Does improved sanitation reduce diarrhea in children in rural India?

Kumar, Santosh and Vollmer, Sebastian

Harvard School of Public Health, Harvard School of Public Health

21 March 2011

Does Improved Sanitation Reduce Diarrhea in Children in Rural India?

Santosh Kumar *   Sebastian Vollmer†

This Draft: March 11, 2011

Abstract

Nearly nine million children under five years of age die annually. Diarrhea is considered to be the second leading cause of Under-5 mortality in developing countries. About one out of five deaths are caused by diarrhea. In this paper, we use the newly available data set DLHS-3 to quantify the impact of access to improved sanitation on diarrheal morbidity for children under five years of age in India. Using Propensity Score Matching (PSM) and propensity-based weighted regression, we find that access to improved sanitation reduces the risk of contracting diarrhea. Access to improved sanitation decreases child diarrhea incidence by 2.2 percentage points. There is considerable heterogeneity in the impacts of improved sanitation. We neither find statistically significant treatment effects for children in poor household nor for girls, however, boys and high socioeconomic status (SES) children experienced larger treatment effects. The results show that it is important to complement public policies on sanitation with policies that alleviate poverty, improve parent's education and promote gender equity.

JEL classification: C35, D1, I1, O12.
Keywords: Water, Sanitation, Diarrhea, Propensity score, Matching, India.

*Research Fellow, Harvard Center for Population and Development Studies, 9 Bow st, Cambridge, MA, 02138 (e-mail: kumar@hsph.harvard.edu)

†Research Fellow, Harvard Center for Population and Development Studies, 9 Bow st, Cambridge, MA, 02138 (e-mail: svollmer@hsph.harvard.edu). We thank Amitabh Chandra, David Canning, and Gunther Fink and seminar participants at Institute for Health Metrics and Evaluation at University of Washington for helpful comments and discussion.

1 Introduction

The United Nations Millennium Development Goal 4 (MDG 4) calls for a reduction of "under-5 mortality by two-third" by 2015 (General Assembly of the United Nations, 2000). However, the level of child mortality remains high in many low and middle income countries. A recent UNICEF report (Countdown to 2015) reports that only 19 of the 68 *countodown* countries are on track to achieve MDG 4. Globally, nine million children die before age five every year. About one out of five under-five deaths is due to diarrhea (Bryce et al. 2005).Diarrheal disease, the second leading cause of child mortality after pneumonia, kills approximately two million children. About one out of five under-five deaths is due to diarrhea (Bryce et al. 2005). The number of diarrheal deaths is higher than the number of deaths due to AIDS, malaria and measles combined (WHO, 2009). The global health burden of diarrheal disease is tremendous and falls disproportionately on Africa and South Asia, with more than 80 percent of diarrheal deaths occurring in these two regions. Within South Asia, India has the greatest burden of diarrhea. In India alone which is the focus of this paper, about 0.4 million children die annually due to diarrhea.

Diarrheal diseases are often described as water-related, but more accurately it is an excreta-related disease since the pathogens are derive from faecal matter (UN factsheet, 2008). The principal route of diarrheal disease infection is fecal-oral cycle, and breaking this cycle which depends primarily upon hand-washing and toilet use, saves children's lives. Hygiene and sanitation are considered as the most cost-effective public health interventions to reduce diarrheal morbidity and mortality.

In developed countries, the provision of piped water and improved sanitation facilities brought substantial health improvements (Cutler and Miller, 2005; Watson, 2006). However, similar evidence for developing countries is lacking and a key question in policy circles is whether investing in the environment health sector alone will be sufficient to reduce the diarrheal disease burden. There is a lack of consensus on the effectiveness of different water, sanitation, and hygiene interventions. On one hand, children in developing countries face very high risk of mortality and morbidity from diarrheal diseases. On the other hand, these countries have also made considerable progress in extending sanitation coverage. In this context, it is an important policy question to examine whether access to improved sanitation is effective in reducing diarrheal morbidity, especially in a large developing country like India.[1]

Quantifying the impact of access to improved sanitation on children's morbidity is an important policy purpose for at least two reasons. First, it can serve as a guide for the allocation of scarce resources to the numerous other interventions competing for the same funds. Second, it will also help us understand the relative importance of various factors that permit certain households in a given socioeconomic environment to achieve greater benefits from access to improved sanitation than others.

India is an ideal country to examine the aforementioned question, because it has made substantial progress in water and sanitation coverage in the last decade. The sanitation coverage in India has almost tripled from 22 percent in 2001 to 57 percent in 2008 (Figure 1). We take advantage of a newly available nationally representative individual-level data set, District Level Household Survey 3 (DLHS-3), to rigorously quantify the effect of improved sanitation on diarrhea incidence among young children in India.[2]

A major challenge in programme evaluation is the non-random distribution of the treatment, which leads to selection bias. For example, we do not know why only some households in a village have access to improved sanitation. It might be the case that households with access to improved sanitation are also more forward-looking and possess better health knowledge and behavior. Thus, the unobserved characteristics of households may be responsible for differences between households and not the actual treatment. The

[1] According to the Joint Monitoring Programme for Water Supply and Sanitation by the World Health Organization and UNICEF, "improved" sanitation includes connection to a public sewer, connection to a septic system, pour-flush latrine, access to a pit latrine, ventilated improved pit latrine.

[2] DLHS has three rounds and was conducted in 1998-99, 2002-04, 2007-08. See data section for more details.

best method to solve this problem is to conduct a Randomized Controlled Trial (RCT). However, it is nearly impossible to randomize infrastructure projects, such as sanitation. We thus use a non-experimental method, Propensity Score Matching (PSM), to evaluate the effect of improved sanitation on diarrheal incidence.[3] We exploit the richness of the DLHS data to create a reasonable counterfactual based on the propensity score and address the issue of observed selection bias.[4] We further check the robustness of PSM results by estimating a weighted least square (WLS) with propensity score as the weights (Hirano, Imbens, and Ridder, 2003).

Our main finding is that children in households with access to improved sanitation have a lower diarrheal incidence than children in households without. The incidence of diarrhea for children living in household with *improved sanitation* is 2.2 percentage points lower than for children living in household without access to improved sanitation. Put differently, the odds of contracting diarrhea for children in households without improved sanitation is 24 percent higher than for children in households with improved sanitation. The treatment has heterogeneous effects. Boys experienced larger benefits more than girls from the treatment. There is also a gradient by socio-economic status (SES). The treatment has a larger effect on children in high-SES households than on children in low and middle SES households. We also find that diarrheal disease burden is less in households that treat water before use.

The remaining sections of the paper are organized as follows. Section 2 provides an overview of sanitation programme in India and related literature. Section 3 presents the empirical framework. Section 4 describes the data. Section 5 presents our findings. Finally, section 6 discusses our main results and provides some concluding remarks.

[3] See Ravallion (2001) for methods employed in programme evaluation.
[4] PSM assumes that there is no unobserved selection bias.

2 Context and Previous Literature

2.1 Total Sanitation Campaign

In 1986, the Indian Ministry of Rural Development launched the first nationwide programme of sanitation, the Central Rural Sanitation Programme (CRSP). CRSP was supply-driven, highly subsidized, and gave emphasis on toilet construction. The programme failed to motivate and sustain high levels of sanitation coverage as there was no perceived need for sanitation among communities. It was based on the erroneous assumption that provision of sanitary facilities would lead to increased coverage and usage. Despite an investment of more than Rs. 6 billion and construction of over nine million latrines in rural areas, rural sanitation grew at just one percent annually throughout the 1990s and the census of 2001 found that only 22 percent of rural households had access to a toilet (India country paper, 2008).

Recognizing this limitation, CRSP was restructured and renamed Total Sanitation Campaign (TSC) in 1999. The focus shifted from infrastructure to behavioral change. TSC is a demand-driven low-cost approach, advocating a shift from a high subsidy to a low subsidy regime and greater community involvement. TSC puts more emphasis on Information, Education, and Communication (IEC), capacity building and hygiene education.[5] TSC places great emphasis on awareness creation and demand generation for sanitary facilities in houses, schools, and in the village community environment. It is not restricted to the construction of toilet facilities.

TSC is being implemented in 590 districts of 30 states in India. As of October 2008, 57 million toilets have been constructed. In addition, 0.68 million school toilets, 14,540 sanitary complexes for women, and 222,267 anganwadi (pre-school) toilets have been constructed. Rural sanitation coverage has almost tripled from 22 percent in 2001 to 57

[5] One of the main objectives of TSC is to eliminate open defecation to minimize risk of contamination of drinking water sources and food. The major components of TSC are start-up activities, IEC activities, Rural Sanitary Marts and Production Centers (RSM), Individual Household Latrines, Community Sanitary Complex, School Sanitation, and Anganwadi Sanitation.

percent in 2008 (Figure 1). The Government of India has allocated about $4 billion for TSC and as of 2009, an expenditure of $1.6 billion has been incurred. That is an average spending of $25 per household. This study attempts to quantify the impact of access to sanitation on children's morbidity.

[Insert Figure 1 here]

2.2 Related Literature

Jalan and Ravallion (2003) is one of the first papers that evaluates the impact of environmental factors, access to piped water in particular, on diarrheal morbidity in rural India. Using a household survey conducted by the National Council of Applied Economic Research (NCAER), New Delhi, India in 1993-94, they measured the child-health effects of access to piped water in rural India. By implementing propensity score matching at the household level, they find a lower incidence and duration of diarrhea for children living in households with access to piped water.[6] Interestingly, the health benefits of piped water bypass poor households and households in which mothers are poorly educated.

Khanna (2008) extends Jalan and Ravallion's work by including access to sanitation as the explanatory variable, in addition to access to piped water. Unlike Jalan and Ravallion, Khanna makes a distinction between type of sanitation and water infrastructure. She uses data from India's second National Family Health Survey (NFHS), conducted in 1998-99. Employing propensity score matching, she finds an increase in diarrheal incidence in households with piped water, which is contrary to findings in Jalan and Ravallion (2003). While estimating the joint impact of access to water and sanitation, she finds a decrease in diarrhea incidence for children living in households with access to well water, handpump water, well water and sanitation, and handpump and sanitation.

Our study builds on these two studies. Given that national sanitation programmes and investments in sanitation are increasing, it is important to quantify the effect of improved

[6] However, matching at village-level does not indicate lower diarrhea incidence in households with piped water.

sanitation on child morbidity. We are not aware of any studies that rigorously estimates the effect of improved sanitation on diarrheal morbidity for the period after the launch of the Total Sanitation Campaign (TSC) in India. Both, Khanna (2008) and Jalan and Ravallion (2003) use data from pre-TSC period and do not cover the effects of recent sanitation improvements due to TSC. Until 2009, the Government of India has spent 700 million dollars on TSC and has provided access to improved sanitation to five million households. Our study is the first study to estimate the impact of improved sanitation on diarrhea with a very recent data set that adequately covers the TSC period. Another advantage of our study is that we uses a large, nationally representative, and recent data set[7], that enables use to estimate the impact of TSC more precisely.

In addition to the aforementioned studies, there are a few more studies on the impact of access to water and sanitation on diarrheal morbidity. However, most of these studies are based on small and unrepresentative samples. For example, Dasgupta (2004) collected own data from 600 households in 14 localities[8] in New Delhi, India. She finds that children who live in households with access to piped water are less vulnerable to diarrheal attack but surprisingly, she finds no effect of sanitation and the education of the household head on diarrhea incidence. Duraiswamy (2001) uses the large-scale nation-wide NCAER-HDI 1994 survey to examine the correlates of children's vulnerability to diseases[9] and finds no significant effect of availability of toilet, hand-washing behavior, and sources of drinking water on children's morbidity. The availability of a separate kitchen turned out to be a significant correlate of morbidity among children. Borooah (2004) finds that the safety of a village's water supply reduces the incidence of diarrhea by 5%. A lack of toilet facilities in the house increases the probability of diarrhea. He also finds a correlation between diarrhea incidence and hand washing habits of the mothers before feeding their

[7]The DLHS-3 was conducted in 2007-2008 and covers all the districts of India with a sample of about 720,000 households.

[8]These 14 localites were selected based on occurrence of 5 or more cases of cholera in the locality in the past three years before the survey i.e. 1996-1998. The occurrence of more than 5 cases of cholera in a locality is taken as a standard benchmark for determining the vulnerability of an area to waterborne diseases by epidemiologists at the municipal health department. The survey was conducted during the summer months of 1999.

[9]Diarrhea was the third most common cause of morbidity among children after cold and cough.

children.[10] It should be noted that Jalan and Ravallion (2003), Duraiswamy (2001), and Borooah (2004) all use the same data set (the NCAER-HDI survey collected in 1993-94), but implement different methods. The adoption of improved sanitation and water facility is not exogenous, thus the results in Duraiswamy (2001) and Borooah (2004) cannot be interpreted as causal. Jalan and Ravallion (2003) on the other hand correct for selection bias by employing propensity score matching and can thus be interpreted as causal.[11]

Bose (2009) analyzes the DHS data for Nepal with propensity score matching and finds a 5 percent reduction in diarrhea incidence for children living in households with improved sanitation. The effect is larger, about 11 percent, for children below 24 months of age. Besides the impact of water and sanitation on children's morbidity, some studies have also looked at other health outcomes, such as infant and child mortality, height-for-age, weight-for age or height-for-weight (Fink et al., 2010). Fink et. al. merged 171 DHS surveys in 70 low and middle income countries over the period 1986-2007 to estimate the effect of water and sanitation on child development. Using a logit model they find that access to improved water and sanitation infrastructure at the household level results in reduction in infant mortality, diarrhea incidence, and stunting among children in low and middle income countries.

There is a growing impact evaluation literature that examines the effects of improved water, sanitation, and hygiene (WSH) in other developing countries. In recent years, a number of surveys have been published examining the impact of WSH interventions on diarrhea morbidity, using systematic literature reviews, meta-analysis, and/or meta-evaluation (Esrey et al. 1991, Curtis and Cairncross 2003, Fewtrell and Colford 2004, Fewtrell et al. 2005, Arnold and Colford 2007, Snilstveit and Waddington 2009).[12] These studies provide overwhelming evidence on the positive impact of hand washing, sanitation, and household and point-of-use water treatment on better health outcomes.

[10] In developing countries, feeding is mostly done by hand and not using cutlery, and India is no exception.
[11] Under the assumption that there is no unobserved selection bias.
[12] See Fewtrell et al. 2009 for detailed discussion.

3 Empirical Framework

The objective of this paper is to estimate the causal impact of "improved sanitation infrastructure" on child morbidity, indicated by diarrheal rates. Estimating the impact of sanitation access is a major methodological challenge because we cannot observe outcomes for the same individuals in both states: treatment and counterfactual state (Heckman and Robb 1985). For example, in this study, we can observe households with either access to improved sanitation or without, but we can not observe outcomes for the same households in both states. The most convincing approach to solve this problem of missing data is to conduct a randomized experiment where the counterfactual is created from a random subset of the eligible population. However, randomizing infrastructure such as, roads, ports, electricity, water and sanitation is not feasible for many reasons.

Therefore, in the absence of experimental data, we rely on observational data and implement non-experimental method, propensity score matching (PSM), to estimate the causal impact of improved sanitation on child morbidity. The estimation of the treatment effect in observational studies may be biased due to confounding factors, because subjects are assigned to the treatment and control groups non-randomly. Propensity score matching is an alternative to "correct" the bias by creating treated and control groups that are not confounded by differences in observed covariate distributions (Rosenbaum and Rubin, 1983). In recent years, matching methods have become increasingly popular and widely used in the evaluation of economic policy interventions (Becker and Ichino, 2002).

The basic idea in PSM is to generate groups of treated and control that have similar characteristics so that comparisons can be made within these matched groups. In the event of a large number of observed characteristics, direct matching becomes infeasible and propensity score (a single-index variable) matching is used. The propensity score $p(X)$ is the estimated probability of receiving treatment given a set of background covariates. The difference in the average outcome of treated and control groups can be attributed to the program under the assumption that selection into program participation is based on observable factors alone.[13]

3.1 Average Treatment Effect on the Treated (ATT)

Let Y_{1i} and Y_{0i} be the outcome variables for treated and control households, respectively, and $D \in \{0,1\}$ the indicator of treatment status. The propensity score *p(X)* is defined by Rosenbaum and Rubin (1983) as the conditional probability of receiving treatment given observed characteristics:

$$p(X) \equiv Pr(D=1 \mid X) = E(D \mid X) \tag{1}$$

where X is the multidimensional vector of observed characteristics.

Given the propensity score *p(X)*, the Average effect of Treatment on the Treated (ATT) can be estimated as follows:

$$\begin{aligned}\widehat{ATT} &\equiv E\{Y_{1i} - Y_{0i} \mid D_i = 1\} \\ &= E[E\{Y_{1i} - Y_{0i} \mid D_i = 1, p(X_i)\}] \\ &= E[E\{Y_{1i} \mid D_i = 1, p(X_i)\} - E\{Y_{0i} \mid D_i = 0, p(X_i)\} \mid D_i = 1]\end{aligned} \tag{2}$$

Equation (2) gives the average program impact under the conditional independence assumption (CIA)[14] and overlap assumption.[15]

[13] See Dehejia and Wahba (1999, 2002), Heckman et al. (1997, 1998a), Smith and Todd (2001, 2005a) for an evaluation of matching estimators.

[14] Conditional independence means that conditional on X, the outcomes are independent of treatment, and can be written as Y_1, $Y_0 \perp D \mid X$.

[15] Overlap means that for each X there are both treated and control units, i.e. $0 < \Pr[D{=}1 \mid X] < 1$.

3.2 Nearest Neighbor Matching Method

In this paper, we employ nearest-neighbor (NN) matching with replacement, which is the most widely used matching algorithm.[16] We matched the treatment household with the nearest-five-neighbors. Formally, the NN matching estimator with replacement within caliper is,

$$\widehat{ATT} = \frac{1}{N1} \sum_{i=I} \{Y_i - Y_j\} \tag{3}$$

For a pre-specified caliper $\delta > 0$, j is chosen such that,

$$\delta >\mid p(X_i) - p(X_j) \mid = min_{k \in I} \{\mid p(X_i) - p(X_j) \mid\}$$

If none of the non-treated units is within δ from the treated unit i, then i is left unmatched. We use the nearest five neighbors, which takes the average outcome measures of the closest five matched control units as the counterfactual for each participant.

3.3 Propensity-based Weighted Regression

Another method widely used in program evaluation literature is estimation of a multivariate regression, using propensity score as sampling weights. Several studies suggest that weighting the data by propensity score balances the distribution of covariates and results in fully efficient estimates (Rosenbaum, 1987; Hirano and Imbens, 2001; Hirano, Imbens, and Ridder, 2003; McCaffrey et al., 2004). This approach uses propensity score ($\hat{\lambda}$) to weight treatment and control groups in order to make the covariate distribution similar across both groups. The weight is defined as the inverse of the propensity score $1\backslash\ \hat{\lambda}$ for treated households and the inverse of one minus the propensity score $1\backslash$ 1-$\hat{\lambda}$ for untreated households.[17] For comparison and robustness, we implement this approach by

[16] With nearest-neighbor matching, the individual from comparison group is chosen as a matching partner for a treated individual that is closest in terms of propensity score.

[17] A variation of the formula with the square root is also used. We prefer the square-rooted version because it scales down the variation in weight.

estimating the following multivariate regression with propensity score as weights:

$$Y_{ijs} = \beta_0 + \beta_1 PROGRAM_{js} + \delta X_{js} + \gamma_s + \epsilon_{ijs} \tag{4}$$

where $PROGRAM_{js}$ is the access to electricity and the equation is estimated using the weight $\hat{\lambda}$. [18]

4 Data

We use data from third wave of the District Level Household Survey (DLHS 3), which is a health survey covering family planning, maternal and child health, reproductive health of ever-married women and adolescent girls, and use of maternal and child health-care services at the district level for all states in India. DLHS-3 was carried out during December 2007-December 2008 in all districts of India, interviewing about 720,320 households from 611 districts in 34 states. However, after dropping observations with missing information, our final analytical sample consists of 206,935 observations.

In addition to socio-demographic information, the survey also collected information on diarrhea prevalence in the past two weeks before the survey for the children born after January 2004. We use this information to construct our outcome variable, that is a dummy for whether a child suffered from diarrhea in the past two weeks. We matched households with access to "improved sanitation" to household without "improved sanitation" using the propensity score generated from the following variables: whether the household has access to piped water, mother's age, father's age, mother's years of schooling, father's years of schooling, whether house structure is pucca/kutcha, number of young children in the household (less than five years old), fraction of boys among young children, average age of young children, whether the panchayat head lives in the respondent's village, total number of males in the household, total number of females in the household, amount of irrigated

[18] Propensity score ($\hat{\lambda}$) are estimated from a logistic regression. We also estimate an ordinary least square (OLS) model for income and linear probability model (LPM) for educational outcomes without reweighting the data.

land (in acres), whether village has health and sanitation committee, distance to district headquarter (in km), household religion, household caste, state of residence, whether household is below the poverty line, household electrification status, and availability of health facility (anganwadi) in the village.[19]

The treatment variable is access to improved sanitation.[20] In our sample, about 26 percent of all households have access to improved sanitation. The mean incidence of diarrhea is 12 percent. The mean diarrheal incidence in the treated households is 10 percent, whereas it is 13 percent in the control households. Younger children are more susceptible to diarrheal risk. The average diarrhea incidence among children under two years of age is 15 percent, for children over two years of age the average incidence is 9 percent.

Table 1 reports the descriptive statistics of variables that were included in the propensity score estimation. It can be noted that treated households are different from control households on many dimensions. For example, treated households have more access to piped water, have higher parents education, have higher landholding size, more likely to to electrified, less likely to be hindu and poor, and more likely to have pucca house structure. This means that treated households enjoy relatively higher economic and social status, and are very different from control households, and thus motivates the propensity score matching as a framework to estimate the treatment effect.

[Insert Table 1 here]

[19] In the absence of baseline date, we resort to ex-post matching by using variables that are presumable time-invariant and might not have changed due to treatment.

[20] The World Health Organization (WHO) defines improved sources of sanitation to be flush toilets connected either to a sewage system or a septic tank, ventilated pits, or composite toilets.

5 Results

5.1 Propensity score estimation

The propensity score is the probability of receiving treatment (here, access to improved sanitation) conditional on the observed characteristics of the child. We estimate the propensity score with a logit regression, which has a treatment dummy as the dependent variable, and a number of covariates as independent variables (see data section for a full list). Table 2 reports the first stage. Almost all variables that entered into the regression significantly predict the treatment.

[Insert Table 2 here]

Following Sianesi (2004), we compared the Pseudo R-squared pre and post matching to examine the quality of matching. Table 3 shows that the post matching Pseudo R-squared is much lower (0.01) than pre matching Pseudo R-squared (0.26). This strong reduction in the Pseudo R-squared indicates a high quality of the match. Likelihood ratio test of the joint significance is insignificant suggesting that matching quality is good. It is evident from Table 3 that matching has achieved significant reduction in observed selection bias. Finally, visual analysis of the density distribution of the propensity score in figure 2 indicates sufficient overlap between treated and control households.

[Insert Table 3 here]

[Insert Figure 2 here]

5.2 Main Results

Table 4 reports the average treatment effects on the treated (ATT) for diarrhea outcome. The ATT is the difference in mean prevalence of diarrhea for children in households with improved sanitation and children in households without access of improved sanitation. Access to sanitation leads to statistically significant reducation in diarrhea for children under age five. The mean incidence of diarrhea in households with access to improved sanitation is 2.2 percentage points lower than in households in the control group. To put this in context, the odds of a child living in household without access to improved sanitation of having diarrhea is 24 percent larger than for a child in the treated group.[21] Additionally, on a comparison group average of 13% of children with diarrhea in the past week, a 2.2 percentage points decline means a drop of 17%.

[Insert Table 4 here]

5.3 Heterogeneous Treatment Effects

So far, we were focussed on the average impact of improved sanitation on diarrhea incidence. However, it is quite likely that the impact of the treatment varies by socioeconomic status (SES), age and gender of the child, and household behavior about water treatment. Results are shown in Table 5.

We find no significant treatment effect for children who live in households that belong to either low or middle SES. The effects are driven mainly by reduced diarrhea among children belonging to high SES. For children in high SES, the estimated reduction in diarrhea is 2.5 percentage points, and this effect is significant at 99% confidence level (Panel A, Table 5). While the treatment effect is negative for middle SES children, it is positive for Low SES children, though neither of the estimated effects are statistically significant at any level of significance. We also estimated if more young children benefitted more from the treatment, possibly due to within household externality. Younger children may also benefit disproportionately because they are more susceptible to diarrheal attack due to their weaker and less developed immune system. We do not find any age gradient in the treatment effect. The effects are quite similar across young (less than two years old) and slightly older children.[22]

We also stratified the sample by gender of children and household behavior about water

[21] The odds for the treated group is 0.101/0.899=0.112, and the odds for control group is 0.123/0.877=0.140, which results in an odds ratio of 1.25 (0.140/0.112).

[22] Results not shown here, but available upon request.

treatment (Table 5, Panel B & C). About 27% households reported treating their water before drinking and boiling was the most common method of treating. The mean incidence of diarrhea for boys in households with access to improved sanitation is 2.0 percentage points lower than the boys in households in the control group, and this effect is significant at 95% confidence level. For girls, the treatment effect is 0.7 percentage points, however, this effect is statistically insignificant. We also find that greater reduction in diarrheal incidence in households that treat water before use. The treatment effect is 3.3 percentage points and is statistically significant at 99% confidence level.

[Insert Table 5 here]

5.4 Robustness Checks

For comparison purposes and as a robustness check, we also estimate a linear probability model(LPM, Table 6, Col 1) and weighted least square regression (WLS, Table 6, Col 2).[23] The treatment effects are consistent in sign, and they are significant as well. The point estimate in LPM and WLS is 0.8 and 1.0 percentage points respectively. It should be observed that point estimates in either LPM or WLS are slightly lower than PSM estimates.

However, as we have discussed before, the estimates from either LPM or WLS linear probability cannot be interpreted as a causal estimate because sources of variation in the treatment are not plausibly exogenous, and that is why the preferred specification in our paper is PSM.

[Insert Table 6 here]

Finally, even though we are confident in the quality of match, our results should be interpreted with a caveat that propensity score matching only provides causal interpretation if the selection into treatment is observed and adequately controlled in the model. The PSM assumptions are violated if the selection into treatment is based on some unobserved

[23] The weight is defined as the inverse of the propensity score $1\backslash\ \lambda$ for treated households and the inverse of one minus the propensity score $1\backslash$ 1-λ for untreated households (see Hirano, Imbens, and Ridder (2003) for more detail).

covariates. We could not combine matching with differences-in-differences method, to remove the bias due to selection on time invariant unobserved variables, due to unavailability of suitable baseline data. We therefore used ex-post matching. Second, our estimates may underestimate the true health effect of sanitation due to spill-over effects/externalities. Diarrhea is an infectious disease, therefore it is quite likely that children in counterfactual households might have benefitted indirectly from the treatment of other households in their neighborhood.

6 Conclusions and Policy Implications

This study employs propensity score matching to quantify the health gains to children from access to improved sanitation in rural India. We find reduced diarrhea incidence for children in household with access to improved sanitation. Access to improved sanitation roughly averts 0.8 diarrhea episodes per household year.[24] 0.8 case per household-year might not seem like big improvement. But given that diarrhea is the second leading cause of death and that there are on average 3.1 diarrhea cases per child-year (or 3.9 diarrhea cases per household-year), this is a non-negligible improvement from a public health perspective.

The sole health outcome studied in this paper is diarrheal morbidity. Clearly, water, sanitation and hygiene interventions are likely to have an impact on other illnesses, such as schistosomiasis, ascariasis and respiratory outcomes as well.

There is considerable heterogeneity in the impact of improved sanitation. Boys benefit more than girls from the treatment. Also, the treatment effects bypassed the poorer children belonging to low and middle SES and high SES children experienced larger benefit from access to sanitation. It is particularly troubling that for girls and for children from low SES households there is no significant effect of improved sanitation on diarrhea.

[24] (0.022 diarrhea case per child in two weeks)*(1.40 children under 5 per household)*(26 two week episodes per year)=(0.4225 diarrhea cases per household year)

This indicate that the average impact may not be best measure to assess the gains from improved sanitation. For lower SES households it might be the case that the potential benefits from improved sanitation are not realized, because of lower parent's education and health awareness. The differences between boys and girls are more puzzling, because in theory all children in the household should have equal access to improved sanitation (if treated). Thus, the only explanation for these large gender differences, that we can think of, are neglect and gender discrimination.

Despite making a huge progress in sanitation coverage in rural parts, India still has the largest burden of child mortality and morbidity related to diarrhea in South Asia. It is plausible to argue that benefits from the access to infrastructure such as piped water, improved sanitation might not have been realized fully in the absence of complementary inputs such as education or income of parents. Also behavior change through community participation, education, awareness, and health promotion activities may go a long way in reducing diarrheal burden. Public policies on sanitation should be complemented with other policies that alleviate poverty, improve parent's education and promote gender equity.

References

Arnold BF, Colford JM. 2007 . Treating water with chlorine at point-of-use to improve water quality and reduce child diarrhea in developing countries: a systematic review and meta-analysis. *American journal of tropical medicine and hygiene* **76**(2): 354-364.

Becker SO, Ichino A. 2002 . Estimation of Average Treatment Effects based on Propensity Scores. *Biometrika* **70**(1): 41-55.

Borooah VK. 2004 . On the Incidence of Diarrhea among Young Indian Children. *Economics and Human Biology* **2**: 119-138.

Bose R. 2009 . The impact of water supply and sanitation interventions on child health: evidence from DHS surveys Mimeo. International Initiative for Impact Evaluation: New Delhi.

Bryce J. et al. 2005 . WHO estimates of the causes of death in children. *Lancet* **365**: 1147-1152.

Curtis V, Cairncross S. 2003 . Effect of washing hands with soap on diarrhoea risk in the community: a systematic review. *Lancet infectious diseases* **3**(5): 275-281.

Cutler D, Miller G. 2005 . The role of public health improvements in health advances: the 20th century United States. *Demography* **42**(1): 1-22

Dehejia RH, Wahba S. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics* **84** (n1):151-61.

Dasgupta P. 2004 . Valuing Health Damages from Water Pollution in Urban Delhi, India. *Environment and Development Economics* **9**: 83-106.

Doraiswamy P. 2001 . Health Status and Curative Health Care in Rural India. National Council of Applied Economic Research (NCAER) Working paper no. 78.

Esrey SA *et al.* 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 L, Colford J. 2004 . Interventions and diarrhea: a systematic review and meta-analysis. *World Bank HNP Discussion Paper No. 34960.* World Bank: Washington, DC.

Fewtrell L *et al.* 2005 . Water, sanitation, and hygiene interventions to combat childhood diarrhea in developing countries. Synthetic Review 001. International Initiative for Impact Evaluation.

Fewtrell L *et al.* 2009 . Water, sanitation, and hygiene interventions to reduce diarrhea in less developed countries: a systematic review and meta-analysis. *Lancet infectious diseases* **5**: 42-52.

Fink G, Gunther I. 2010 . Water, Sanitation, and Children's Health: Evidence from 172 DHS Surveys. *World Bank Policy Research Working Paper No 5275.* World Bank: Washington DC.

Heckman JJ, Robb R. 1985 . Alternative Methods for Evaluating the Impact of Interventions: An Overview. *Journal of Econometrics* **30**(1-2): 239-67.

Hirano K, Imbens G. 2001 Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Hear Catherization. *Health Services and Outcomes Research Methodology* 2(3-4): 259-278.

Hirano K, Imbens G, Ridder G. 2003 . Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score. *Econometrica* **71**: 1161-1189.

India Country Paper. 2008 . Sustaining the Sanitation Revolution. Department of Drinking Water, Ministry of Rural Development, Government of India, New Delhi.

Jalan J, Ravallion M. 2003b . Does Piped Water Reduce Diarrhea for Children in Rural India? *Journal of Econometrics* **112**: 153-173

Khanna G. 2008 . The Impact of Child Health from Access to Water and Sanitation and other Socioeconomic Factors. *HEI Working Paper No: 02/2008* Graduate Institute of International studies, Geneva.

Kremer M, Zwane A. 2007 . Cost-Effective Prevention of Diarrheal Diseases: A Critical Review. Working Paper, Center for Global Development

McCaffrey DF *et al.* 2004 . Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods* **9**(4).

Ravallion M. 2001 . The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. *World Bank Economic Review* **15**(1): 115-140.

Ravallion M. 2005 . Evaluating anti-poverty programs. The World Bank, Policy Research Working Paper Series: 3625.

Rosenbaum PR, Rubin DB. 1983 . The Central Role of the Propensity Score in Observational Studies for Causal Effects. *The Stata Journal* **4**: 358-377.

Rosenbaum PR. 1987 . Sensitivity analysis for certain permutation inferences in matched observational studies *Biometrika* **74**(1): 13-26.

Sianesi B. 2004 . An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *Review of Economics and Statistics.* **86** (n1): 133-55.

Snilstveit B, Waddington H. 2010 . Effectiveness and sustainability of water, sanitation, and hygiene interventions in combating diarrhea. *Journal of Development Effectiveness* **1**(3): 295-335.

Watson T. 2006 . Public health investments and the infant mortality gap: evidence from federal sanitation interventions and hospitals on U.S. Indian reservations. *Journal of*

Public Economics. **90**(8-9): 1537-60

UNICEF/WHO. 2009 . Diarrhoea: Why children are still dying and what can be done.

Table 1: Descriptive statistics of household characteristics

Household characteristics	With access to improved sanitation		Without access to improved sanitation	
	Mean	S.D.	Mean	S.D.
	(1)	(2)	(3)	(4)
Piped water (1=yes)	0.33	0.47	0.16	0.37
Mother's age	26.39	5.02	26.30	5.46
Father's age	31.74	6.21	31.26	6.29
Mother's education	6.78	4.64	2.74	3.85
Father's education	8.34	4.56	5.40	4.62
House structure (1=pucca)	0.35	0.49	0.12	0.32
No of young children	1.35	0.48	1.42	0.50
Fraction of young boys	0.53	0.45	0.52	0.44
Average age of young children (months)	24.74	12.21	23.59	11.16
Panchayat head lives in respondent's village	0.68	0.47	0.56	0.50
No of males in the household	3.48	1.94	3.44	1.89
No of females in the household	3.64	1.97	3.65	1.94
Amount of irrigated land	1.53	5.86	0.87	2.80
Health and sanitation committee in the village	0.34	0.47	0.22	0.41
Distance to district	43.11	37.35	46.07	33.34
Hindu	0.59	0.49	0.83	0.38
Muslim	0.18	0.38	0.11	0.32
Schedule caste (SC)	0.13	0.34	0.22	0.41
Schedule tribe (ST)	0.21	0.41	0.19	0.39
Other backward caste (OBC)	0.31	0.47	0.43	0.50
Below poverty line status	0.25	0.43	0.37	0.48
Electrified	0.77	0.42	0.45	0.50
Anganwadi in the village	0.95	0.22	0.91	0.29
N	54629		152306	

Notes: Parents education is years of schooling.

Table 2: Determinant of access to improved sanitation estimated for propensity score

Household characteristics	
Access to piped water	0.429***
	(0.019)
Mother's age	0.067***
	(0.012)
Father's age	0.025***
	(0.009)
Mother's education (Years of schooling)	0.113***
	(0.002)
Father's education (Years of schooling)	0.062***
	(0.002)
House structure (pucca=1)	0.954***
	(0.017)
No of young children	-0.115***
	(0.014)
Fraction of boys among young	-0.003
	(0.017)
Average age of young children	0.002***
	(0.001)
Panchayat head lives in the respondent's village	0.242***
	(0.016)
Total no of male in the household	0.023***
	(0.004)
Total no of female in the household	0.027***
	(0.004)
Amount of irrigated land (in acres)	0.031***
	(0.002)
Health and sanitation committee in the village	0.251***
	(0.017)
Distance to district headquarter	-0.002***
	(0.000)
Hindu	-0.682***
	(0.034)
Muslim	0.278***
	(0.040)
Schedule caste	-0.353***
	(0.022)

continued from Table 1

Schedule tribe	-0.692***
	(0.027)
Other backward caste	-0.257***
	(0.018)
Health personnel in the village	0.141***
	(0.027)
Below poverty line	-0.165***
	(0.016)
Whether household is electrified	0.868***
	(0.018)
Wald	48362.42
P-value	0.00
McFadden's Pseudo-R	0.36
N	206935

Notes: Standard errors are clustered by state and are robust to heteroskedasticity. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 3: Absolute Bias, pseudo-R^2 and LR χ^2

	Pseudo-R^2	LR χ^2	p $> \chi^2$
Unmatched	0.263	62722.89	0.000
Matched	0.010	1509.12	0.000

Table 4: Average Treatment Effect of Improved Sanitation: PSM Estimates

Variables	Treatment	Control	Difference	S.E.	T-stat
	(1)	(2)	(3)	(4)	(5)
Unmatched	0.101	0.125	-0.024***	0.002	-14.91
Matched	0.101	0.123	-0.022***	0.007	-3.22

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Heterogeneous Effects of Improved Sanitation

Panel A: Stratified by wealth index quintiles

Variables	Treatment	Control	Difference	S.E.	T-stat
	(1)	(2)	(3)	(4)	(5)
Low SES	0.091	0.090	0.001	0.009	0.12
Middle SES	0.095	0.104	-0.008	0.006	-1.33
High SES	0.106	0.131	-0.025***	0.008	-2.99

Panel B: Stratified by gender of children

Variables	Treatment	Control	Difference	S.E.	T-stat
	(1)	(2)	(3)	(4)	(5)
Boy	0.104	0.125	-0.020**	0.009	-2.35
Girl	0.099	0.105	-0.007	0.008	-0.84

Panel C: Whether households treat water

Variables	Treatment	Control	Difference	S.E.	T-stat
	(1)	(2)	(3)	(4)	(5)
Treat water (yes)	0.095	0.129	-0.033***	0.011	-3.11
Treat water (no)	0.107	0.107	0.0005	0.005	0.10

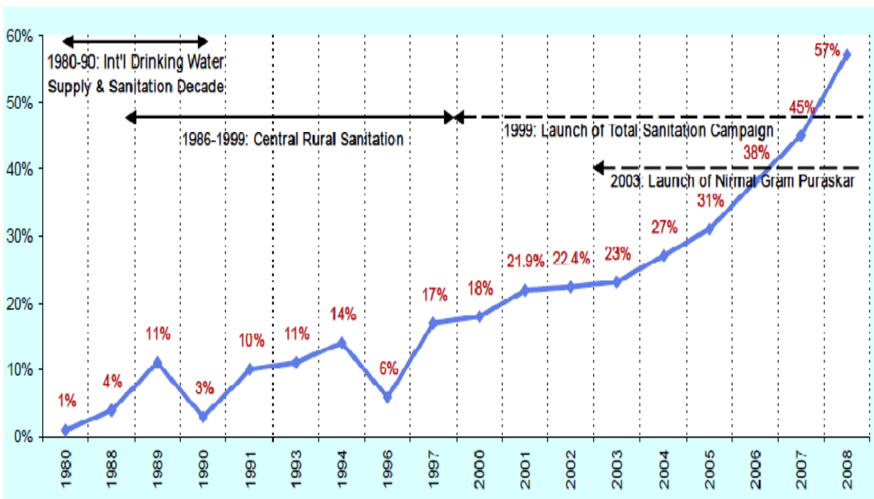
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6: Average Treatment Effect of Improved Sanitation: Robustness check

	Linear Probability Model	Weighted Regression
Independent Variables	(1)	(2)
Improved sanitation	-0.008**	-0.010***
	(0.004)	(0.003)
N	206,935	206,935
R Square	0.03	0.03

Notes: Standard errors are clustered by state and are robust to heteroskedasticity. Household characteristics include mother's age, mother's education, father's age, father's education, poverty status, whether village has health worker, caste, electrification status, house type (pucca), and religion. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure 1: Progress in Coverage of Rural Sanitation in India



Source: Govt. of India, Dept. of Drinking Water Supply http://ddws.nic.in Accessed 16 Oct-08

Figure 2: Distribution of Propensity Score