Graduate Institute of
International Studies | Geneva
Economics

HEI Working Paper No: 02/2008

The Impact on Child Health from Access to Water and Sanitation and other Socioeconomic Factors

Gauri Khanna
Graduate Institute of International Studies

Abstract
In this paper we examine the impacts on child health, using diarrhoea as the health outcome, (amongst children living in households) with access to different types of water and sanitation facilities, and from socio-economic and child specific factors. Using cross-sectional health survey data for India, we employ the propensity score method to match children belonging to different treatment groups, defined by water types and sanitation facilities, with children in a control group. We also employ non-matching techniques to compare our results and to check for their robustness. Our results indicate that disease-specific awareness has strong marginal effects on reducing the predicted probabilities of diarrhoeal outcomes in young children, which are consistent across the models utilised. We also find disease-specific awareness to have the largest impact on reducing the burden of disease from diarrhoea across a select group of predictors.

The Impact on Child Health from Access to Water and Sanitation and other Socioeconomic Factors*

Gauri Khanna[§]
Graduate Institute of International Studies

Abstract

In this paper we examine the impacts on child health, using diarrhoea as the health outcome, (amongst children living in households) with access to different types of water and sanitation facilities, and from socio-economic and child specific factors. Using cross-sectional health survey data for India, we employ the propensity score method to match children belonging to different treatment groups, defined by water types and sanitation facilities, with children in a control group. We also employ non-matching techniques to compare our results and to check for their robustness. Our results indicate that disease-specific awareness has strong marginal effects on reducing the predicted probabilities of diarrhoeal outcomes in young children, which are consistent across the models utilised. We also find disease-specific awareness to have the largest impact on reducing the burden of disease from diarrhoea across a select group of predictors.

Keywords: Diarrhoea, Water, Sanitation, Propensity Score, Matching Techniques

JEL Codes: I1, D1, C35

---

* I would like to thank Prof. John Cuddy, Prof. Neha Khanna, Dr. Krishna Rao and Mr. Will Reidhead for their advice and comments

§ Graduate Institute of International Studies, 11A Avenue de la Paix, 1202 Genève.
Correspondence: khanna1@hei.unige.ch; gauri.khanna@gmail.com

1. Introduction

India's latest National Water Policy 2002 gives clean drinking water prime importance over competing objectives such as irrigation and generation of hydropower (Government of India, 2002). Meeting the twin challenges of water and sanitation, especially for rural households in India, has been the focus of many of the government's efforts. For example, the central government launched the Accelerated Rural Water Supply Programme (ARWSP) in 1972-73 (Government of India, 2003) and the Central Rural Sanitation Programme (CRSP) in 1986 (which was subsequently restructured in 1999), to assist States and Union Territories in their work of making drinking water and sanitation facilities available in villages (Government of India, 2002b; Government of India 2003).[1] More recently, the government launched the Total Sanitation Campaign (TSC) in 1999 and the *Swajaldhara* Scheme in 2002, which adopt a "participatory and demand driven approach" by integrating village communities (Government of India, 2003b). Furthermore, the TSC scheme includes the dissemination of information and education on sanitation and domestic hygiene as part of its Information, Education and Communication efforts to bring about behavioural change amongst villagers.

While considerable progress has been made in the provision of drinking water to rural households, with full coverage of 94.37% of rural habitations, little progress has been made in the sanitation sector, where a mere 22% of the total rural population has access to basic sanitation facilities (Government of India, 2004). In urban India, between 75 and 81% of the population has access to sanitation. However, these figures do not reflect the wide regional disparities present. Furthermore, between 26% and 31% of the rural population and between 7% and 9% of the urban population continues to take water from unprotected sources (Government of India, 2002b). Unless access to water, both in terms of quantity *and* quality, and sanitation facilities are available universally, reducing the health burden from water-borne diseases and unhygienic practices will continue to remain a challenge in India.

Globally, the importance of clean drinking water and improved sanitation to reducing mortality and morbidity is widely documented. According to the World Health Organisation (2002), 3.1% of deaths (1.7 million people annually) are attributable to unsafe water, sanitation and hygiene. It is estimated that environmental risks, such as those posed by unsafe

[1] Water and sanitation is a State subject in India and comes under the purview of the State administrative authorities.

water and poor sanitation, account for a burden of disease of 3.7% (54.2 million) of disability adjusted life years (DALYs), where a DALY equals loss of one healthy year. This figure is even higher (5.5%) for developing countries, such as India, with high child mortality. Poor quality drinking water and sanitation pose a risk to health and are channels for transmission of infectious diseases like diarrhoea. The WHO estimates that diarrhoea is the "single largest contribution to the burden of disease" linked with water, sanitation and hygiene, with children in developing countries comprising the majority of sufferers. In India, an estimated 0.4 to 0.5 million children under five die due to diarrhoea each year (Government of India, 2002b). Likewise, it is estimated that the improvement of water and sanitation services would result in a 17% reduction in the number of diarrhoea cases (an annual reduction of 1.8 billion cases globally) (WHO, 2002).

The paper is organised as follows. Section I describes the research objectives and introduces the health outcome under study. The specific questions raised in this paper are described in section 2. A literature review is presented in section 3 followed by a description of the methodology and a theoretical framework in section 4. In section 5 we present the estimation techniques adopted with a brief description of the data used in section 6. we present my results in section 7 followed by conclusions and policy implications of the study in section 8.

1.1 Research Objectives

The objective of this research paper is to examine the impact of environmental variables - water and sanitation - and socioeconomic factors on child health at the household level. The research has been motivated by several questions relating to the efficacy of public investments in water availability, sanitation facilities, education campaigns and related factors, some of which fall in the domain of private behavioural inputs which are complementary to public action. In addition, other factors as diverse as occupation, independence of the mother in deciding healthcare, and general awareness are equally important in determining health outcomes. The influence of socioeconomic variables and their interactive effects with household environmental variables are examined to determine their impact on health outcomes in children.

We draw inspiration from the Government of India’s Tenth Five Year Plan (2002-2007) which emphasises Rural Water Supply and Sanitation. The Plan also reflects the importance of reducing water-borne diseases by the provision of safe drinking water and sanitation and states that “70%-80% of illnesses are related to water contamination and poor sanitation.” Reduction in the incidence of diarrhoea and jaundice are seen as the main conduits for meeting the national objective of reduced morbidity (Government of India, 2002c). Much of this concern has been echoed in the guidelines of the *Swajaldhara* Scheme and the ARWSP. The analysis in this study thus hopes to render useful policy prescriptions to reduce morbidity in children from diarrhoea by examining factors that influence their health outcomes.

1.2 Health Indicator

In this paper, we use diarrhoea as the health indicator. Diarrhoea, a symptom of gastrointestinal infection, is defined as the passage of loose watery stools more frequently than for a normal healthy person. A severe case of diarrhoea or acute diarrhoea is characterised by the passage of three or more watery or loose stools in a span of 24 hours (WHO, 2004; Manatsathit, S. et al 2002). Young children and infants are more vulnerable to diarrhoeal disease than adults due to their lower body mass ratio and high risks of mortality from dehydration. The intake of water contaminated by human or animal excreta, especially faeces, poses a high risk to health and is one significant channel of transmission of diarrhoea pathogens. Other possible channels include contaminated food and direct faecal oral contamination.

There are several factors that determine the transmission of diarrhoea-causing pathogens: 1) hygiene standards and sanitation practices at the household level; 2) availability of drinking water facilities, whether contaminated or not, and availability of a continuous or interrupted supply of water; and 3) sanitation services and sewage facilities. Thus, interventions can be made at several points and can be combined for greater effect. Esrey et al (1985) and Esrey et al (1991) reviewed a total of 84 studies on diarrhoeal morbidity and improvements on water supply and sanitation, with a separate analysis for each of the components of water supply, water availability, water quality, sanitation and hygiene. The results of the review reveal that improvements in one or more of these components are associated with a substantial decline in diarrhoeal morbidity with an expected reduction in diarrhoea ranging from 16% to 33% for each of the components across the studies reviewed.

In India, sanitation facilities are poor and often lacking, and the availability of water is highly variable. Some households have in-compound access to water, other households are a short walking distance to public water facilities, and yet others are several hours from the nearest facility. At the same time, households may experience interruptions in water supply, which confound sanitation practices. Sanitation is also highly variable across households and is generally a strong correlate of wealth. Wealthy households have a higher probability of having some form of sanitation facility, whereas those belonging to lower income groups or backward communities with few assets are likely to have no sanitation facility. Garbage disposal and collection, another component of sanitation, again, is highly variable, both in rural and urban India (especially urban slums), with some areas having no proper disposal or disposal site, with others having a confirmed deposit place and pick up routine. This paper will raise four questions to be analysed at the household level in children using data from India.

2. Research Questions

This paper assesses the impacts of socioeconomic and environmental factors on diarrhoeal incidence in children. It is important to recognise that morbidity, i.e. illness, is highly subjective and is based on the individual's perception of what constitutes sickness (Doraiswamy, 1998). Hence, individuals could vary in their perception of what is regarded as being sick and what is good health. In an ideal situation, individuals' perceptions or reports on sickness should be matched by medical diagnosis. However, in rural India and urban slums, often no formal medical treatment is sought due to poor accessibility to hospitals and health care centres, to expense, and to lack of medical insurance. In this study, an incidence of diarrhoea will be understood to be one that is reported, and hence perceived, by the respondent (i.e. mother of the child).

- Question 1: What are the impacts of different sources of water, considered individually and jointly with sanitation facilities on child health? Does proximity of the water source influence health outcomes? In their study, Jalan and Ravallion (2003b) do not find much difference in prevalence of diarrhoea between children in households that have piped water inside the premises versus those that draw water from a public tap located outside the periphery of the house. However, they do find a longer duration of diarrhoeal occurrence in households using water from a public tap suggesting contamination from handling and storage.

- Question 2: What are the impacts on child health of public investments in infrastructure across different wealth strata and in rural and urban areas? Are the health gains from different water and sanitation facilities the same for the poor as for the rich? Jalan and Ravallion (2003b) find evidence in favour of reduced diarrhoeal diseases in wealthier households, while Alberini et al (1996) find the effect to be weak due to interruptions in piped water supply.

- Question 3: What are the impacts of education, exposure to mass media, and disease-specific awareness of the homemaker and her spouse on child health? Jalan, Somanathan and Choudhri (2003a) show that improvements in awareness of the health risks associated with environmental factors significantly increase household responses to water purification and in fact exceed the impacts of increased income. This question therefore seeks to analyse the importance of education, the general and disease-specific awareness of the homemaker and the education of her spouse on children's health in a non-experimental situation.

- Question 4: Are the health outcomes from access to different water and sanitation types consistent across girls and boys? Separate to this, we also investigate whether age of the child and child height (to capture anthropometric measures) influences health outcomes.

3. Literature Review

Several studies have examined environmental and socioeconomic factors, including behavioural patterns (such as hand-washing practices), on diarrhoeal morbidity. Notable contributions to the health literature covering environmental and socio-economic variables are reviewed here.

Dasgupta (2004) uses the health production model to evaluate health damages from diarrhoeal morbidity, including children, in the slums of Delhi. Applying probit regression techniques to household-level survey data, the author examines the relationship between diarrhoeal illness, household behaviour and environmental risk factors captured by household -and neighbourhood-level characteristics, such as sources of drinking water, presence of sanitation facilities, availability of water, presence of garbage dumps, income and education of household head. The results indicate that those with access to piped water are less vulnerable to diarrhoeal illness, though access by itself does nothing to modify behaviour. Interestingly, she concludes that presence of sanitation facilities and education of household head also plays no significant part in illness. On the other hand, per capita income plays a part in determining illness, with better off families experiencing reduced illness. Income is correlated with behavioural inputs that are difficult to quantify, such as the ability of households to exercise choice in improving their environmental conditions.

Jalan and Ravallion (2003b) estimate the impacts of piped water on diarrhoeal outcomes in children in rural India using matching techniques. They find a lower incidence of diarrhoea among children living in piped water households but find that these benefits largely bypass low income families and where mothers are poorly educated. This paper extends Jalan and Ravallion's work to cover other water types and sanitation.

Using household-level data in a survey conducted in Jakarta, Indonesia, Alberini et al (1996) examine the impact of environmental variables together with an individual's behaviour on diarrhoeal disease. To measure the incidence of diarrhoeal outcomes in households, they estimate a model using defensive action adopted by individuals and employ proxies for risk of contamination such as water quality and presence of sanitary facilities. Their results indicate that water supply rather than water quality impact diarrhoeal illness by disrupting defensive behaviour. Further, they find that in Jakarta, investments in education and improvements to income (contrary to Dasgupta's findings above) would not achieve the desired reduction in incidence even though these households demonstrate a greater potential for adoption of defensive behaviour.

Feachem (1984) reviews studies examining behavioural factors, specifically hand-washing practices, on diarrhoeal outcomes and hygiene education that potentially alters such behaviour. The review, based on a comparison between treatment and control group behaviour in Bangladesh, USA and Guatemala, suggests that hygiene education is instrumental in reducing illness from diarrhoea.

Using longitudinal data from a household survey conducted in Cebu, Philippines, Akin (1992) examines factors influencing behavioural inputs and further, examines the impact of these inputs together with other exogenous factors on diarrhoeal outcomes in children. The author attempts to correct for unobserved heterogeneity by using family specific fixed effects across individuals. The estimation results show that population density, age of the child, exposure to contaminated water and faeces, and rainfall increase the incidence of diarrhoea. However, anthropometric measures of children do not affect diarrhoeal outcomes.

Bozkurt et al (2003) have looked at the mean incidence of diarrhoea in children under five years in Giziantep, Turkey, using individual and household level variables. They find incidence to be higher amongst children in poor housing conditions, in children with poor individual status defined over a number variables relating to their general health (such as immunisation, birth weight and duration of breast feeding) , amongst mothers without health counselling on diarrhoea, and amongst poorly educated fathers.

Rao, Mishra and Retherford (1998) assess the effects of awareness via the electronic mass media on women's knowledge and thus on their ability to treat and safeguard their children against diarrhoea. Applying multinomial logistic regressions to all India data generated by the National Family Health Survey 1992-93, the authors' analysis leads them to find an inverse relationship between exposure to mass media and knowledge. They also find discrimination in use of oral rehydration treatment, favouring boys over girls.

Van der Hoek et al (2001) assess the impact of water consumption from irrigation sources on diarrhoeal outcomes in southern Punjab, Pakistan. Their results corroborate much of the evidence reviewed by others as they find that risk factors associated with water availability, presence of sanitary facilities, income, education and hygiene practices affect diarrhoeal outcomes.

Duraisamy (2001) examines morbidity in children, adults and elderly, using survey data conducted at the national level in 1994 in rural India. The survey revealed that children are more prone to diarrhoea, thus suggesting the vulnerability of children to the disease. The author estimates a reduced form health production function, incorporated within a utility maximising framework. Using the tobit method of estimation, health outcomes are measured by the number of days ill for all age cohorts and are a function of household level demographic and socioeconomic factors, environmental factors, and village-level infrastructure facilities. The results for children highlight the importance of mother's education, incomes, and presence of a separate kitchen. Environmental variables such as availability of a toilet, hand-washing behaviour, and sources of drinking water were not significant on morbidity. Many were correlated with other infrastructure variables such as with piped water, separate kitchen and separate toilet. Interestingly, the estimates revealed that boys had a higher probability of sickness than girls.

Fewtrell and Colford (2004) have presented an excellent overview of water, sanitation and hygiene interventions on diarrhoeal morbidity. Using a meta-analysis framework, their study examines in detail the analyses presented in 64 papers in developing and developed countries. Furthermore, they examine the evidence on diarrhoeal morbidity for each of the interventions separately as well as for multiple interventions. For developing countries (48 papers spanning 24 countries), the literature reviewed suggests that interventions in water quality, particularly disinfection at point of use, and water supply, particularly for household connections, are effective in reducing diarrhoea. A large number of studies reviewed on hygiene interventions focussed on hand-washing practices and suggested a greater impact on reduced illness of behaviour over hygiene education.

Table 1. Summary of Studies on Child Morbidity from Diarrhoea*

Authors	Data Source	Dependent Variable	Method	Environmental Variables	Other Covariates
Feachem, 1984	Review of 3 studies in USA, Guatemala, & Bangladesh	Diarrhoea morbidity	Review	Not included	Hygiene education is significant, Education, religion are significant for behaviour
Akin, 1992	Household Survey, Cebu, Philippines, 1983-1984.	Incidence of Diarrhoea	Random effects Probit model	In urban areas - contaminated water, exposure to faeces & in rural areas-rainfall significantly increase diarrhea	Breast feeding, age of child, low population density are significant in reducing diarrhoea,
Alberini, Eskeland, Krupnick & McGranahan, 1996	Jakarta Household Survey, 1991	Diarrhoea Illness	Bivariate Probit Model	Interruptions in water supply significantly affects diarrhoea	Weak income effect, No Education effect Behaviour i.e. hand-washing is significant
Rao, Mishra & Retherford, 1998	National Family Health Survey, India, 1992-1993	Diarrhoea Morbidity	Multinomial Logit regression	Safe drinking water , significant for ORS/RHS treatment	Mother's education, Awareness & discrimination against girls for ORS/RHS) in treatment, significant
Van der Hoek, Konradsen, Ensink, Mudasser &	Pakistan Household Survey, 1998-1999	Diarrhoeal Incidence	Logistic Regression	Access to water, presence of a toilet are significant	Age of child, Mother's literacy (negative sign), socio-economic status are

Jensen, 2001					significant
Duraisamy, 2001	National Council of Applied Economic Research, India Rural Survey, 1993-1994	Short Duration Illness	Tobit Model	Presence of toilet & source of drinking water, not significant	Mother's education, income & separate kitchen, significant; hand-washing not significant
Bozkurt, Özgür & Özcirpici 2003	Household survey in Gaziantep, Turkey, 1998-1999.	Diarrhoea morbidity	Mean Incidence	Poor housing conditions significant	Education of parents & individual status of child, significant
Jalan & Ravallion, 2003b	National Council of Applied Economic Research, India Rural Survey, 1993-1994	Diarrhoea Morbidity	Matching techniques using the propensity score	Access to piped water is significant	Income & mother's, education are significant. Strong interaction effects between income and mother's education.
Dasgupta, 2004	Household survey in urban slums of Delhi, 1998	Diarrhoeal Illness	Bivariate Probit Model	Source of drinking water & removal of garbage significant, presence of latrines not significant	Income significant education of household head, not significant
Fewtrell & Colford, 2004	Meta Analysis	Diarrhoea Morbidity	Pooling using Random & Fixed Effects Model	Water quantity & sanitation interventions are effective	Disinfection of water, hygiene education are significant

*** *In the summary presented above, only those parts of the studies relevant to diarrhoea in children are highlighted. Hence, in some cases they only form a part of the unit under analysis and part of the research study by authors.*

3.1 Contribution to the Literature

The research in this paper adds to the body of work reviewed by examining health outcomes in children through the use of propensity score and matching techniques that compare outcomes between those households that have certain facilities with those that do not, in *similar* environments. Since diarrhoea can be spread due to a number of reasons not all relating to infrastructure variables, the results stemming from a comparison of households along socioeconomic and environmental dimensions provide for a richer analysis.

In particular we build on Jalan and Ravallion's (2003b) work and look at the impacts of water variables individually and jointly with the presence of sanitation facilities to assess impacts in child health using matching techniques. We also assess the importance of social and economic variables that potentially impact diarrhoeal outcomes in children. Further, we also adopt alternate methods to check for the robustness of results. Finally, we assess marginal impacts and interaction effects between the various water types and sanitation facilities on diarrhoeal outcomes in children.

The results show the importance of disease-specific awareness in reducing diarrhoeal incidence in young children. The importance of this finding suggests orienting public health campaigns that raise diarrhoeal awareness in mothers in order to effectively reduce morbidity in young children from this disease. The results also indicate that piped water households have a higher incidence of diarrhoeal outcomes in children. We relate this finding to the fact that piped water, which is supplied by government run utilities in India, can have several sources of contamination even after being treated such as unclean storage tanks and water distribution networks.

4. Methodology

The incidence of diarrhoea is estimated using a health model, the origin of which can be traced back to Grossman (1972), where health is treated as a form of human capital and where "good health" produces the output of healthy days. Health enters the individual's intertemporal utility function directly, with ill-health as a measure of disutility, and indirectly, by reducing the total time available for work and leisure. Subsequent applications and revisions to the health model have been made since (Cropper, 1981; Gerking and Stanley, 1986; Harrington and Portney, 1987; Harrington et al, 1989; Akin, 1992; Alberini et al, 1996; Dasgupta, 2004).

Most notably, Cropper (1981) built on Grossman's work and introduced an air pollution variable to the health model to estimate individuals' willingness to pay for reduced air pollution. Gerking and Stanley (1986) use an alternate specification of the health production model, with medical expenditures described as a function of health stock. Harrington and Portney (1987) employ a static model of constrained utility maximisation in which time spent ill is included as a choice variable in the utility function, and which depends on an individual's defensive behaviour and environmental factors. An application of this model to contaminated water sources and its effect on worker productivity is examined in Harrington et al (1989). Health related consumer goods and defensive behaviour have also been employed in determining diarrhoeal outcomes (Alberini et al, 1996; Dasgupta, 2004). Akin

(1992) uses a child health production function to estimate the impact of behavioural inputs and exogenous factors on diarrhoeal outcomes. Children's health enters the utility function of their parents who maximise utility subject to income constraints.

4.1 The Health Model in a Static Framework

The health production function relates health outcomes to exogenous variables, such as those relating to existing environmental conditions like the presence of underground sewage systems, and over which individuals do not have control. It also relates health to choice variables such as preventive action and medical expenses that are consciously determined by agents' behaviour, and are influenced by information, education, age and income.

In the health model, individuals are utility maximisers and choose optimal health outcomes based on their input choices that minimise their costs of producing good health. In this setting, it is assumed that parents know the health production functions of their children. In the author's opinion, this is not unreasonable as mothers know their child's health status and make decisions based on the information available to them alongside other constraints.

Jalan and Ravallion (2003b) use the utility maximising framework to theorise the impacts of piped water on children's diarrhoeal outcomes. We apply and extend their theoretical framework to cover the impact of environmental variables on diarrhoeal outcomes in children. The environmental variables considered are drinking water sources and sanitation facilities. The objective of the study is to estimate the individual and joint impacts of these environmental variables on health with parents as utility maximisers. Following Jalan and Ravallion (2003b) the child health production function is described by

$$h = h\ (e, s, z, \mu) \qquad (1)$$

where e is the vector of environmental variables, e = (water, sanitation); s is expenditure by parents on their children's health and represents that part of income spent on goods that improve their children's health such as in having a certain water and sanitation facility; z is a vector of child-specific and socioeconomic variables such as age, gender of the child, height of the child, education or awareness of the mother, income of the household; and μ is a vector of unobserved variables. The function h is assumed to be twice continuously differentiable,

$\partial h/\partial e > 0$, $\partial h/\partial s > 0$, $h_{ee} < 0$, $h_{ss} < 0$ and $\partial h/\partial z >< 0$ (depending on the particular variable) Parents derive utility from improved access to environmental variables, from net income for non-health consumer goods and from other factors embodied in *z*. Expenditure *s* impacts child health, and parents indirectly derive utility in consumption from children's health. Children's health is thus added separately to that of their parents. As a variation to this, others in the literature reviewed have assumed that children's health directly enters the utility function (Rosenzweig and Wolpin, 1986; Akin, 1992). Expenses, *s*, incurred by parents on their children's health and consumption *c* of non-health consumer goods together make up exogenous income *y*. Hence, $y - s = c$ is left for parents to spend on non-health consumption goods and forms the budget constraint. The parents' problem is described by choice of *s* which maximises their utility

$$Max_{s} : U = u\,(e, y\text{-}s\text{=}c, z) + h\,(e, s, z, \mu) \qquad (2)$$

where *U* is assumed to be twice continuously differentiable in $(y - s)$ and strictly increasing in $(y - s)$ and *e*. (Note: *e* is treated as a continuous variable for the sake of exposition)

$\partial U/\partial e > 0$, $\partial U/\partial\,(y\text{-}s) > 0$, $U_{y\text{-}s,\,y\text{-}s} < 0$ (diminishing marginal utility from income)

The solution to the maximisation problem is solved from the first order condition by setting $\partial U / \partial s = 0$ to give

$$-U_y(e, y - s, z) + h_s(e, s, z) = 0 \text{ or} \qquad (3a)$$

$$U_y(e, y - s, z) = h_s(e, s, z) \qquad (3b)$$

which equates the marginal utility of income to the marginal gain to children's health from a unit change in expenditure, *s*. The optimal $\hat{s} = s(e, y, z)$ is associated with maximum utility produced by substituting in the parents' utility function

$$\upsilon(e, \hat{s}, z) \equiv u\,[\,e, y - s(e,y,z), z] + H(e, s, z)\text{.....parent} \qquad (4)$$

and optimal child health described by

$$H(e, y, z) \equiv h(e, \hat{s}, z) = h(e, s(e,y,z), z)\text{.........child} \quad (5)$$

4.1.1 Environmental Effect on Child Health

Using the envelope theorem (applied to points in the neighbourhood around the maximum), an increase in *e*, arising from an improvement in drinking water source and/or access to sanitation facilities, is associated with an increase in *U*. However, the directional change across the two components of the parents' utility function may not be similar. Decomposing the effect of an improvement in *e*, the impact on children's health at the optimum

$$\partial H / \partial e = H_e = h_e + h_s s_e \quad (6)$$

where s_e is obtained from the first order condition (equation (3b))

$$U_y(e, y - s, z) = h_s(e, s, z) \quad (3b)$$

$$U_{ye} - U_{yy} s_e = h_{Se} + h_{SS} s_e \quad (7)$$

$$s_e = \frac{U_{ye} - h_{Se}}{U_{yy} + h_{SS}} \quad (8)$$

Recalling that $U_{yy} < 0$ and $h_{ss} < 0$, the denominator for s_e is negative. The outcome therefore depends on the numerator which can be positive, negative or zero. Since h_e and h_s is positive in equation 6 (from our assumptions), all that is needed for an improvement in child health is for s_e >0 arising from $U_{ye} - h_{Se} < 0$ or the impact of access to improved water or sanitation facilities to parents' marginal utility from income is exceeded by the impact of improved access to the marginal gain in child health from unit expenditure on health goods.

4.1.2 Income Effect on Child Health

The income effect of the health gain to children from environmental variables is given by

$$\frac{\partial H / \partial e}{\partial y} = \frac{1}{\partial y}\left(H_e = h_e + h_s s_e\right) \quad (9)$$

$$\equiv H_{ey} = h_{eS}s_y + h_s s_{ey} + s_e h_{ss} s_y \tag{10a}$$

$$\equiv H_{ey} = s_y\left(h_{eS} + s_e h_{ss}\right) + h_s s_{ey} \tag{10b}$$

where s_Y is obtained from the first order condition (equation (3b))

$$U_y(e, y - s, z) = h_s(e, s, z) \tag{3b}$$

$$U_{yy} - U_{yy}s_y = h_{ss}s_y \tag{11}$$

$$0 < s_y = \frac{U_{yy}}{U_{yy} + h_{ss}} \leq 1 \tag{12}$$

An additional theoretical assessment of the impact of income on child health using alternate specifications of child health and parents utility is treated below separately.

4.1.3 Socioeconomic Effect on Child Health

The effect of socioeconomic variables, captured by *z*, on the health gain to children from environmental variables is

$$\frac{\partial H / \partial e}{\partial z} = \frac{1}{\partial z}\left(H_e = h_e + h_s s_e\right) \tag{13}$$

$$\equiv H_{eZ} = h_{eZ} + h_{eS}s_Z + h_s s_{eZ} + s_e\left(h_{ss}s_z + h_{sZ}\right) \tag{14a}$$

$$\equiv H_{eZ} = \left(h_{eS} + s_e h_{ss}\right)s_Z + h_{eZ} + h_s s_{eZ} + h_{sZ} \tag{14b}$$

where h_S >0 and h_{SS}< 0 and where s_z is obtained from the first order condition, equation (3b)

$$U_y(e, y - s, z) = h_s(e, s, z) \tag{3b}$$

$$U_{yZ} - U_{yy}s_Z = h_{sZ} + h_{ss}s_Z \tag{15a}$$

$$\equiv s_Z = \frac{\left(U_{yz} - h_{Sz}\right)}{U_{yy} + h_{ss}} \tag{15b}$$

The sign of s_Z therefore depends on whether $U_{yZ} - h_{sZ}$ is positive, negative or zero (the denominator is negative as before). The impact of socioeconomic variables, in equation (14b) is not so straightforward. Not only does it depend on the sign of s_Z but also on the signs

h_{eZ}, s_{eZ}, h_{eS} and h_{sZ}. The signs in turn will also depend on the particular socioeconomic variable chosen such as education, awareness, and age.

4.1.4 Child-Specific Effect on Child Health

Child-specific effect such as the age or height of the child is assessed by specifying a child health production function that includes this variable separately: $h = h(e, s, z, n, \mu)$; where n = child-specific status captured by child height. The parents' problem:

$$Max_s : U = u(e, y - s, z) + h(e, s, z, n, \mu) \tag{16}$$

From the first order conditions

$$U_Y(e, y - s, z) = h_s(e, s, z, n) \tag{17}$$

Optimal s is characterised by $\hat{s} = s(e, y, z, n)$, which when substituted into the child health production function gives optimal child health

$$H(e, s, z, n) \equiv h\left(e, \hat{s}, z, n\right) = h[e, s(e, y, z, n), z, n] \tag{18}$$

The child-specific impact of a health gain to children from environmental variables is captured by

$$\frac{\partial H / \partial e}{\partial n} = \frac{1}{\partial n}\left(H_e = h_e + h_s s_e\right) \tag{19}$$

$$H_{en} = h_{en} + h_{eS} s_n + h_s s_{en} + s_e\left(h_{ss} s_n + h_{Sn}\right) \tag{20}$$

where s_n is obtained from the first order condition, in equation (17)

$$U_Y(e, y - s, z) = h_s(e, s, z, n) \tag{17}$$

$$-U_{yy} s_n = h_{ss} s_n + h_{Sn} \tag{21}$$

$$\equiv s_n = \frac{-h_{Sn}}{U_{yy} + h_{ss}} > 0 \tag{22}$$

As before, $U_{yy} + h_{SS} < 0$, and $s_n > 0$, if $h_{Sn} > 0$
In equation (22) s_n shows how the optimal s or expenditure varies with child-specific status of the child. If a child is very sick, the family may choose to spend a lot more[2] on this child. Similarly, h_{Sn} shows how the health impact of s varies with the child-specific status.

Hence we see from equation (20) that the effect of child-specific status to the health gain to children from environmental variables depends on a number of factors: h_{en}, h_{eS}, s_n, s_{en}, h_{Sn} and h_{ss}. We consider three cases stemming from three possible scenarios for s_e:

- If $s_e = 0$ then equation (20) is reduced to

$$H_{en} = h_{en} + h_{eS} s_n \tag{23}$$

and the child specific impact depends on h_{en} and h_{eS}

- If $s_e > 0$ then rewriting equation (20) below

$$H_{en} = h_{en} + h_{eS} s_n + h_s s_{en} + s_e (h_{ss} s_n + h_{Sn}) \tag{24}$$

where $s_e h_{ss} s_n < 0$.[3] The final outcome, H_{en}, in equation (24) will depend on h_{en}, h_{eS}, s_{en} and h_{Sn} respectively and the net impact of each of these interactions. For $H_{en} > 0$, the net impact must be positive (and vice versa) and will imply that the healthier the child, the greater the gain to child health from an improvement in water and sanitation conditions.

- If $s_e < 0$, then rewriting equation (20) below

[2] It can be argued that parents may choose to spend less on a dying child and spend it on other living and healthy children.
[3] This result stems from $h_{ss} < 0$ and $s_n > 0$ assuming that $h_{sn} > 0$.

$$H_{en} = h_{en} + h_{eS}s_n + h_s s_{en} + s_e\left(h_{ss}s_n + h_{Sn}\right) \quad (25)$$

where $s_e h_{ss} s_n > 0$. As before, the final outcome H_{en} will depend on h_{en}, h_{eS}, s_{en} and h_{Sn} respectively and the net impact of each of these interactions. For $H_{en} > 0$, the net impact must be positive and vice versa.

The latest population census in India shows large variations in the sex ratio with some states showing a preference for boys over girls (e.g. in the north Indian states of Punjab and Haryana). Keeping such a bias in mind, we attempt to determine the impact of a child's gender on their health. We do not model it rather we determine it empirically since gender is not a continuous variable that can be used to show the marginal impacts of the gender of a child.

4.2 Alternate Specifications of the Health Models

Additional cases are introduced below that alter children's health functions and parents' utility models for a richer and more varied analysis of the possible effects of access to different types of environmental variables. Section 4.2.1 looks at the income effect on child health. Under certain conditions defined by the nature of parents' utility function and the child health function, there will be no pronounced effects of income on diarrhoeal outcomes, thus mitigating any predicted impacts. Section 4.2.2 discusses an alternate situation where the income effect can be positive or negative. Finally, section 4.2.3 specifies another form of the parents' utility function where child health enters their utility function directly.

4.2.1 No Interaction effects

(a) no interaction effects between parental utility and *e* i.e. $U = (y - s, z)$ e.g. when parents care about their children's health but not their own.

(b) no interaction effects between expenditure *s*, on child health i.e. $h(e, z, \mu)$ e.g. hand washing soap is not used by children but by their parents.

In this case, parents' utility function is $U = u(y - s, z) + h(e, z, \mu)$, where parent's utility does not depend on *e* and the child's health does not depend on *s*. Parents maximise utility and the solution to the maximisation problem is obtained from the first order condition by setting $\partial U / \partial s = 0$

$$\Rightarrow U_y(y-s,z)=0 \tag{26}$$

Since *U(.)* is not a function of *e* and *h(.)* is not a function of *s*, we can say that that U_{ye} and h_{se} do not exist in equation (8). (Note: U_e does not exist and so $U_{ey} = U_{ye}$ does not exist; similarly, h_s does not exist and so h_{se} does not exist.) Using equation (8), we can say s_e does not exist, as the numerator is zero, and so s_{ey} does not exist.

Solving for H_{ey} using equation (10b), $H_{ey} = s_y(h_{eS} + s_e h_{ss}) + h_s s_{ey}$ and inserting the relationships observed, $s_e = 0$; $U_{ey} = U_{ye} = 0$; $h_s = 0$; $h_{se} = 0$; $s_{ey} = 0$, we obtain

$$H_{ey} = 0 \tag{27}$$

The health gain to children from income is zero or independent of it. This could happen in a situation where interrupted supply of water is the main impediment to child health and not income, as was found by Alberini et al (1996) in Jakarta.

Similar results are obtained by assuming no interaction effects between expenditure *s*, on child health i.e. *h(e, z, μ)* while maintaining interaction effects between parental utility and *e* i.e. *u(e, y-s, z)*. Parents' utility function is described by:

$$U = u(e, y-s, z) + h(e, z, \mu) \tag{28}$$

In this case h_S does not exist, and hence the first order conditions obtained by setting $\partial U / \partial s = 0$ give

$$U_y(e, y-s, z) = 0 \tag{29}$$

Optimal *s* is characterised by $\hat{s} = s(e,y,z)$, which is no longer substituted in children's' health function as it no longer effects their health.

The income effect on child health in this case can be solved for H_{ey} using equation (10b), $H_{ey} = s_y\left(h_{eS} + s_e h_{ss}\right) + h_s s_{ey}$. Since *h(.)* is not a function of *s*, we can say $h_s = 0$; and that h_{eS} does not exist, all elements of H_{ey} will be zero irrespective of the sign of s_y and of s_{ey}.

4.2.2 Additive Separability and Expenditure on Health Goods

(c) parental utility is additively separable in consumption and environmental variables: *U(y-s, z) + U(e)*; this is relevant for the case where the marginal utility of *e* is independent of other arguments in *U*. An example of this is provided by the introduction of an improved source of drinking water that benefits the entire community.

(d) and the marginal propensity to spend on inputs to children's health is unaffected by access to the particular water / sanitation source $s_{ye} = 0$

In this case, parents' utility function is

$$U = U(y\text{-}s, z) + U(e) + h(e, s, z) \qquad (30)$$

From equation (8), $s_e = \dfrac{-h_{Se}}{U_{yy} + h_{ss}}$ as $U_{ye} = 0$ from the utility function

From equation (12), $s_y = \dfrac{U_{yy}}{U_{yy} + h_{ss}} \Rightarrow \left(U_{yy} + h_{ss}\right) = \dfrac{U_{yy}}{s_y}$ and substituting in s_e above,

$$\Rightarrow s_e = \frac{-h_{Se}}{U_{yy}} s_y \qquad (31)$$

Solving for H_{ey} using equation (10b), $H_{ey} = s_y\ (h_{eS} + s_e h_{SS}) + h_S s_{ey}$, and substituting in equation (31)

$$H_{ey} = s_y\left(h_{eS} + s_e h_{ss}\right) \text{ as } s_{ye} = 0 \text{ (by assumption)}$$

and by substituting for s_e, we get

$$H_{ey} = s_y \left(h_{eS} - \frac{h_{Se}}{U_{yy}} s_y h_{ss} \right) \tag{32}$$

From equation (12) we get

$$s_y = \frac{U_{yy}}{U_{yy} + h_{ss}} \Rightarrow s_y \left(U_{yy} + h_{ss} \right) = U_{yy} \Rightarrow U_{yy} = \frac{h_{ss} s_y}{\left(1 - s_y\right)} \tag{33}$$

Substituting back into equation (32)

$$H_{ey} = s_y \left[h_{eS} - \frac{h_{Se}}{h_{ss} s_y} s_y h_{ss} \left(1 - s_y\right) \right] \tag{34}$$

$$H_{ey} = s_y^{\ 2} h_{Se} \tag{35}$$

In this case, the effect of environmental variables on the marginal impact of income on child health can be either positive or negative and will depend on h_{Se} in equation (35). If $h_{Se} > 0$ implying that better access to water or sanitation facilities will increase the impact of expenditure on child health, then impacts of income are positive. For example, the presence of a sewage system will greatly enhance the impact of an investment in sanitation facilities. On the other hand if $h_{Se} < 0$ then an increase in environmental variables can lower the income impacts on child health. For example, in-house access to water may lower the impact of disinfection measures

4.2.3 Modified Utility Function

In this case I present a modified utility function for parents where child health directly enters their utility function i.e. $U = u(e, y\text{-}s, z, h\ (e, s, z, \mu))$. By expressing the utility function in this way, we are introducing the possibility of interaction effects between child health and other arguments in parents' utility function. In the previous case where child health was additively separable, such interaction effects are assumed away. The solution to the maximisation problem is defined by setting $\partial U / \partial s = 0$ which gives

$$-U_y\left(e, y - s, z, h(e, s, z)\right) + U_h h_s(e, s, z) = 0 \tag{36}$$

$$U_y(e, y-s, z, h(e,s,z)) = U_h h_s(e,s,z) \quad (37)$$

At equilibrium, marginal utility of income = marginal utility due to gains in child health accrued due to an increase in health expenditure.

Optimal *s* is therefore given by $\hat{s} = s(e, y, z)$ which gives the maximum utility for parents obtained through substitution in their utility function and optimal child health

$$v(e,\ s,\ z, h(.)) \equiv u\ [\ e,\ y - s(e,y,z),\ z, h(e,s(e,y,z),z)] \quad \text{.....parent} \quad (38)$$

$$H(e,\ y,\ z) \equiv h(e,\ \hat{s}\ ,\ z) = h(\ e,\ s(e,y,z),\ z) \quad \text{.....child} \quad (39)$$

Using the modified utility function for parents I assess the environmental effect on child health and the impacts of income and socioeconomic variables on child health from improved access to environmental variables.

4.2.3.1 Environmental Effect on Child Health

The impact on children's health at the optimum is given by

$$\partial H / \partial e = H_e = h_e + h_s s_e \quad (40)$$

where s_e is obtained from the first order condition in equation (37)

$$U_y(e, y-s, z, h(e,s,z)) = U_h h_s(e,s,z) \quad (37)$$

$$U_{ye} - U_{yy} s_e + U_{yh} h_e + U_{yh} h_s s_e = U_h (h_{se} + h_{ss} s_e) \quad (41)$$

$$U_{ye} + U_{yh} h_e - U_h h_{se} = U_h h_{ss} s_e + U_{yy} s_e - U_{yh} h_s s_e \quad (42)$$

$$s_e = \frac{U_{ye} + U_{yh} h_e - U_h h_{se}}{U_h h_{ss} + U_{yy} - U_{yh} h_s} \quad (43)$$

Comparing equation (43) with equation (8), we find the expression in equation (43) to be fairly complex without a clear direction on the final outcome. Using $U_{yy} < 0$, $h_{ss} < 0$, $h_s > 0$ and assuming $U_h > 0$ the sign of the denominator depends on U_{yh}. If $U_{yh} > 0$ then the denominator is negative. If $U_{yh} < 0$ is large enough to outweigh the negative impacts arising from the rest of the terms then the denominator will be positive. Less can be said about the numerator where a positive or negative outcome depends on the effects of U_{ye}, U_{yh}, and h_{se}

4.2.3.2 Income Effect on Child Health

The income effect of a health gain to children from environmental variables is given by

$$\frac{\partial H / \partial e}{\partial y} = \frac{1}{\partial y}(H_e = h_e + h_s s_e) \tag{44}$$

$$H_{ey} = h_{ey} + h_{es} s_y + h_{ss} s_y s_e + h_s s_{ey} + s_e h_{sy} \tag{45}$$

Substituting for s_y from the first order conditions in equation 37

$$U_y\big(e, y - s, z, h(e, s, z)\big) = U_h h_s\big(e, s, z\big) \tag{37}$$

$$U_{yy} - U_{yy} s_y + U_{yh} h_s s_y = U_h h_s s_y \tag{46}$$

$$U_{yy} = s_y (U_{yy} - U_{yh} h_s + U_h h_s) \tag{47}$$

$$s_y = \frac{U_{yy}}{U_{yy} - U_{yh} h_s + U_h h_s} \tag{48}$$

The numerator is negative as $U_{yy} < 0$. In the denominator, $U_{yy} < 0$ and $h_s > 0$. The final impact will rest on U_{yh} assuming $U_h > 0$. The impact on child health H_{ey} therefore is less clear as it depends on a number of effects including on s_y and s_e that can be seen in equation

(45). We find similar complications when obtaining the socioeconomic and child-specific effect.

4.2.3.3 Socioeconomic Effect on Child Health

The socioeconomic effect of a health gain to children from environmental variables is given by:

$$\frac{\partial H / \partial e}{\partial z} = \frac{1}{\partial z}(H_e = h_e + h_s s_e) \tag{49}$$

$$H_{ez} = h_{ez} + h_{es} s_z + h_{ss} s_z s_e + h_s s_{ez} + s_e h_{sz} \tag{50}$$

Substituting for s_z from the first order conditions in equation (37)

$$U_{yy}(e, y - s, z, h(e,s,z)) = U_h h_s(e,s,z) \tag{37}$$

$$U_{yz} - U_{yy} s_z + U_{yh} h_s s_z + U_{yh} h_z = U_h h_s s_z + U_h h_{sz} \tag{51}$$

$$U_{yz} + U_{yh} h_z - U_h h_{sz} = U_h h_s s_z + U_{yy} s_z - U_{yh} h_s s_z \tag{52}$$

Collecting terms for s_z we get

$$s_z = \frac{U_{yz} + U_{yh} h_z - U_h h_{sz}}{U_h h_s + U_{yy} - U_{yh} h_s} \tag{53}$$

The sign of the denominator depends on U_{yh} as well as on the magnitude of the other terms. Less so can be said for the numerator. The impact on child health H_{ez} depends on a number of factors and the exact outcome depends on the interplay and magnitude of each of the factors in equation 53.

4.2.3.4 Child-Specific Effect on Child Health

The child-specific effect of a health gain to children from environmental variables is given by

$$\frac{\partial H / \partial e}{\partial n} = \frac{1}{\partial n}\left(H_e = h_e + h_s s_e\right) \tag{54}$$

$$H_{en} = h_{en} + h_{eS} s_n + h_s s_{en} + s_e\left(h_{ss} s_n + h_{Sn}\right) \tag{55}$$

Substituting for s_n from the first order conditions in equation 37

$$U_y\left(e, y - s, z, h(e, s, z, n)\right) = U_h h_s\left(e, s, z, n\right) \tag{37}$$

$$-U_{yy} s_n + U_{yh} h_s s_n + U_{yh} h_n = U_h h_s s_n + U_h h_{sn} \tag{56a}$$

$$-U_{yy} s_n + U_{yh} h_s s_n - U_h h_s s_n = U_h h_{sn} - U_{yh} h_n \tag{56b}$$

Collecting terms for s_n we get

$$s_n = \frac{U_h h_{sn} - U_{yh} h_n}{-U_{yy} + U_{yh} h_s - U_h h_s} \tag{57}$$

As was with the case of socioeconomic variables, we arrive at a similar conclusion that the impact on child health stemming from child-specific status H_{en} depends on a number of factors including s_n and s_e.

We thus discover that by using a utility function for parents which is not additively separable in child health results in fairly complex outcomes with no clear theoretical conclusions on the direction of impacts. Utility functions that are additively separable are frequently used for computational and intuitive simplicity and our theoretical exercise above indicates the same.

5. Estimation Methods and Techniques

Each of the research questions raised in Section 2 are analysed by employing the matching technique using the propensity score method (PSM), where households with access to different water sources (piped water, well water, and handpump water) and presence of sanitation facilities will be matched across a vector of characteristics to facilitate comparison with *observationally identical* households in the control group that did not have access to these facilities. Thus the effect of environmental variables will be isolated. The advantage of using PSM lies in reducing biased estimates due to observed heterogeneity.[4] The fundamental question that summarises the research hypothesis raised above is the following: if water facility X had not been present, what would have been the health outcomes in children?

5.1 Programme Evaluation: Finding the Right Counterfactual

Programme evaluation has been an important tool in determining the impact of various public interventions. Evaluations inform the policy maker of the benefits to the program participants through a comparison of participants and non-participants, thereby allowing an assessment of the returns to investing in the particular programme. The evaluation can only be examined if it is known what would have happened in the absence of the programme. While it is possible to observe gains to participants before and after programme implementation, often researchers do not have the benefit of such farsighted data collection process. More often than not, researchers have to contend with data that arrives post-intervention and to analyse its benefits to programme participants. Since it is difficult to answer post-intervention the question, what if the programme had not been implemented and the participants had not participated, the counterfactual non-participant group assumes utmost importance.

Naïve estimates of programme evaluation based on a comparison of outcomes between participant and counterfactual groups suffer from two main sources of bias. These biases stem from differences in observable characteristics and from unobservable ones (also called as selection bias), respectively, thus giving rise to distinctions between the treatment and the non-treatment groups. The two sources of bias may not work in the same direction. Hence, eliminating bias from one source may not solve the problem.

Various methods exist that attempt to reduce these sources of bias. These are randomised methods, matching, double differencing and instrumental variables (Ravallion, 2001). In this paper we employ the matching technique as the sample selection procedure was stratified following a number of select variables.[5] For example, rural areas were divided into geographic regions and villages were stratified according to female literacy, percentage of males working in the non-agricultural sector, and percentage of the population belonging to a

[4] Refer to section 5.3 for a comparison of the propensity score method with other techniques

[5] Note that we cannot perform the likelihood ratio test for independent equations for a Heckman probit model as the latter is derived with an assumption of simple random sampling which is violated when using complex survey data.

scheduled castes or tribes. Likewise, for sample selection in urban areas, districts were arranged according to female literacy levels and wards were then chosen systematically with probability proportional to size. Further, for this study, the author believes that households select themselves to have a certain type of water and sanitation facility which is likely to determine health outcomes in children. Using the jargon in the evaluation literature, this means that programme placement is non-random and the assumption of conditional independence would be violated i.e. programme participation (in our case having a certain type of water / sanitation facility) is not independent of outcomes.

Hence, drawing a comparison between *treatments* defined as those children living in households with facility W conditional on certain characteristics X versus *controls* defined as those children living in households that do not have facility W to assess programme impacts could bias comparisons. In this case, we would be attempting to assess programme impacts of water / sanitation types by comparing children living in households that are not similar. Recall that the main question asked is: what would health outcomes in children be had they not had access to facility W? The propensity score and matching method attempts to balance treatments with controls by selecting children from those households with facility W with those children in observationally identical households without facility W.

5.2 Matching Techniques

Matching techniques involve the selection of a non-participant group from a larger survey which forms a comparison group to answer the question of what if the participants had not enrolled in the program. In other words, the impacts of intervention are measured by comparing a treatment group with a counterfactual or control group chosen on the basis of observable characteristics that are likely to affect programme participation. Comparison can be made on a single or a variety of characteristics. While such matching across units of observations is feasible for one or two characteristics, for several variables it is a complex task. However, the propensity score method of matching provides researchers with a feasible solution. The technique involves matching based on the predicted probabilities of participation, conditional on observed characteristics. Each non-participant thus receives a score which is then matched with the scores of the treatment units. Programme impacts are then assessed by obtaining mean impacts of outcomes for treatments over their matched counterfactuals.

5.3 Other Evaluation Methods versus the Propensity Score

Before describing the propensity score method in detail, a brief exposition of other non-experimental methods that are commonly used in evaluating programme impacts and the specific advantages of the propensity score merits attention. For this, we draw on the work of Jalan and Ravallion (2003b) and Ravallion (2001) where a summary of the different methods adopted to evaluate programme impacts and the advantages of the propensity score are outlined.

Frequently, average outcomes are assessed between households that have a certain facility with those that do not by employing some form of matched comparison, often using control variables to match households to tackle the problem of heterogeneity between them. Failure to take account of this heterogeneity could make impact assessment inaccurate. Drawing a comparison across households using one or two control variables is feasible but it is clearly difficult when using a range of characteristics. The propensity score method allows a convenient way to compare households for a wider dimension of observed characteristics rather than a limited number. Instead of finding a matched control with a participating household that has the same value of the particular control variable(s), the propensity score provides the same result by matching on the probability of participating given the control variable(s).

The second method employed is to regress outcomes on a range of independent variables and a dummy variable for programme participation. Or in other instances, programme placement is treated as endogenous and further regressed on a number of variables that are correlated with placement but are exogenous to outcomes (commonly referred to as instrumental variable regression).[6] Both these methods impose a particular functional form and are run for the whole sample. The propensity score method does away with this requirement by obtaining mean impacts between households selected on the basis of propensity scores generated for the matched sample only.

[6] In the case of instruments, one has also to respect the exclusion restriction i.e. to find a valid instrument that is independent of outcome given participation. Jalan and Ravallion (2003b) state this to be a particular challenge for cross-sectional data as it is often the case that one can find variables that are correlated with programme participation (or households having a certain water / sanitation facility) to address the endogeneity issue, but question on *a priori* grounds whether those variables are uncorrelated with programme outcomes given participation.

5.4 The Propensity Score Method

The primary question being asked is what would have been the health outcomes in children in the absence of a particular water /sanitation facility? Consider child *j* from a treatment group, which for the purpose of the study is characterised by access to a certain kind of water / sanitation facility, *e*. Children belonging to households with access to *e* are denoted by D = 1, and those with no access to *e* are grouped as D = 0. Child *j* has a health outcome h_{J1} if she receives treatment and h_{J0} if she does not. The question therefore seeks to determine the gain in health $h_{J1} - h_{J0}$. A popular method of estimating impacts is to determine the average treatment effect or the conditional mean impact of treatment, described by the expected gain in health, $E(h_{J1} - h_{J0}) \mid D = 1$.

Often researchers do not have the benefit of obtaining data prior to the implementation of a project that would generate data on h_{J0}. This poses a problem when evaluating the impacts of a certain intervention across programme beneficiaries. The advantage of the propensity score method lies in its ability to facilitate an evaluation of projects in the absence of such baseline pre-intervention data and by comparing groups along certain characteristics post-intervention.

This method allows the comparison of a treatment group, which in this study is described by households with access to certain water facilities and sanitation infrastructure, with a control or non-treatment group. The two groups are matched using a vector of characteristics, X_c= (education, wealth, age etc). Matching takes place by comparing the distributions of the vector of observed covariates and finding the closest match based on their propensity scores (Rosenbaum and Rubin, 1983). The score is the predicted probability of the household to participate in a given programme, which, in this study would be the predicted probability of having access to the particular water /and sanitation facility.

5.4.1 Modelling the Propensity Score

Consider a treatment group *j = 1….T* households and a control group *i = 1……NT* households.[7] The treatment group is characterised by access to a particular water / and sanitation facility whereas the control group is not. Using a discrete variable, the two groups are classified as *D=1* for treatment households and *D=0* for control group households.

[7] T= treatment: NT=non-treatment

Programme outcomes for the two groups are described by H_1 and H_0. Let X_c be the vector of *observed* pre-intervention covariates for the i^{th} household. A balancing score is defined as a function of covariates $b(X_c)$ such that the conditional distribution of X_c given $b(X_c)$ is the same for treated ($D=1$) and control groups ($D=0$).

$$(X_c) \perp\!\!\!\perp D \mid b(X_c) \qquad \text{(P1)}$$

The balancing score can be used to compare treatments with controls in non-experimental settings. Rosenbaum and Rubin (1983) show that the balancing score is a propensity score.[8] The propensity score is described by

$$P(X_{ci}) = pr\,(D = 1 \mid X_{ci}), \qquad 0 < pr\,(D = 1 \mid X_{ci}) < 1 \qquad \text{(P2)}$$

which is the conditional probability of household i having access to water/and sanitation facility, conditional on X_c. They also establish that if treatment assignment is "strongly ignorable" or independent over all units and if outcomes are independent ($\perp\!\!\!\perp$) of participation given X_c, then outcomes are also independent of participation given $P(X_{ci})$ (Jalan and Ravallion, 2003b).

$$(H_1, H_0) \perp\!\!\!\perp D \mid X_{Ci} \;\text{ and }\; 0 < pr\,(D = 1 \mid X_{Ci}) < 1 \;\text{ for all } X_C \qquad \text{(P3)}$$

then $$(H_1, H_0) \perp\!\!\!\perp D \mid p(X_{Ci}) \text{ and } 0 < pr\,(D = 1 \mid p(X_{Ci}) < 1 \qquad \text{(P4)}$$

Furthermore, they establish that under these conditions, the expected difference in outcomes for the treatment and comparison groups equals the average treatment effect or

$$E\,(H_1 \mid p(X_{Ci}), D=1) - E\,(H_0 \mid p(X_{Ci}), D=0) = E\,(H_1 - H_0 \mid p(X_{Ci})) \qquad \text{(P5)}$$

Under these conditions, matching using propensity scores over the entire sample will produce unbiased estimates of treatment effects. As a consequence of propensity score matching, the distribution of the observed X_C across the two groups will be the same. Thus any bias arising

[8] Rosenbaum and Rubin (1983), Theorem 1, pg 43.

from differences in observed characteristics will be eliminated. Bias from unobservables will still remain.

For non-experimental data, Heckman, Ichimura and Todd (1997) show that if outcomes are independent of participation, then non-participants' outcomes have the same distribution as participants'

$$F\,(h_0 \mid X_C\,,\, D=1) = F\,(h_0 \mid X_C\,,\, D=0) \qquad \text{(P6)}$$

Using mean values, this can be written as

$$E\,(h_0 \mid X_C\,,\, D=1) = E\,(h_0 \mid X_C\,,\, D=0) \qquad \text{(P7)}$$

implying that health outcomes for the control group can serve as a counterfactual to show health outcomes in participants had they not participated.

To estimate propensity scores, any standard probability model can be used such as the logit or probit model. A logit distribution function is described by

$$\Pr\left(D_I = 1 \mid \mathrm{X_{Ci}}\right) = \frac{e^{Z=\lambda h(X_{Ci})}}{1+e^{Z=\lambda h(X_{Ci})}} \qquad \text{(P8)}$$

where *D* is the treatment status, *e* is the natural base of the logarithm and $h(X_{Ci})$ consists of linear and higher order terms of covariates. As *Z* ranges from – ∞ to + ∞, *P* ranges from 0 to 1. The logit model is described by

$$Log\,[P/(1-P)] = Z = \lambda h(X_{Ci}); \qquad \text{(P9)}$$

Propensity scores for the two groups are thus estimated using predicted values from logit regression models. The dependent variable in the logit model is the predicted log odds ratio, $p(X_{ci})=P(X_{ci})/(1-P(X_{ci}))$, which is interpreted as the odds in favour of the household having access to water / sanitation facility. Propensity scores thus generated are matched to assess health outcomes in children across the treatment and control groups. Matching takes place by comparing the propensity scores of the control group households with the treatment group households. Matching can be performed for the entire sample or can be classified along different characteristics within the two groups such as by income quartiles, education levels, gender and geographical areas.

5.5 Matching Methods

There are several ways to select control group households to match with the treatment households. Heckman et al (1997) and Heckman et al (1998), describe several methods: nearest neighbour match, kernel-based matching, regression adjusted matching and conditional difference-in-difference matching. We provide a brief description of each of these methods.

5.5.1 Nearest Neighbour Matching

Briefly, nearest neighbour matching selects the match that minimizes the distance $\| P(X_{Cj}) - P(X_{Ci}) \|$ (where *j* and *i* are households that belong to treatments and controls respectively) thereby assigning all the weight to this match. $\| . \|$ is the norm. Here too, there are two possibilities, one in which non-participants can be matched only once (sampling without replacement) (see Pradhan and Rawlings (2002) for an example of single time matching) and the other in which they can be reused (sampling with replacement). Nearest neighbour matching can be extended to the nearest *n* neighbours where the average outcome of the closest *n* neighbours is chosen. Jalan and Ravallion (2003b) and (2003c) use the nearest five neighbours. It is possible that the distance $\| P(X_{Cj}) - P(X_{Ci}) \|$ may in fact be quite large. Hence a tolerance limit can be applied whereby those matches satisfying $\| P(X_{Cj}) - P(X_{Ci}) \| < \varepsilon$ are accepted, where ε is arbitrarily small. This is known as Caliper matching.

5.5.2 Kernel- Based Matching

The second method is the kernel-based matching estimator where weights averaged over health outcomes for all non-participants are formed. Higher weights are placed on those individuals who are closer to participants than those more distant in terms of X_C.[9] Thus, control observations with propensity scores that are closer to scores of the treatment observations are given more weight and those with propensity scores farther away from scores of treatment observations are given lower weights.

[9] Kernel function is a weighting function that assigns higher weights to closer data points (An exposition is available in Greene,W.H. 2000 and Becker and Ichino 2002)

5.5.3 Regression Methods

Regression methods have also been incorporated in matching techniques based on the propensity score method. Outcomes for non-treatment households are described by a linear model comprised of observable variables and unobservables: $H_o = X\beta_o + U_o$. The regression is run only for the matched comparison group and an OLS estimate $\beta_o^{'}$ is obtained to estimate the regression adjusted mean impact in the following way.

$$M(S) = \Delta\mathrm{H} = \sum_{j=1\ldots T} \omega_J \left[\left(h_{j1} - x_j \beta_o^{'} \right) - \sum_{i \in D=0} W_{T,NT(i,j)} \left(h_{jo} - x_i \beta_o^{'} \right) \right] \qquad \text{(P10)}$$

where h_{J1} is health outcome in treatment group and h_{J0} is the health outcome for the control group, *T* are the children in the treatment households and *NT* are the children in the comparison group households, and sample weights are shown by ω_J and $W_{(i,j)}$ is a weight where $\sum_{i \in D=0} W_{T,NT}(i,j) = 1$

5.5.4 Conditional Difference-in-Difference Matching

Conditional difference-in-difference matching is data-intensive as it requires pre- and post-intervention information for the treatment and comparison groups. Conditional on the probability of participation (or having access to a particular water /and sanitation facility), health outcomes of the participants prior to and after the intervention are differenced along with the outcomes of the comparison group over the same.

In this study, we use two methods, the nearest neighbour with replacement and the kernel-based matching to check the robustness of results. We select these two methods as all the treated units are kept and a match is found. We use the Stata pscore programme (Becker and Ichino, 2002) to estimate propensity scores and to conduct nearest neighbour and kernel matching.

Propensity score matching was pioneered by Rosenbaum and Rubin (1983, 1985) and has been subsequently used to measure the impact of a variety of programmes: the impact of labour enhancing programmes in the United States (Dehejia and Wahba (1998)); the impact of social investments in water, sanitation, sewage, primary schools and health posts in Nicaragua on diarrhoea and malnutrition in beneficiary households (Pradhan and Rawlings

(2002)); the impact of piped water on diarrhoea in young children and infants in rural India (Jalan and Ravallion (2003b)); the benefits of an anti-poverty programme through its effect on increased labour incomes (Jalan and Ravallion (2003c)); the impact of infrastructure rehabilitation in Georgia (Lokshin and Yemstov (2003)); gains from infrastructure rehabilitation, (Rao, and Ibáñez (2003)); evaluation of women's role as decision makers, (Chattopadhyay and Duflo (2004)); and the impact of education reform in Mexico (Shapiro and Trevino (2004)).

6. Data

We use data from India's second National Family Health Survey (NFHS), conducted in 1998-1999. The data collection was funded by the United States Agency for International Development (USAID) with additional funds from UNICEF. The data covers more than 90,000 women in the age group 15-49 across 26 states in both rural and urban areas. It also provides regional estimates for select states. Data in Union Territories is not collected. The data spans a large number of questions pertaining to women and their immediate family and environment such as fertility, family planning practices, morbidity in children and their health status, child mortality, maternal health, and the status of women. For the survey, a total of 91,196 households were interviewed with a high response rate (the number of households interviewed per 100 households) of 98 percent.

7. Results and Discussion

We start the analysis by presenting summary statistics for our entire dataset and for our six main categories of water sources considered individually and jointly with sanitation facilities: households with piped water (pipewater); households with well water (wellwater); households with handpump water (handpumpwater); households with piped water and presence of sanitation facility (pipesan); households with well water and presence of sanitation facility (wellsan); and households with handpump water and presence of sanitation facility (pumpsan). Piped water is considered as the best source of water in terms of ready access followed by handpump water which is sourced from under the ground and finally well water.[10] Since many of our variables are dummies, the mean statistics show the fraction of the population where the dummy value is 1. The unit of analysis is at the level of an individual child within the age group of one to three belonging to a household. In our analysis we do not

[10] Inferior sources of water are rainwater, surface water such as from nearby ponds and lakes, and tanker truck which are less desirable due to unreliable quality or intermittent access.

include two children from the same household. Hence, each household corresponds to one child.

7.1 Summary Statistics

In Table 2 we describe the percentage distribution of the sample population by water type and sanitation coverage. Less than half of the population has some form of sanitation irrespective of the water source and the presence of piped water is greater than well water and water sourced from handpumps. The proportion declines when water sources are considered jointly with sanitation with a drastic decline for well water and handpump water.

Table 2. Distribution of Water Type and Sanitation (Percentage of Sample Population)

Dummies	**Sanitation**	**Pipewater**	**Wellwater**	**Handpumpwater**	**Pipesan**	**Wellsan**	**Pumpsan**
Fullsample N=24800	.38 (.48)	.38 (.48)	.19 (.39)	.35 (.47)	.23 (.42)	.04 (.20)	.08 (.27)

Figures in brackets show standard deviation;

In Table 3, we present summary statistics for diarrhoeal incidence by water / sanitation type. Starting with diarrhoea we see that nearly 20% of our entire sample of young children suffered from diarrhoea in the past two weeks.[11] This number is highest for those children with access to piped water among the pure water categories (piped water, well water and handpump water) and is also highest for those belonging to households with both access to piped water and sanitation facilities. We will see later on that this result is fairly robust across our analysis for the various categories of water types and water and sanitation categories taken jointly.

7.1.1 Summary Statistics: Socioeconomic Indicators

We also summarise select socio-economic indicators that describes our sample population across the different water / sanitation categories. We find age of the husband to be higher than those of their spouses and the average size of the household to be close to seven members in all six categories. The years of education for the father of the child is highest in the piped water / piped water with sanitation category which may indicate that increased education brings with it increased wealth and access to an improved water source. Literacy of the mother indicates that the percentage of illiterate mothers was higher for the well water and handpump category of households with this fraction declining for higher education levels

[11] In the questionnaire, the mother of the child was asked a yes or no question on whether her child suffered from diarrhoea in the past two weeks.

within the six categories. Most of the children belonged to households that were headed by a male whose average age was 43 years.

In our analysis we include two awareness measures: 'aware', 'ors-heard of oral rehydration'. The former captures exposure to mass media (by watching television once a week, reading a newspaper once a week or listening to the radio once a week) and the latter describes whether the mother of the child has heard of oral rehydration and serves as a proxy for disease-specific awareness. We find both these measures to be the highest for the piped water category and slightly lower for pipesan category over wellsan. households.

We also present summary statistics for select wealth measures such as the type of house (*pucca* house is made entirely of concrete whereas a *kaccha* house refers to a mud habitation), ownership of agricultural land, presence of electricity, and a standard of living index constructed in the NFHS Survey[12] (International Institute for Population Sciences and ORC Macro, 2000). As expected, we find that a larger fraction of households with piped water were *pucca* houses and had electricity. However, we do not have data on the number of hours of electricity supply so we cannot determine interruptions in water supply for children living in households with piped water which can potentially confound sanitation practices. Across the three levels of the standard of living index for each of the six categories of water and sanitation types, we find that most of the children live in households that belong to the middle category with a comparatively lower distribution of children in piped water households in the low category (18.89% for piped water, 6.38% for pipesan). Ownership of agricultural land and livestock is greater for children living in households with access to well water and handpump water, indicating the importance of agriculture as a livelihood.

It is interesting to note that healthcare responsibilities fall on the mother of the child in piped water households much more frequently than in well water and handpump water households, whereas this figure is higher for well water households with sanitation than for the other two water and sanitation categories. A look at three behavioural inputs for hygiene suggests that children living in households with piped water / pipesan treat their water, such as by straining

[12] Standard of Living Index is computed as a sum of various assets where the ownership of each asset is assigned a value. Hence, for sources of lighting, "2" is assigned for electricity, "1" for kerosene, gas or oil, and "0" for other sources. A low SSLI comprised values ranging from 0-14, a medium SSL had values between 15-24 and a high SSLI had values of 25-67.

through a cloth, more than their counterparts. This trend continues for boiling water whereby piped water households have a higher fraction of households than the other two water categories, but when considering water jointly with sanitation, a larger fraction of households in the wellsan category boil water than do their counterparts.

Table 3: Summary Statistics

Variable	Full	WaterPipe	Waterwell	Waterpump	Pipesan	Wellsan	Pumpsan
Diarrhoea*	.19 (.39)	.20 (.40)	.19 (.39)	.19 (.39)	.19 (.39)	.14 (.35)	.17 (.37)
Age of hh head(yrs)	43.27 (15.07)	43.35 (15.03)	43.51 (15.25)	43.32 (15.11)	43.88 (15.00)	46.47 (15.61)	44.93 15.34
No. of hh Members	7.15 (3.38)	6.90 (3.24)	7.30 (3.46)	7.41 (3.53)	6.92 (3.34)	7.22 (3.50)	7.59 (3.94)
Husband's education(yrs)	6.40 (5.02)	7.83 (5.02)	5.51 (4.69)	5.60 (4.94)	9.34 (4.71)	7.86 (4.39)	8.16 (4.99)
Working mother*	.28 (.45)	.25 (.43)	.35 (.47)	.26 (.44)	.18 (.39)	.16 (.37)	.11 (.31)
Male HH head*	.93 (.24)	.92 (.25)	.93 (.24)	.94 (.22)	.93 (.25)	.87 (.32)	.92 (.25)
Age of mother(yrs)	25.75 (5.44)	25.72 (5.08)	25.70 (5.58)	25.61 (5.60)	26.22 (4.95)	26.33 (5.36)	25.13 (5.17)
Age of husband(yrs)	31.63 (7.34)	31.29 (6.15)	31.91 (8.22)	31.65 (7.86)	31.61 (5.98)	32.90 (6.25)	31.77 (6.90)
Aware*	.56 (.49)	.75 (.42)	.43 (.49)	.44 (.49)	.87 (.33)	.77 (.41)	.74 (.43)
Heard of ORS*	.37 (.48)	.46 (.49)	.33 (.47)	.29 (.45)	.54 (.49)	.57 (.49)	.47 (.49)
Has electricity*	.61 (.48)	.86 (.34)	.47 (.49)	.44 (.49)	.94 (.22)	.69 (.45)	.73 (.44)
Hh owns Agricultural land*	.53 (.49)	.42 (.49)	.63 (.48)	.56 (.49)	.33 (.47)	.47 (.49)	.44 (.49)
Hh owns livestock*	.53 (.49)	.38 (.48)	.64 (.47)	.61 (.48)	.27 (.44)	.44 (.49)	.41 (.49)
Mother decides on obtaining healthcare*	.21 (.41)	.25 (.43)	.20 (.40)	.18 (.39)	.27 (.44)	.31 (.46)	.23 (.42)
Pucca House*	.29 (.45)	.47 (.49)	.19 (.39)	.18 (.39)	.62 (.48)	.46 (.49)	.47 (.49)
Semi-pucca*	.37 (.48)	.32 (.46)	.39 (.48)	.40 (.49)	.26 (.43)	.24 (.42)	.31 (.46)
Kaccha*	.33 (.47)	.19 (.39)	.41 (.49)	.40 (.49)	.10 (.31)	.29 (.45)	.21 (.41)
Water purification: Strain by	.16 (.37)	.23 (.42)	.18 (.39)	.09 (.28)	.22 (.41)	.10 (.30)	.06 (.24)

cloth*							
Boil Water*	.13 (.33)	.19 (.39)	.13 (.34)	.03 (.19)	.27 (.44)	.39 (.49)	.10 (.30)
No water purification*	.66 (.47)	.52 (.49)	.63 (.48)	.84 (.35)	.42 (.49)	.38 (.48)	.75 (.43)
Standard of Living -Low	33.22	18.89	40.94	41.35	6.38	16.38	14.31
Standard of Living – Medium	48.22	50.82	48.62	45.40	49.17	54.64	48.69
Standard of Living -High	18.56	30.29	10.44	13.25	44.44	28.98	37
Mother Education: Illiterate*	.54 (.49)	.38 (.48)	.61 (.48)	.65 (.47)	.24 (.43)	.20 (.40)	.34 (.47)
Literate to Middle School*	.18 (.38)	.17 (.40)	.17 (.38)	.16 (.36)	.20 (.40)	.25 (.43)	.21 (.41)
Completed Middle School*	.10 (.30)	.12 (.33)	.09 (.28)	.07 (.26)	.15 (.36)	.19 (.39)	.14 (.35)
High School and Above*	.16 (.37)	.27 (.44)	.11 (.31)	.10 (.30)	.39 (.48)	.35 (.47)	.28 (.45)
N	24800	9453	4844	8890	5766	1114	2036

*Figures in brackets show standard deviations. Variables with a * next to them refer to dummies.*

7.2 Average Treatment Effects

We now present average treatment effects on the treated (ATT) i.e. the difference in mean outcomes in diarrhoeal incidence in children between treatment and control groups, using the results from the propensity score logit regressions.

$$\text{ATT: } E\,(H_1 \mid p(X_c), D{=}1) - E\,(H_0 \mid p(X_c), D{=}0) = E\,(H_1 - H_0 \mid p(X_c)) \qquad \text{(P11)}$$

The treatment in each case refers to the presence of a particular water type and is a binary variable; for example treatment =1 if household has piped water and presence of sanitation facility, 0 otherwise. The ATT effects capture the environmental effect on child health outlined in the theoretical analysis in section 4.1.1. The steps in the estimation of the propensity score follow those outlined in Becker and Ichino, 2002. We present a brief synopsis of the procedure before analysing the results obtained.

7.2.1 Estimating the Propensity Score

As a first step, the propensity score is estimated using a standard probability model (the logit model in our case) with a binary dependant variable to indicate presence (or absence) of a water / sanitation facility with a number of independent covariates entered linearly or with higher order terms in the regression. The choice of higher order terms is driven only by the need to satisfy the balancing property (ibid).[13] The satisfaction of the balancing property within each stratum of the estimated propensity score implies that the mean of the propensity score is the same across the two groups, and that the distribution of the observable characteristics is the same across the two groups, i.e. they are observationally identical. The covariates chosen in generating the propensity scores are those that are strongly related to the treatment assignment even if they are possibly poor predictors of outcomes. [14]

The average treatment effect on the treated is then estimated by taking the average difference in mean outcomes between the treatment and the control group observations. In our study we restrict the ATT to a region of common support where observations in the treatment groups with scores outside the boundary of the highest and lowest propensity scores of the control group are excluded. This has the advantage of improving the quality of matches (Becker and Ichino, 2002) (See Rao K.D, 2004 for a complete discussion of issues related to common support).

7.2.2 Average Treatment Effects for Water / Sanitation Types

The ATT are shown in Table 4a for the six categories of water and sanitation types using nearest neighbour matching (including with bootstrapped (BS) standard errors) and kernel matching. The results are fairly robust across the two matching methods. We find that the incidence of diarrhoea in children living in households with access to well water, handpump water, well water and sanitation, and handpump water and sanitation is lower than in piped water/piped water with sanitation households and is statistically significant. We find that mean scores are higher for handpump and well water considered jointly with sanitation and show a reduction of diarrhoeal outcomes when sanitation facilities are present. What is

---

[13] Balancing property, also known as the "strong ignorability" or "condition independence assumption", describes $D \perp\!\!\!\perp X \mid p(X)$ where X is a vector of pre-treatment characteristics, p(X) is the mono-dimensional propensity score and D={0,1} refers to exposure to treatment or not.

[14] The propensity scores generated for each water / sanitation category are available on request from the author. Note that in each case, the balancing property was satisfied.

interesting is the higher incidence of diarrhoea in children living in households with access to piped water, which is contrary to findings in the literature (Jalan and Ravallion, 2003b, Dasgupta 2004). Although there is a major decline in magnitude in the mean score in diarrhoeal incidence between piped water households and pipesan households, the incidence is still higher in the treatment households than in the control group, however it is no longer significant in the pipesan category.

Table 4a. Average Treatment Effects for Water / Sanitation Type

Matching Method	**Piped Water**	**Well Water**	**Handpump Water**	**Piped water & Sanitation**	**Well water & Sanitation**	**Handpump water & Sanitation**
Nearest neighbour	0.028 (0.008)* [9453, 4983] #	-0.019 (0.009)** [4844, 3987]	-0.016 (0.008)** [8890, 5712]	0.009 (0.010) [5766, 3399]	-0.050 (0.017)** [1114, 1020]	-0.031 (0.013)** [2036, 1875]
Nearest Neighbour BS	0.028 (0.007)* [9453, 4983]	-0.019 (0.011)*** [4844, 3987]	-0.016 (0.008) [8890, 5712]	0.009 (0.010) [5766, 3399]	-0.050 (0.017)** [1114, 1020]	-0.031 (0.012)** [2036, 1875]
Kernel	0.025 (0.005)* [9453, 15324]	-0.013 (0.007)** [4844 19863]	-0.011 (0.005)** [8890, 15910]	0.009 (0.008) [5766, 18359]	-0.053 (0.011)* [1114, 23375]	-0.029 (0.009)* [2036, 22563]

** Significant at 1% level, ** Significant at 5% level, *** Significant at 10% level; #[Number of treatments, Number of controls], Standard error are shown in parenthesis. First row shows the average treatment effects.*

A quick look at the summary statistics (Table 3) for households with piped water shows that these households on average tend to treat their water more than those that do not have piped water. However, the practice of treating water is not frequent, with only 16% of households in the entire sample reporting a ‘yes’ for water purification using a cloth and only 13% for boiling their water. A look at the awareness variables and literacy variables in Table 3 also seems to suggest that piped water households have a higher awareness and higher education of the mother.

7.3 Average Treatment Effects: A Robustness Check

We try three alternate methods to check the robustness of our results not only for piped water but for all six categories.

Note that for our empirical exercise on income impacts on child health, we use wealth as a proxy indicator for income. It has been argued that wealth is a more accurate measure of economic status rather than income due to under-reporting of the latter. Further, in developing countries, households often earn income from multiple sources which can vary not only by year but across seasons as well. Therefore, all these multiple sources must be accounted for when generating a single measure of income. Also, this variation and uncertainty associated with employment and income earned makes it difficult to rely on any cross-sectional measure of income as representative (Montgomery et al 2000).

An alternate to income is expenditure data which is often used to circumvent the problem of under-reporting. However, expenditure data also relies on memory recall and can thus be subject to recall errors. In their pioneering work on wealth analysis, Filmer and Pritchett (2001)[15] compare the impact of wealth status with a more conventional measure of consumption expenditure on education levels in India. They find that wealth, a proxy for economic status in predicting enrolment rates, performs as at least as well and in some cases even better than expenditure. The advantage of wealth is that it can be viewed as a long run measure of economic status whereas expenditure or income are typically collected for a particular interval during which time there might be fluctuations. Filmer and Pritchett also assert that the wealth index, just as with expenditure (or income), is a measure for something unobserved: the long run economic status of a household. While both measures may be closely correlated, any discrepancies in household classification deriving from the wealth index cannot be considered as a mistake as each measure has empirically and conceptually distinct limitations.

7.3.1 Method 1: Refining Water Placement and Stratification of Average Treatment Effects

We start with a further classification of households with piped water[16] to those that have a private tap and those that source it from a public tap (Table 4b). We also look at the impact of proximity of the water source irrespective of the kind of water used.

[15] This paper was originally published in 1994 as a World Bank Working Paper. Since then many examples incorporating the use of wealth measures for economic status can be found such as Gwatkin et al, 2000 and Rutstein and Johnson, 2004. There are also studies that have created a Standard of Living Measure in demographic analysis. A list of 24 papers, often using DHS data, are cited in Montgomery et al 2000

[16] As a quick check to measure the robustness of our results for piped water we generate propensity scores and estimate average treatment effects by restricting the sample to rural areas only and separately to urban areas only. In both cases we find the piped water results to hold i.e. higher mean outcome for diarrhoeal incidence in children living in households with piped water in rural areas and in urban areas. The results are available on request from the author.

Our results show that our findings on piped water are fairly robust with both private and public tap households showing a higher mean incidence of diarrhoea. However, the results for proximity show that children had a lower mean incidence of diarrhoea in households that were closer to their water source.

Table 4b. Average Treatment Effects for Access to Piped Water from a Private/Public Tap and Proximity of Water Source

Matching Method	**Pipewater Private**	**Pipewater Public**	**Water in Residence**
Nearest neighbour	0.021 (0.009)** [5392,5635][#]	0.015 (0.008)** [4061,18393]	-0.018 (0.007)** [10964, 6446]
Nearest Neighbour BS	0.021 (0.009)** [5392,5635]	0.015 (0.007)** [4061,18393]	-0.018 (0.008)** [10964, 6446]

** Significant at 1% level, ** Significant at 5% level, *** Significant at 10% level; # [Number of treatments, Number of controls], Standard error are shown in parenthesis. First row shows the average treatment effects*

We then look at the impact of select variables on diarrhoeal incidence. After generating propensity scores, we stratify our data according to wealth categories,[17] gender of the child and rural and urban areas. The survey does not collect information on income or on expenditure but does collect information on ownership of assets that are subsequently used to generate a wealth index. Households are then divided into wealth quintiles (low, second, middle, fourth and highest).

The findings from stratification of our data by wealth and other select indicators are shown in Table 4c. Our results show that the incidence of diarrhoea is positive and significant for children living in households with access to piped water for the poorest (0.061) and fourth quintile (0.046) with mean incidence falling in the fourth quintile. As a quick check we also investigate for each category of water type the distribution of households across wealth

[17] We use a wealth index rather than the standard of living index (SSLI) as the latter is more subjective. The wealth index was created by Rutsein and Johnson (based on Filmer and Pritchett's work), and is a pure economic measure of economic status where level of education and type of occupation are excluded. Unlike the SSLI where weights were assigned such as "3" for a motorcycle and "1" for a bicycle, and to some extent are arbitrary, the DHS measure uses the principal component analysis to assign weights. To start with, the variables (or assets) used in the index are standardised (using z scores), then the factor coefficient scores are calculated. The two are then multiplied for each household and summed to arrive at a final index which is a standardised score with mean zero and variance of one. The wealth index thus calculated is then divided into quintiles, and is based on the distribution of the household population rather than on households, The cut-off points are calculated by obtaining a weighted frequency distribution of households (weight = no. of dejure members of the household * sampling weight of the household). Thus each member is given the wealth index score of his household. Finally the scores are ordered and cut off points are marked to obtain the five quintiles (20% sections). (Rutstein and Johnson, 2004)

quintiles (see Table 9 in Appendix A) and do not find any dominance of households belonging to one quintile over another.[18] For households with access to well water and piped water and sanitation, wealth does not play a significant role in influencing diarrhoeal outcomes. In households with handpump water, mean incidence of diarrhoea is negative and significant in the highest wealth quintile (-.034) and is lowest in the second wealth quintile (-.097) for the well water and sanitation category. In households with handpump water and sanitation, there is a switch in sign for the average treatment effects when moving across from the lowest to the highest wealth quintile with a significant reduction in incidence in the middle or third wealth category.

Table 4c. Average Treatment Effects by Wealth, Gender of Child and Location

Variable	Piped Water	Well Water	Handpump Water	Piped water & Sanitation	Well water & Sanitation	Handpump water & Sanitation
Poorest quintile	0.061 (0.022)**	-0.010 (0.023)	-0.006 (0.019)	0.028 (0.028)	0.001 (0.062)	0.002 (0.035)
Second quintile	0.028 (0.019)	0.018 (0.022)	-0.017 (0.019)	-0.005 (0.025)	-0.097 (0.042)**	0.012 (0.032)
Middle quintile	0.013 (0.018)	-0.012 (0.019)	0.010 (0.017)	0.004 (0.023)	-0.004 (0.033)	-0.058 (0.028)**
Fourth quintile	0.046 (0.017)**	-0.025 (0.018)	-0.010 (0.016)	0.027 (0.020)	-0.042 (0.032)	-0.023 (0.027)
Highest quintile	0.035 (0.015)**	0.006 (0.018)	-0.034 (0.016)**	0.018 (0.019)	0.016 (0.032)	-0.036 (0.024)
Female Child	0.029 (0.012)**	-0.008 (0.013)	-0.011 (0.011)	0.003 (0.014)	-0.053 (0.024)**	-0.001 (0.018)
Male Child	0.020 (0.011)***	-0.032 (0.013)**	-0.018 (0.011)***	-0.004 (0.014)	-0.037 (0.023)***	-0.031 (0.018)***
Urban	0.041 (0.015)**	-0.012 (0.028)	-0.029 (0.018)***	0.003 (0.015)	-0.037 (0.034)	-0.038 (0.020)**
Rural	0.023 (0.010)**	-0.022 (0.010)**	-0.012 (0.008)	0.023 (0.015)	-0.060 (0.020)*	-0.033 (0.017)**

** Significant at 1% level, ** Significant at 5% level, *** Significant at 10% level; Standard errors are provided in parenthesis, First rows shows the average treatment effects*

[18] The only exception is the case for the poorest quintile for well water and sanitation where only 9% of the sub-sample falls in this category.

We find that urban households have a lower and significant incidence of diarrhoea for households with access to handpump water with and without sanitation. For rural households, we find a negative and significant incidence of diarrhoea for those with handpump water and sanitation, well water with and without sanitation. Male children report a lower and significant incidence of diarrhoea in households with well water and handpump water considered individually as well as jointly with sanitation. Female children have a lower and significant incidence of diarrhoea only in households with well water and sanitation. For piped water households there is a positive and significant incidence of diarrhoea for both rural and urban areas and for male and female children.

A comparison of the variables used to stratify the data in Table 4c suggests that, except for the poorest quintile, children living with wellsan or pumpsan facilities experience the largest benefit in terms of a reduction in diarrhoeal incidence across the same variables. For example, children in the second and fourth wealth quintile with well water and sanitation facilities have the highest reduction in diarrhoeal incidence mean scores across all water / sanitation types in those wealth categories. Likewise a highest reduction in diarrhoeal incidence is observed for male and female children and in rural areas in the wellsan category than in other water / sanitation types. For children living with pumpsan facilities, reduction in diarrhoeal incidence is highest for the middle and highest quintile and for urban areas than in other water / sanitation categories.

7.3.2 Method 2: Marginal Effects for the Matched Group

Our second alternate approach restricts our sample to the matched households only and runs a probit model using diarrhoea as the dependant variable for each of the six water / sanitation categories and using select independent variables[19] that address the research questions raised in section 2. We then estimate the marginal effects for the independent variables for each category.[20] The results are shown in Tables 5a-5f. Note that the stratification and restriction to matched households method corresponds with our theoretical exercise where we illustrate

[19] The select variables are: whether the mother of the child has a literacy level between literate to middle school, whether the mother has completed middle school, whether the mother has completed high school and above, illiteracy of the mother is the excluded category; child specific variables such as age and gender of the child; rural – urban location of the household; age of the husband; whether husband has completed high school and above; whether household belongs to the second, third, fourth and highest wealth quintile where the poorest wealth quintile is the excluded category; whether the mother has had exposure to media services; and whether the mother of the child has heard of oral rehydration and specific water dummies
[20] Probit regressions for each water / sanitation category are available on request from the author

the impact of environmental variables on child health as well as the effect of socio-economic and child-specific factors conditional on a certain type of water / and sanitation facility (sections 4.1.2, 4.1.3 and 4.1.4 and the corresponding section with alternative specifications).

7.3.2.1 Piped Water / Sanitation

Starting with households having access to piped water (Table 5a) we find that mothers with literate to middle school category (variable litmiddle) of education and high school and above education (which is of the expected sign) do not have a significant impact on the predicted probability of diarrhoeal incidence. We find adverse impacts of middle school education and exposure to mass media which increase the predicted probability by .029 and .034. On the other hand we find a strong impact of reduction in probability of diarrhoea arising from disease-specific awareness captured by the 'ORS' variable (of -.15) which has the largest marginal effect amongst all the independent variables. Both age and high school education of the husband point toward their importance in reducing the probability of diarrhoeal incidence in children. We find male children to have a higher probability of diarrhoeal incidence and a reduction in diarrhoeal incidence with an increase with the age of the child, which is consistent with the literature that emphasises the vulnerability of infants. We find no significant impacts for children living in rural households. Wealth impacts are consistently negative across all quintiles and are largest for the second quintile. We also consider the impacts of piped water and find them to be positive and significant and thus robust and consistent with our earlier ATT findings.

Considering the joint impacts of piped water and sanitation (Table 5d), we find that the results for the piped water category by and large hold except for mother's literacy levels where children with the literate to middle school mothers have a higher predicted probability of diarrhoeal incidence which is now significant. For mothers with middle school and high school and above education, the impacts are not significant. The impact of the pipesan variable indicates that, as with piped water, children living in households with both piped water and sanitation have an increased predicted probability of diarhoeal incidence. (Note that in Table 4a we found positive but insignificant ATT for this category). Wealth impacts are consistently negative and are also of the highest magnitude for children belonging to households in the second quintile. Disease-specific awareness continues to have a significant impact and is the largest amongst all variables.

7.3.2.2 Well Water / Sanitation

Contrary to children living in households with piped water, we find that for well water households (Table 5b) high school education of the mother significantly reduces predicted probability of diarrhoeal incidence by .03. We find no significant impacts of the wealth variables, which is also consistent with our findings in table 4a for this category. Likewise, there are no significant impacts associated with the gender of the child or the location of the household. The disease-specific awareness and husbands' age and education continue to benefit children through reduced predicted probabilities of diarrhoeal incidence. As with piped water, disease-specific awareness dominates marginal effects amongst all explanatory variables in this category.

For children living in households with access to well water and sanitation (Table 5e), we find no significant impact of mothers' or father's literacy level, nor of exposure to mass media. We continue to find strong reduction in predicted probability of diarrhoeal incidence for disease specific awareness (i.e. ORS). As with the results in Table 4a, we find negative and reduced impacts on diarrhoeal incidence for the variable well water taken individually and jointly with sanitation (Table 5b and Table 5e).

7.3.2.3 Handpump Water / Sanitation

As with well water households, the results are similar for children living in households with access to handpump water, except for the highest two echelons of the wealth variable which significantly reduces the predicted probability of diarrhoeal incidence (the result for the highest wealth quintile also holds true when we had stratified by wealth in Table 4c). When considered jointly with sanitation facilities, we find exposure to mass media no longer has a significant impact on predicted probability of diarrhoeal incidence nor do any of the literacy and education variables for the parents of the child. However, disease-specific awareness continues to play an important role in reducing predicted probability of diarrhoeal incidence in young children. The results for the variable water pump and pumpsan that indicate access to handpump water (and with sanitation) are negative and significant and are consistent with the results found with the propensity score matching method in Table 4a. We continue to find the strong impact of disease-specific awareness in reduced probability in diarrhoeal outcomes in children living in households with water sourced from handpumps (including with sanitation) which has the largest marginal effect across all explanatory variables.

Table 5a. Marginal Effects for Piped Water: Matched Group

Probit regression, reporting marginal effects

Number of obs = 14317
LR chi2(15) = 660.87
Prob > chi2 = 0.0000
Log likelihood = -6849.8799
Pseudo R2 = 0.0460

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0135797	.0094033	1.46	0.144	.207725	-.00485	.03201
middlecomp*	.0295014	.0121799	2.50	0.012	.129077	.005629	.053374
highschool*	-.003532	.0113052	-0.31	0.755	.242718	-.02569	.018626
malechild*	.0182391	.0066043	2.76	0.006	.530139	.005295	.031183
childage	-.02372	.0041402	-5.72	0.000	.99085	-.031835	-.015605
rural*	-.0017961	.0074388	-0.24	0.809	.599078	-.016376	.012784
husbhighsch*	-.0225706	.0082105	-2.73	0.006	.396591	-.038663	-.006478
husbandage	-.0017165	.0005188	-3.31	0.001	31.3233	-.002733	-.0007
waterpipe*	.0254967	.0071108	3.54	0.000	.653629	.01156	.039434
wealthsec*	-.0394948	.0103334	-3.64	0.000	.169938	-.059748	-.019242
wealthmidd*	-.0163401	.0105103	-1.53	0.126	.192778	-.03694	.00426
wealthfour*	-.0209327	.010182	-2.01	0.044	.216456	-.04089	-.00098
wealthhigh*	-.0235894	.0098593	-2.34	0.019	.253545	-.042913	-.004266
ors*	-.1550491	.0066953	-22.03	0.000	.444367	-.168172	-.141927
aware*	.0345445	.0080119	4.19	0.000	.725361	.018842	.050247
obs. P	.2008102						
pred. P	.1889854	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Table 5b. Marginal Effects for Well Water: Matched Group

Probit regression, reporting marginal effects

Number of obs = 8793
LR chi2(15) = 253.33
Prob > chi2 = 0.0000
Log likelihood = -4239.0527
Pseudo R2 = 0.0290

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0028682	.0123027	0.23	0.815	.16968	-.021245	.026981
middlecomp*	.0015103	.0172378	0.09	0.930	.08438	-.032275	.035296
highschool*	-.0325695	.0164052	-1.90	0.058	.12055	-.064723	-.000416
malechild*	.0023863	.0084139	0.28	0.777	.523939	-.014105	.018877
childage	-.0232682	.0052713	-4.41	0.000	.986012	-.0336	-.012937
rural*	.0050942	.0129755	0.39	0.696	.855112	-.020337	.030526
husbhighsch*	-.0231663	.0111476	-2.04	0.042	.246446	-.045015	-.001317
husbandage	-.0017448	.0005573	-3.13	0.002	31.8808	-.002837	-.000653
waterwell*	-.0161943	.0086336	-1.88	0.060	.548846	-.033116	.000727
wealthsec*	-.0217752	.0139566	-1.52	0.128	.161947	-.04913	.005579
wealthmid*	-.0128007	.0136359	-0.93	0.354	.19743	-.039527	.013925
wealthfour*	-.0051802	.0132253	-0.39	0.697	.235642	-.031101	.020741
wealthhigh*	-.0177571	.0129714	-1.35	0.178	.240646	-.043181	.007666
ors*	-.1188245	.0086382	-12.58	0.000	.331059	-.135755	-.101894
aware*	.0467463	.0102789	4.58	0.000	.442511	.0266	.066893
obs. P	.1972023						
pred. P	.1894565	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Table 5c. Marginal Effects for Handpump Water: Matched Group

Probit regression, reporting marginal effects

Number of obs = 14530
LR chi2(15) = 305.80
Prob > chi2 = 0.0000
Log likelihood = -7080.3611
Pseudo R2 = 0.0211

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0055605	.0095301	0.59	0.557	.17426	-.013118	.024239
middlecomp*	.0042492	.0137011	0.31	0.755	.081693	-.022604	.031103
highschool*	-.0238019	.0130546	-1.77	0.077	.114591	-.049388	.001785
malechild*	.0090598	.0065775	1.38	0.169	.52691	-.003832	.021951
childage	-.0225683	.0041248	-5.47	0.000	.985272	-.030653	-.014484
rural*	-.007586	.0093808	-0.81	0.415	.819202	-.025972	.0108
husbhighsch*	-.024935	.0086035	-2.84	0.004	.267447	-.041798	-.008072
husbandage	-.0011433	.0004428	-2.58	0.010	31.6448	-.002011	-.000276
waterpump*	-.0148612	.0068845	-2.17	0.030	.60977	-.028355	-.001368
wealthsec*	-.0138962	.0107972	-1.27	0.205	.16510	-.035058	.007266
wealthmid*	.0013398	.0106982	0.13	0.900	.19022	-.019628	.022308
wealthfour*	-.0194888	.0099936	-1.92	0.055	.22470	-.039076	.000098
wealthhigh*	-.0184593	.0098956	-1.84	0.066	.23978	-.037854	.000936
ors*	-.1024977	.006797	-13.95	0.000	.31927	-.11582	-.089176
aware*	.0423142	.0078398	5.42	0.000	.463799	.026948	.05768
obs. P	.1981418						
pred. P	.1926741	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Table 5d. Marginal Effects for Pipewater and Sanitation: Matched Group

Probit regression, reporting marginal effects

Number of obs = 9071
LR chi2(15) = 591.55
Prob > chi2 = 0.0000
Log likelihood = -4129.9514
Pseudo R2 = 0.0668

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0195172	.0119715	1.66	0.096	.213538	-.003946	.042981
middlecomp*	.0191645	.0142832	1.37	0.170	.149046	-.00883	.047159
highschool*	.0104153	.0131894	0.79	0.427	.341638	-.015435	.036266
malechild*	.0223997	.0080606	2.77	0.006	.536986	.006601	.038198
childage	-.0259767	.0050722	-5.11	0.000	.998898	-.035918	-.016035
rural*	-.0047421	.009043	-0.52	0.600	.455848	-.022466	.012982
husbhih*	-.0270386	.0097824	-2.76	0.006	.498401	-.046212	-.007865
husbdage	-.0019193	.0006646	-2.89	0.004	31.5704	-.003222	-.000617
pipesan*	.0254446	.0090433	2.78	0.005	.626833	.00772	.043169
wealthc*	-.0451333	.012259	-3.46	0.001	.171536	-.06916	-.021106
wealthd*	-.0253422	.0125938	-1.95	0.051	.190277	-.050025	-.000659
wealthr*	-.0285148	.012153	-2.27	0.023	.216294	-.052334	-.004695
wealthh*	-.0271404	.0118525	-2.23	0.026	.258516	-.050371	-.00391
ors*	-.1840404	.0084432	-21.33	0.000	.514938	-.200589	-.167492
aware*	.0235807	.0114341	2.00	0.046	.854151	.00117	.045991
obs. P	.1911586						
pred. P	.1741213	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Table 5e. Marginal Effects for Well Water and Sanitation: Matched Group

Probit regression, reporting marginal effects

Number of obs	= 2129
LR chi2(15)	= 126.27
Prob > chi2	= 0.0000
Pseudo R2	= 0.0648

Log likelihood = -910.72169

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0170247	.0235011	0.74	0.460	.220761	-.029037	.063086
middlecomp*	.0047718	.0268682	0.18	0.858	.168154	-.047889	.057433
highschool*	-.0184706	.0266378	-0.68	0.495	.315641	-.07068	.033738
malechild*	.0130046	.0158483	0.82	0.413	.529357	-.018058	.044067
childage	-.0187883	.0100474	-1.87	0.062	1.01315	-.038481	.000904
rural*	.011632	.0184939	0.62	0.534	.712071	-.024615	.047879
husbhighsch*	-.0154463	.0194403	-0.79	0.430	.415218	-.053549	.022656
husbandage	-.0036835	.0011508	-3.19	0.001	32.9018	-.005939	-001428
wellsan*	-.0391681	.0167623	-2.34	0.019	.521372	-.072022	-006315
wealthsec*	-.0303888	.0267063	-1.08	0.278	.168154	-.082732	.021955
wealthmid*	-.0206248	.0268801	-0.75	0.455	.208079	-.073309	.032059
wealthfour*	-.0253436	.0257394	-0.96	0.339	.246595	-.075792	.025105
wealthhigh*	-.0190575	.0259421	-0.72	0.472	.258807	-.069903	.031788
ors*	-.1477047	.0170897	-8.57	0.000	.519962	-.1812	-.114209
aware*	.0271809	.0195192	1.35	0.178	.764209	-.011076	.065438
obs. P	.1709723						
pred. P	.1542402	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

Table 5f. Marginal Effects for Handpump Water and Sanitation: Matched Group

Probit regression, reporting marginal effects

Number of obs	= 3892
LR chi2(15)	= 131.67
Prob > chi2	= 0.0000
Pseudo R2	= 0.0350

Log likelihood = -1812.7259

diarrhoea	dF/dx	Std. Err.	z	P>|z|	x-bar	[95% C.I.	]
litmiddle*	.0056646	.0177964	0.32	0.749	.213258	-.029216	.040545
middlecomp*	.0266448	.0227278	1.21	0.227	.138746	-.017901	.07119
highschool*	-.0086779	.0205698	-0.42	0.675	.266958	-.048994	.031638
malechild*	.0181223	.0123807	1.46	0.144	.526978	-.006144	.042388
childage	-.0289365	.0077431	-3.73	0.000	.991264	-.044113	-.01376
rural*	-.0068493	.0135032	-0.51	0.611	.615622	-.033315	.019616
husbhighsch*	-.0009777	.0151793	-0.06	0.949	.450411	-.030729	.028773
husbdage	-.000763	.000874	-0.87	0.383	31.7857	-.002476	.00095
pumpsan*	-.0298572	.0124587	-2.40	0.016	.520041	-.054276	-.005439
wealthsec*	-.033515	.0195985	-1.64	0.102	.163412	-.071927	.004897
wealthmid*	-.0566169	.0182914	-2.87	0.004	.18705	-.092467	-.020766
wealthfour*	-.0162197	.0190883	-0.84	0.403	.225591	-.053632	.021193
wealthhigh*	-.0506807	.0177249	-2.73	0.006	.269527	-.085421	-.01594
ors*	-.1169737	.0126213	-8.99	0.000	.458376	-.141711	-.092236
aware*	.0015682	.0160257	0.10	0.922	.728931	-.029842	.032978
obs. P	.1875642						
pred. P	.1786234	(at x-bar)					

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| correspond to the test of the underlying coefficient being 0

With the exception of children living in households with piped water and sanitation, we observe a switch in sign from positive to negative when moving from middle school to high school education of the mother for the remaining five categories. In the pipesan category, there is no change in sign but there is a reduction in marginal impact when moving from the lowest to highest literacy levels of mothers.

7.3.3 Method 3: Marginal and Interaction Effects Without Matching

Our third approach is independent of the matching technique. In this case, we employ a probit model on select variables that are good predictors of diarrhoeal incidence by using the 'survey' commands in Stata. We then compute marginal and interaction effects for the six categories of water and sanitation types to check for the robustness of our results[21] found using the propensity score method.

In this regression model we initially used all the explanatory variables employed in Tables 5 and key additional ones and then selected those that were significant to arrive at a reduced number of variables.[22] We then interacted the independent variables with each of the water type dummies.[23] We utilise two separate probit models with the first using the water type dummies (Model 1) and the second where the water type dummies are considered jointly with sanitation facilities (Model 2). The results for the marginal and interaction effects are presented in Tables 6a and 7a.[24]

[21] I thank Professor John Cuddy for suggesting this approach.

[22] One can also also employ the 'stepwise' approach in Stata where insignificant variables are automatically dropped depending on the level of signficance chosen as a basis for elimination. Unfortunately, the 'survey' commands in Stata do not support the stepwise approach (as the degrees of freedom are limited by the number of cluster). Instead it is suggested to group variables together (in our case wealth and education variables) and perform an F-test on the variables. We do such a test and find that the model we finally employ based on the significance of variables taken individually, is not vastly different from the pseudo stepwise approach using groups of variables. We find that the education variables for mothers when taken together are jointly significant but of them only highschool education is significant.

[23] We use the explanatory variables employed in Tables 5 and some key additional variables and select those that are significant. It is also noteworthy to mention that using the significant variables with the pipewater dummy only results in a positive and significant coefficient for this dummy. We also rerun the model with all three water type dummies (piped water, well water and handpump water) and find that sign on the pipewater dummy changes to negative and is not significant. We follow the same procedure for the water and sanitation type dummies and find that when using only piped water and sanitation dummy the coefficient is negative and not significant.

[24] Probit regressions for each water / sanitation category are available on request from the author

7.3.3.1 Marginal Effects by Water / Sanitation Type

Marginal effects in both models show that all six water categories indicate a reduction in the incidence of diarrhoea with the results significant for well water with and without sanitation facilities (first row of Tables 6a and 7a). This forms a departure from the results found using propensity score matching where both well water and handpump water categories including with sanitation had negative and significant results (Table 4a). On the other hand, piped water (including with sanitation) has a negative sign indicating a reduction in the predicted probability of diarrhoeal incidence but is not significant, whereas this result was found to be positive when employing the score and matching technique method. With respect to discerning a trend, we find that across water types only, marginal impacts are the highest (that is the greatest reduction in predicted probability of diarrhoeal incidence) for children living in households with piped water, followed by well water and finally handpump water (first row of Table 6a). For the water categories with sanitation, the impacts are highest for well water, followed by piped water and then handpump water (first row of Table 7a).

7.3.3.2 Interaction Effects by Water Type

A look at interaction effects in Table 6a shows that the proportion of adult women in the household (defined as those falling between the ages 15-49) does not have any impact on diarrhoea for all three water categories. We find that the age of the husband and child-specific factors of height (a proxy for child-specific anthropometric status) are consistently negative and child age is consistently positive for all categories.[25] However, husband's age is only significant for children living in households with piped water. Contrary to the finding in Table 5a, 5b and 5c, we find that vulnerability of children to diarrhoeal incidence increases with age. Mother's high school education, rural location (except for well water), media exposure, and whether the mother of the child works, play no significant role in diarrhoeal incidence. Mother's disease-specific awareness relating to oral rehydration shows that children benefit by a reduction in their predicted probability of diarrhoeal incidence in piped water and well water households with a larger impact in the former.

[25] We considered a child's weight as an additional variable to measure anthropometric status but it was not significant.

Table 6a. Marginal Effects and Interaction Effects for the Three Water Types

	Piped Water	**Well Water**	**Handpump Water**
Marginal effects	-.0639024 (.0609212)	-.0309758 (.0192223)***	-.0170715 (.0188398)
Interaction Effects with water type dummies changing from 0 to 1 – continuous variables			
Propwomen	.0003209 (.0013734)	.0013439 (.0013393)	.0006709 (.0013162)
Husband Age	-.0041993 (.0017888)**	-.0023056 (.0018743)	-.0026476 (.0017905)
Child Height	-.0040206 (.0015656)**	-.0037898 (.0018462)**	-.0037566 (.0015515)**
Child age	.0518185 (.0231288)**	.047417 (.024678)***	.0648828 (.0226779)*
Interaction Effects with water type dummies changing from 0 to 1 – dummy variables			
Mother high school dummy=1	.0094153 (.0530457)	-.0091189 (.0552266)	.0203487 (.0544235)
Rural Dummy=1	-.0106463 (.0196956)	-.0032576 (.0186553)	-.0014194 (.0181586)
Rural Dummy=0 (Urban areas)	-.0431658 (.0425119)	-.0991642 (.0474551)**	-.0540999 (.0421415)
Aware dummy =1	.0085688 (.0209565)	-.0020224 (.0211357)	.006458 (.0200575)
Mother works dummy = 1	-.0252493 (.0269569)	-.0406894 (.0278799)	-.0196569 (.0255708)
Heard of ors dummy=1	-.0720847 (.0289603)**	-.0626995 (.0280833)**	-.0354542 (.0287013)
Second wealth quintile dummy=1	-.0967622 (.0439671)**	-.0833912 (.0432627)**	-.0656179 (.043347)
Third wealth quintile dummy=1	-.0514338 (.0413388)	-.0574338 (.042766)	-.0198196 (.041114)
Fourth wealth quintile dummy=1	-.0440697 (.0379773)	-.0542616 (.0410536)	-.0372376 (.0313433)
Fifth wealth quintile dummy=1	-.0604124 (.0387141)***	-.0626642 (.0420321)	-.0564444 (.0388918)

* significant at1% ** significant at5% *** significant at10% level

Wealth effects are negative for all quintiles and across all categories. Significant wealth impacts are found for children with access to piped water and well water for the second quintiles where the impact in terms of reduction in probability of diarrhoeal incidence is also the highest (-.097 and -.083 respectively). We also find significant and negative impacts for the highest wealth quintile in the piped water category. Among the three water categories of piped, well and handpump water, the interaction effects are the most well-behaved for well water households with respect to expected signs on coefficients for variables on high school education of the mother, husband's age, child height, ors and awareness and wealth.

Examining wealth impacts across the three types of households we find that for the second quintile the impacts are largest for the piped water category (-.096), while the estimated impacts for the third, fourth and fifth quintiles are fairly close to each other for piped water and well water categories.

7.3.3.3 Interaction Effects by Water and Sanitation Type

Interaction effects for the three water and sanitation categories (Table 7a) show that an increase in the proportion of adult women significantly reduces the probability of diarrhoeal incidence in children in pipesan and pumpsan categories, though the effects are very weak. More women in the household are able to provide greater child care, devote more attention and benefit from each other's advice. Likewise, age of the husband also reduces diarrhoeal incidence in children in pipesan and wellsan categories. Age of the child is consistently negative across all three categories but not significant. Child height has a positive but weak effect in households in the pumpsan category in which we also find an adverse impact of mother's high school education on child health. Children in urban areas in the wellsan category have a reduced probability of diarrhoeal incidence. Exposure to media shows consistently negative and significant results, which represents a departure over what was observed for only the water categories (Table 6a) except for well water where it is negative. Disease-specific awareness continues to have strong impacts on children living in pipesan and wellsan households with a larger impact in the latter category and a negative but not significant effect in pumpsan habitations.

Wealth impacts are consistently negative for all quintiles across all the three water and sanitation categories and have a significant impact in pipesan households for the second quintile and for children in pumpsan households belonging to the third and fifth quintiles. Looking at wealth impacts across the three household types we find that for the second quintile the interaction effects decline from pipesan to pumpsan categories (-.039 in pipesan to -.017 in pumpsan households) and a reverse pattern is true for the third wealth quintile (-.044 in pumpsan to -.022 in pipesan households); for the fourth wealth quintile, the largest impact is in the wellsan category (-.043) and the least in the pipesan category (-.005); for the fifth quintile the largest reduction in predicted probability of diarrhoeal incidence is in pumpsan households (-.047) and the least is in wellsan households (-.011)

Table 7a. Marginal Effects and Interaction Effects for the Three Water-Sanitation Types

	Piped Water & Sanitation	**Well Water & Sanitation**	**Handpump Water & Sanitation**
Marginal effects	-.0164715 (.0159272)	-.0393329 (.0203204)***	.0016461 (.012746)
Interaction Effects with water and sanitation type dummies changing from 0 to 1 – continuous variables			
Propwomen	-.0014445 (.0008216)***	.0000303 (.001494)	-.0024147 (.0012838)***
Husband age	-.0036838 (.0013509)**	-.0066409 (.0033036)**	-.0016676 (.0018702)
Child height	-.0007081 (.0008913)	.0011004 (.0015402)	.0024343 (.0014633)***
Child age	-.0020131 (.0139623)	-.005702 (.0283508)	-.0310604 (.0212313)
Interaction Effects with water and sanitation type dummies changing from 0 to 1 – dummy variables			
Mother high school dummy=1	.0132812 (.0230379)	-.0089324 (.0360947)	.0441329 (.0260688)***
Rural Dummy=1	-.0164554 (.0181324)	-.0278429 (.0219105)	.007746 (.0122695)
Rural Dummy=0 (Urban areas)	-.0096176 (.0164737)	-.0617519 (.0392598)***	-.0154802 (.02349)
Aware dummy =1	-.0326992 (.0147281)**	-.0454315 (.0246691)***	-.0303256 (.0140318)**
Mother works dummy = 1	-.0042898 (.0206136)	-.0337026 (.0407884)	.0220098 (.0256143)
Heard of ors dummy=1	-.0562074 (.0154459)*	-.073767 (.0234086)*	-.0083984 (.0178481)
Second wealth quintile dummy=1	-.0395192 (.0246254)***	-.0185743 (.0519151)	-.0176589 (.0358273)
Third wealth quintile dummy=1	-.0225895 (.0246832)	-.0309539 (.0485127)	-.044928 (.0286581)***
Fourth wealth quintile dummy=1	-.005604 (.0233571)	-.0439904 (.0422369)	-.013637 (.0307792)
Fifth wealth quintile dummy=1	-.0215647 (.0213535)	-.0111424 (.0461161)	-.0470854 (.0270713)***

8. Conclusions and Policy Implications

This paper examines the health impacts, using diarrhoea – a water related disease – as the health outcome, in children that live in households distinguished along six main water and sanitation categories. The data used for the analysis stems from a nation-wide survey conducted in India. The paper starts with a theoretical exercise to model health impacts in children using the utility maximization framework. The model is then utilised to make empirical assessments of health impacts in children using the propensity score and matching techniques for non-random and complex survey data. This method allows comparison of outcomes with and without programme participation using post programme intervention (or facility type) data. The main advantage of the propensity score and matching technique is that it reduces bias from heterogeneity across observables. We then enrich our analysis and verify our results using alternate methods. In particular we look at (1) ATT by wealth, gender of the child and a household's location; (2) compute marginal effects by restricting the sample to the matched group only and (3) use an alternate probit model using Stata's survey commands and compute marginal and interaction effects for all six water and sanitation categories. Although having more than one model allows us to compare our results, nevertheless it makes the task of drawing specific conclusions difficult due to a variation in results. In Table 8, we present a summary of our findings from all the models utilised in this paper.[26]

Table 8. Summary of Results from All Methods

Method	**Piped Water**	**Well Water**	**Handpump Water**	**Pipesan**	**Wellsan**	**Pumpsan**
Propensity Score &Matching	Higher incidence of diarrhoea	Lower incidence of diarrhoea	Lower incidence of diarrhoea	Higher incidence of diarrhoea- insignificant	Lower incidence of diarrhoea	Lower incidence of diarrhoea
Propensity Score & Stratification	Higher incidence of diarrhoea for: poorest, fourth and fifth wealth quintile households; male and female children; urban and rural households	Lower incidence of diarrhoea in male children and in rural households; no wealth effects	Lower incidence of diarrhoea in highest wealth quintile, male child and urban households	No significant impacts	Lower incidence of diarrhoea in second wealth quintile, in both male and female children, and in rural households	Lower incidence of diarrhoea in middle wealth quintile, in male children, and in both urban and rural households
Propensity Score & Matched Sample: Marginal Effects	Higher predicted probability of diarrhoea for children: in piped water households, with mothers with up to middle school education; that are male, with	Lower predicted probability of diarrhoea for children: in well water households; with mothers with high school education and above; with	Lower predicted probability of diarrhoea for children: in handpump water households; with mothers with high school education and	Higher predicted probability of diarrhoea for children: in pipesan households; that are male; with mothers exposed to mass media; Lower	Lower predicted probability of diarrhoea for children: in wellsan households; with increase in child age; with increase in father's age; with	Lower predicted probability of diarrhoea: for children: in pumpsan households; with increase in child age; belonging to the third and highest

[26] The results summarized in Table 8 are statistically significant unless other otherwise stated

	mothers exposed to mass media. Lower predicted probability of diarrhoea for children: with increasing age, with mothers with disease-specific awareness, with increasing age of the father, with increasing education of the father; belonging to second, third and highest wealth echelons.	increase in child age; with fathers with high school education; with increase in father's age; with mothers having disease-specific knowledge. Higher predicted probability of diarrhoea for children: with mothers exposed to mass media.	above; with increase in child age; with fathers with high school education; with increase in father's age; with children belonging in the fourth and highest wealth echelons; mothers having disease-specific knowledge. Higher predicted probability of diarrhoea for children: with mothers exposed to mass media.	predicted probability of diarrhoea for children: with increase in child age; with fathers with high school education; with increase in father's age; for children living in progressively wealthier households with the largest impact for the second wealth echelon; with mothers having disease-specific knowledge.	mothers having disease-specific knowledge.	wealth echelon; with mothers having disease-specific knowledge.
Probit Model for survey data: Marginal and Interaction effects when water type dummies change from 0 to 1.	Lower predicted probability of diarrhoea for children: with an increase in their height; with an increase in father's age; with mothers that have disease-specific awareness; belonging to the second and fifth wealth quintile Higher predicted probability of diarrhoea for children: with an increase in their age	Lower predicted probability of diarrhoea for children; in well water households; with an increase in their height; with mothers that have disease-specific awareness; living in urban households; belonging to the second wealth quintile Higher predicted probability of diarrhoea for children: with an increase in their age	Lower predicted probability of diarrhoea for children: with an increase in their height; Higher predicted probability of diarrhoea for children: with an increase in their age	Lower predicted probability of diarrhoea for children: with a greater proportion of adult women; with an increase in father's age; with mothers exposed to mass media and disease-specific awareness; belonging to the second wealth quintile	Lower predicted probability of diarrhoea for children: in well water and sanitation households; with an increase in father's age; in urban areas; with mothers exposed to mass media and disease-specific awareness; living in urban households.	Lower predicted probability of diarrhoea for children: with a greater proportion of adult women, with mothers exposed to mass media; belonging to the third and highest wealth quintile Higher predicted probability of diarrhoea for children: with mothers with high school and above education; with increase child height though the effect is weak

Our main conclusion derives from the importance of disease specific awareness which relates to mothers that have heard of oral rehydration – it shows a strong impact in reducing the (predicted) probability of diarrhoeal incidence in children. Not only is this result consistent across all categories but is the largest in terms of marginal impacts (between 10% and 18% reduction in predicted probability of diarrhoea) when compared with other variables (Tables 5a-5f). This result is also robust when considering the interaction effects of the six water categories with select independent variables (Tables 6a and 7a) where it continues to have strong impacts within a water / sanitation category. This result brings out the importance of disease-related awareness that can go a long way in reducing morbidity in children from diarrhoea. Such an awareness can be achieved through public education campaigns on prevention and remedial action (such as those conducted for vaccination against polio in India, or the Stop TB campaign) through the media in urban settings and on-the-ground campaigns in rural areas. For greater outreach and efficiency in the utilisation of public funds, these campaigns could adopt an integrated approach where more than one disease could be the target of education programmes.

A comparison of health impacts using the propensity score method point toward a greater reduction in diarrhoeal incidence in children when considering sanitation effects jointly with well water and handpump water. While it can be argued that children belonging to such a young age group would not benefit from the existence of sanitation per se. However, it the presence of these facilities that improves overall hygienic practices pursued by their parents and from which children benefit indirectly.

Contrary to the literature we find an increase in diarrhoeal incidence in households with piped water (Table 4a) and piped water and sanitation (no effect in Table 4a and strong effect in Table 5d) using the propensity score method. We test this result by evaluating the marginal and interaction effects for our water categories without employing matching techniques and find the results to be in the correct direction i.e. a negative sign on the coefficient for water dummies in Table 7a but not significant. We do not have data on electricity supply which serves well as a proxy for running water and the lack of which can confound sanitation practices. Nor do we have data on hand-washing practices as unclean hygienic practices can be important in transmitting pathogens. Alberini's (1996) work shows that water supply, specifically its reliability, impact diarrhoeal illness by disrupting defensive behaviour. She finds that water supply sources serving high income households (such as piped water) have the highest interruption rates making these households more susceptible to diarrhoea. Hence, our findings of increased incidence of diarrhoea in the piped water / pipesan category could arise from discontinuous water supply.

Even though households with piped water may have a greater potential to adopt purification, there could be other reasons due to which diarrhoeal incidence could be higher in children belonging to this category of households. First, households that use well water and handpump water as their drinking water source have a confirmed physical source of water whereas piped water households are more vulnerable to interruptions. Second, summary statistics in Table 3 also show that households with both well water and sanitation boil water more than households with piped water and sanitation. Such a phenomenon could be explained by a tacit complacence arising in families with piped water. In India, all piped water is supplied by government utilities and is treated before it is distributed to households. Water is obtained from an untreated source such as from rivers or water stored in dams. It is then treated (such as by purification) and subsequently stored in water storage reservoirs. It is then transported in pipes to households. The good health effects of piped water would thus depend on the quality of this treated water which could have several sources of contamination even *after* it is treated such as from rusting pipes and unclean water storage tanks. Jalan and Ravallion (2003b) note that for rural areas in developing countries it is often the case that piped water is generally safer than other sources but which still does not guarantee a germ-free source and needs to be boiled or purified to ensure quality.

The policy implications from our results point toward educating households that use piped water on potential sources of contamination and disinfection methods that they could adopt to ensure clean drinking water for their children. At the same time, possible contamination points such as water storage tanks and distribution networks need to be maintained to ensure a germ free environment of water. A report on the assessment of water and supply and sanitation in India asserts that transmission and distribution networks are of poor quality, are outdated and poorly maintained (Government of India, 2002b). The results from this paper emphasise the importance of maintaining these networks to reduce contamination and the incidence of water-borne diseases.

We also find that diarrhoeal incidence in children is significantly lower in households that receive water that is delivered or present within the premises of the household. Hence, proximity is an important source of reducing the diarrhoeal burden of disease in children as it allows members of the household to have limited interruptions in water availability on account of distance. Proximity can also have other indirect health impacts such as more time available for mothers to look after their children and more productive time available to them to generate income.

Our results also point toward the importance of husband's education and age where the latter serves as a proxy for experience. Education has a much stronger (marginal) impact on diarrhoeal outcomes in children than age and suggests that fathers should not be excluded from education campaigns. Father's education could serve to complement mother's education especially for women that are poorly educated and rely on their husband's advice for health care.

A comparison of wealth impacts suggests that children living in households that belong to the poorest quintile are, in general, the worst off when compared with higher echelons (Tables 5a-5f). Wealth measures serve as a proxy for choice that allows households to exercise it in terms of the kind of investments that they wish to make to ensure preventative behaviour against disease and to ensure a healthy environment for themselves and for their children. Hence, targeting children living in the poorest group will lead to a reduction in the overall burden of diarrhoeal incidence.

Our study points towards the importance of ensuring surveillance of water supply systems and brings out the importance of awareness and education to combat diseases. At present, India lacks strong water quality monitoring systems. A comparison of water quality reporting standards between the United Kingdom (UK) and India shoes that in the latter, the adequacy of water quality monitoring systems and periodicity of monitoring water quality (both defined as responses to a yes or no question) were used as indicators. On the other hand in the UK, non-compliance of water supply zones with contaminants using seven separate indicators (such as levels of iron, pesticides and nitrates) and the number of times enforcement was considered for breaching micro-bacterial standards facilities. Furthermore, in India there are no indicators for customer service (World Bank, 2006).

There is also an undue emphasis on coverage with little regard to the reliability of the service. The Government of India uses litres per capita per day as a measure for access which does not include the number of hours that water was available per day or per week. In 2001, 74% of the urban population in India had access to piped water; however such systems that served mega cities or towns did not distribute water for more than a few hours a day, irrespective of the quantity of water that was available for distribution (ibid). This implies that households must find coping mechanisms to ensure regular water supply. Wealthy households are more able to do this by finding alternate sources of water (by substituting their water supply from groundwater sources or having better storage facilities) while the poorer households suffer either in terms of time to collect water, or by having inadequate quantities of water.

Awareness and education are also important conduits of reducing morbidity. A study of mothers and health practitioners in 1992 revealed that only 14% of doctors and 6% of mothers used ORS as a first reaction on diagnosing diarrhoea. Further, mothers perceived diarrhoea to be common but not a serious problem (United Nations Children's Fund, 2000). Additional studies show a weak understanding of contamination from exposed faeces and a misplaced judgement that children's faeces are not harmful and are often disposed close to the household. The lack of awareness of ORS was also found to exist amongst mothers *and* health care providers (Rao et al, 1998). The Information, Education and Communication initiative adopted by the Government of India in its Total Sanitation Campaign, is thus a step in the right direction – providing potable water alone will not be enough to combat diarrhoea. The IEC however needs to have some measure to evaluate its effectiveness and outreach which it currently lacks (Government of India, 2002d)

Unless there are focussed efforts to tackle these misplaced notions combined with adequate water and sanitation infrastructure, diarrhoea will continue to remain a major cause of morbidity in children. India has committed to the Millennium Development Goals and reports on the progress made show that for sanitation facilities the percentage of population with access to improved sanitation services[27] was 38% in 2005 (compared to 21% in 1990). For drinking water, the percentage of population with access to improved water sources[28] was 75% in 2005 (over 64 in 1990) (World Bank, 2007). Clearly, sanitation lags behind water supply and calls for concerted action. With respect to drinking water, coverage is not enough – it has to be coverage that is reliable.

---

[27] Improved sanitation services represent the percentage of population with at least adequate excreta disposal facilities that can prevent human, animal and insect contact with excreta.
[28] Improved water source represents the percentage of population with reasonable access to adequate amounts of water.

References

Akin, John. 1992. A Child Health Production Function estimated from Longitudinal Data. Journal of Development Economics. Vol. 38, 323-351.

Alberini, Anna., Gunnar S. Eskeland, Alan Krupnick and Gordon McGranahan. 1996. Determinants of Diarrhoeal Disease in Jakarta. Water Resources Research, Vol. 32, No.7, 2259-2269.

Becker Sascha O and Andrea Ichino. 2002. Estimation of Average Treatment Effects Based on Propensity Scores. The Stata Journal, Vol. 2, no. 4, pp. 358-377

Bozkurt, Ali Ihsan., Servet Özgür, and Birgül Özcirpici. 2003. Association between Household Conditions and Diarrhoeal Diseases among Children in Turkey: A Cohort Study. Pediatrics International. Vol. 45, 443-451.

Chattopadhyay, Raghabendra and Esther Duflo. 2004. Women as Policy Makers: Evidence from a Randomised Policy Experiment in India. Econometrica, Vol. 72, No. 5, 1409-1443.

Cropper, Maureen. 1981. Measuring the Benefits from Reduced Morbidity. The American Economic Review. Vol. 71, No.2, 234-240.

Dehejia, Rajeev H. and Sadek Wahba. 1998. Propensity Score Matching Methods for Non-Experimental Causal Studies. National Bureau of Economic Research, Working Paper 6829, www.nber.org/papers/w6829.

Dasgupta, Purnamita. 2004. Valuing Health Damages from Water Pollution in Urban Delhi, India: A Health Production Function Approach. Environment and Development Economics, Vol. 9, 83-106.

Doraiswamy, P. 1998. Morbidity in Tamil Nadu: Levels, Differentials and Determinants. Economic and Political Weekly, Vol. 33, pp 982-990.

Duraisamy, P. 2001. Health Status and Curative Health Care in Rural India. National Council of Applied Economic Research, Working Paper, No. 78. www.ncaer.org

Esrey, S.A., R.G. Feachem, and J.M. Hughes. 1985. Interventions for the Control of Diarrhoeal Diseases among Young Children: Improving Water Supplies and Excreta Disposal Facilities. Bulletin of the World Health Organisation, Vol. 63 (4), 757-772.

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

Feachem, R.G. 1984. Interventions for the Control of Diarrhoeal Diseases among Young Children: Promotion of Personal and Domestic Hygiene. Bulletin of the World Health Organisation, Vol. 62, No. 3, 467-476.

Fewtrell, Lorna and John M Colford, Jr. 2004. Water, Sanitation and Hygiene: Interventions and Diarrhoea. HNP Working Paper, The World Bank.

Filmer, Deon and Lant H. Pritchett. 2001.Estimating Wealth Effects without Expenditure Data or Tears: An Application to Education Enrollments in States of India. Demography, Vol. 38, No.1 (115-132)

Gerking, Shelby and Linda R. Stanley. 1986. An Economic Analysis of Air Pollution and Health: The Case of St. Louis. The Review of Economics and Statistics, Vol. 68, No.1, 115-121.

Government of India. 2002. National Water Policy.

Government of India. 2002b. India Assessment 2002: Water Supply and Sanitation. Planning Commission of India

Government of India. 2002c. Sectoral Policies and Programmes. Tenth Five Year Plan 2002-2007. Vol. II. Planning Commission of India.
Government of India. 2002d. National Health Policy 2002. Ministry of Family Health and Welfare. www. http://mohfw.nic.in

Government of India. 2003. Guidelines for Implementation of Rural Water Supply Programme. Rajiv Gandhi National Drinking Water Mission, Ministry of Rural Development.

Government of India. 2003b. Guidelines on Swajaldhara: Swajaldhara Gram Hamara. Department of Drinking Water Supply.

Government of India. 2004. Annual Report 2003-2004. Ministry of Rural Development.

Grossman, Michael. 1972. On the concept of Health Capital and the Demand for Health. The Journal of Political Economy, Vol. 80, No. 2, 223-255.

Greene, William H. 2000. Econometric Analysis. Fourth Edition. Prentice Hall International, Inc. USA.

Gwatkin, Davidson R., Shea Rutstein, Kiersten Johnson, Rohini P. Pande and Adam Wagstaff. 2000. Socio-Economic Differences in Health, Nutrition and Population in India. HNP/Poverty Thematic Group of The World Bank, Washington DC.

Harrington, Winston and Paul R. Portney. 1987. Valuing the Benefits of Health and Safety Regulation. Journal of Urban Economics, Vol. 22, 101-112.

Harrington, Winston, Alan J. Krupnick and Walter O. Spofford, Jr. 1989. The Economic Losses of a Waterborne Disease Outbreak. Journal of Urban Economics, Vol. 25, 116-137.

Heckman, James J., Hidehiko Ichimura and Petra E. Todd. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. Review of Economic Studies, Vol. 64, 605-654.

Heckman, James J., Hidehiko Ichimura, Jeffrey Smith and Petra E. Todd. 1998. Characterising Selection Bias Using Experimental Data. Econometrica, Vol. 66, No. 5, 1017-1098.

Jalan, Jyotsna, E. Somanathan and Saraswata Choudhri. 2003a. Awareness and the Demand for Environmental Quality: Drinking Water in Urban India. South Asian Network for Development and Environmental Economics, Working Paper No. 4-03. www.sandee.org

Jalan, Jyotsna and Martin Ravallion. 2003b. Does Piped Water Reduce Diarrhoea for Children in Rural India? Journal of Econometrics, Vol. 112, 153-173.

Jalan, Jyotsna and Martin Ravallion. 2003c. Estimating the Benefit Incidence of an Antipoverty Programme by Propensity Score Matching. Journal of Business and Economics Statistics, Vol. 21, 19-30.

Lokshin, Michael and Ruslan Yemstov. 2003. Evaluating the Impact of Infrastructure Rehabilitation Projects on Household Welfare in Rural Georgia. World Bank Policy Research Working Paper 3155, econ.worldbank.org

Manatsathit, S., H.L. Dupont, M. Farthing, C Kositchaiwat, S Leelakusolvong, B.S. Ramakrishna, A. Sabra, P. Speelman and S. Surangsrirat. 2002. Working Party Report: Guideline for the Management of Acute Diarrhoea in Adults. Journal of Gastroentrology and Hepatology, Vol. 17 (Suppl), S54-S71.

International Institute for Population Sciences (IIS) and ORC Macro. 2000. National Family Health Survey (NFHS-2), 1998-99: India, Mumbai, India.

Montgomery, Mark R., Michel Gragnolati, Kathleen A. Burke and Edmundo Paredes. 2000. Measuring Living Standards with Proxy Variables. Demography Vol. 37, No.2, 155-174.

Pradhan, Menno and Laura B. Rawlings. 2002. The Impact and Targeting of Social Infrastructure Investments: Lessons from the Nicaraguan Social Fund. World Bank Economic Review, Vol. 16, No. 2, 275-195.

Rao, Vijayendra and Ana María Ibáñez. 2003. The Social Impact of Social Funds in Jamaica: A Mixed Methods Analysis of Participation, Targeting and Collective Action in Community Driven Development. World Bank Policy Research Working Paper 2970, econ.worldbank.org

Rao, K.V., Vinod K Mishra and Robert D. Retherford. 1998. Knowledge and Use of Oral Rehydration Therapy for Childhood Diarrhoea in India: Effects of Exposure to Mass Media. National Family Health Survey Subject Report. No.10.

Rao, Krishna D, 2004. Quality Improvement and its Effect on Patient Perceptions and Demand for Health Services: Evidence from Uttar Pradesh, India.. PhD Thesis, John Hopkins University, Baltimore.

Ravallion, Martin. 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation. World Bank Economic Review, Vol. 15, No. 1, 115-140.

Rosenbaum, Paul R and Donald B. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika, Vol. 70, No.1, 41-55.

Rosenbaum, Paul R and Donald B. Rubin. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. The American Statistician, Vol. 39, No.1, 33-38.

Rosenzwieg, Mark R. and Kenneth Wolpin. 1986. Evaluating the Effects of Optimally Distributed Public Programmes: Child Health and Family Planning Interventions. The American Economic Review, Vol. 76, No. 3, 470-482.

Rutstein Shea O and Kiersten Johnson, 2004, The DHS Wealth Index, DHS Comparative Reports No.6, Calverton, Maryland: ORC Macro

Shapiro, Joseph and Jorge Moreno Trevino. 2004. Compensatory Education for Disadvantaged Mexican Students: An Impact Evaluation Using Propensity Score Matching. World Bank Policy Research *Working Paper 3334*, econ.worldbank.org

The World Bank. 2006. India: Water Supply and Sanitation: Bridging the Gap Between Infrastructure and Service. Background Paper: Urban Water Supply and Sanitation. World Bank Report. No 35836, India.

The World Bank. 2007. India: Data and Statistics (Sourced from World Development Indicators 2007) in http://ddp-ext.worldbank.org/ext/ddpreports/ViewSharedReport?REPORT_ID=1336&REQUEST_TYPE=VIEWADVANCED

The World Health Organisation. 2002. The World Health Report. Geneva, Switzerland.

The World Health Organisation. 2004. Water Related Diseases www.who.int/water_sanitation_heatlh/diseases

United Nations Children's Fund. 2000. WATSAN India 2000. Water and Environment Sanitation Section, New Delhi.

Van der Hoek, Wim., Flemming Konradsen, Jeroen H.J.Ensink, Muhammad Mudasser and Peter K Jensen. 2004. Irrigation Water as a Source of Drinking Water: Is Safe Use Possible? Tropical Medicine and International Health. Vol. 6, No. 1, pp 46-54.

Appendix A: Distribution of Households Across Wealth Category by Water Type

Table 9. Distribution of Households Across Wealth Category and Location (%)

Variable	**Piped Water**	**Well Water**	**Handpump Water**	**Piped water & Sanitation**	**Well water & Sanitation**	**Handpump water & Sanitation**
Poorest quintile	16.81	15.81	17.61	15.93	9.22	13.72
Second quintile	17.23	15.35	16.35	16.71	15.53	15.63
Middle quintile	18.95	20.71	18.82	18.97	22.24	19.34
Fourth quintile	21.44	24.60	23.26	21.52	25.45	22.47
Highest quintile	25.57	23.52	23.96	26.87	27.56	28.84
Urban	50.46	8.63	13.45	70.42	22.44	40.88