TV | -0.00701 | -0.00908 | -0.00699 | -0.00332 | -0.00444 | -0.00325 |
| Bicycle | -0.00347 | -0.0027 | -0.00354 | -0.00483 | -0.00423 | -0.00479 |
| Fridge | -0.00946 | -0.0082 | -0.00932 | -0.00617 | -0.00701 | -0.00628 |
| Survey Month | -0.00843 | -0.0076 | -0.0084 | -0.0148* | -0.0145* | -0.0148* |
| Urban | -0.00811 | -0.0105 | -0.00873 | -0.0114 | -0.0138 | -0.0111 |
| Constant | 0.0405** | 0.0409** | 0.0336* | 0.00643 | 0.0112 | 0.00829 |
| State FE | no | no | no | yes | yes | yes |
| Observations | 15,330 | 15,330 | 15,330 | 15,330 | 15,330 | 15,330 |
| Model df | 20 | 22 | 22 | 56 | 58 | 58 |
| F | 8.886 | 8.246 | 8.067 | 7.626 | 7.481 | 7.399 |

*** p<0.01, ** p<0.05, * p<0.1