<u>Do not quote or cite</u>

VALUING HEALTH DAMAGES FROM WATER POLLUTION IN URBAN DELHI, INDIA : A HEALTH PRODUCTION FUNCTION APPROACH

PURNAMITA DASGUPTA

INSTITUTE OF ECONOMIC GROWTH

Introduction[1]

The rapid growth of population and economic activity in a developing economy, increases the demands on the natural system to absorb the fall-outs of enhanced economic activity while reducing the per capita raw material (natural resource) availability. The demands on the ecosystem increase both in terms of its ability to act as a sink and in its ability to regenerate and provide a pool of natural resources. This gives rise to the potential for creating stress on the natural resources and leads one to focus on the concept of scarcity of a natural resource. Economics therefore seeks to put a value on the effects of changes in a natural resource.

Environmental pollution affects human health and well being in several ways. Medical expenses associated with treatment costs of pollution-induced diseases, lost wages, defensive expenditures to prevent the occurrence of pollution-induced illnesses, disutility arising from the illness due to lost opportunities for leisure and, changes in life expectancy due to illness on exposure to pollution are all economically quantifiable aspects of environmental health. Water and disease are closely related. Ensuring adequate supplies of safe water sources is of paramount importance for environment and health considerations.

While there are several dimensions to the development process and its linkages with the ecosystem, rapid urbanisation and its impact has become an all-important aspect of the development process. Rapid urbanisation creates pressures for provision of adequate infrastructural services, such as water supply, sanitation and waste disposal. The present paper addresses the problem of water stress in terms of the valuation of water as a resource in the context of the low-income, infrastructurally disadvantaged urban household, by exploring the links between water quality, water borne disease and the preference patterns of households in urban Delhi.

Among the major Indian metropolises with a population of over one million by 1981, Delhi has experienced the highest demographic growth during the past fifty years. Such

---

[1] This work comprises a part of the author's doctoral thesis submitted to Jawaharlal Nehru University, New Delhi. The author gratefully acknowledges the help received from K. Chopra, R. Dasgupta, B.N. Goldar, M.N. Murty and R.P. Sengupta. The author is grateful for the financial support received for the study from the Environmental Economics Unit at the Institute of Economic Growth.

demographic expansion and economic activity has had obvious implications for the provision of infrastructure, especially for basic services such as water supply. Currently, the consumption of piped water supplies ranges from 313 litres per capita day for the affluent to a mere 16 litres per capita day for the slum dweller (Barah, et al. 1998). While a large proportion of the population is still dependent on public hydrants for their water needs, the inadequacy or non-availability of sufficient quantities of piped water supply forces individuals to resort to alternatives that can be harmful. During summer, the demand for water goes up while the supply of water is reduced due to a fall in the raw water made available from the river (Yamuna) upstream. The poorer households are hit hardest by shortfalls in piped supply and the risks of contamination faced by them are compounded since they can least afford the better means of storage or of purification. During the rainy season, the soluble nature of faecal and other waste matters implies greater and faster dispersion of wastes. As a result the groundwater is contaminated more easily than during the dry months, affecting all those who are dependent on it. Even amongst those using municipal piped supplies, interruptions in supply and improper maintenance of pipes causes frequent exposure to contaminants.

It is in this urban environment that waterborne diseases like cholera and gastroenteritis flourish. At the theoretical level, diarrhoeal diseases are directly attributable to the ingestion of contaminated water or food, contamination being usually due to enteric pathogens, or through faecal-oral contamination. Its causes therefore involve both the household and the public sector as a provider of public goods such as water and sanitation services. Diarrhoeal diseases are endemic in Delhi. Questions of both the adequacy and the quality of the water supply available to the household for drinking purposes have attained crucial importance in this setting.

The present study conducts an objective assessment of the damages incurred by urban households by adopting a health production function approach. The monetisation of health losses borne by households reporting diarrhoeal illness provides a useful measure of the minimum willingness-to-accept of the households for avoiding damages from contaminated water. Such a valuation measure provides a theoretically acceptable as well as a practicable alternative to subjective assessment methods such as those based on contingent valuation exercises. The results obtained on the cost of illness lead in turn to policy implications for the water sector.

Methodology for Evaluating Health Costs

An economic evaluation of the health effects of a change in an environmental or resource good would focus attention on the changes in individual behaviour stemming from the perceived adverse effects (reduction) on the individual's utility. A model for valuing the damages from contaminated water supplies, based on the theory of utility maximizing consumer behaviour was developed for estimating the probability of illness for a household.

Theoretical Construct of the Model

A utility function is defined as

$$U = U(X, L, S) \qquad \text{.... (1)}$$

Where X represents expenditures on all non-health related goods (household's aggregate consumption) and has a price of unity, L is leisure time per period and S is time spent ill (number of days sick). The household derives utility from the consumption of X and L, while S causes disutility. Thus, the first and second order derivatives are of the following nature: $U_X, U_L > 0; U_S < 0$

$$U_{XX}, U_{LL}, USS < 0$$

Time spent ill (S) is modeled as a function of the exposure to contaminants (P) and averting or defensive behaviour (D) to reduce the likelihood of illness. Thus, the health production function can be specified as:

$$S = S(D,P) \qquad \text{....(2)}$$

and $S_D < 0$; $S_P > 0$; $S_{DD}, S_{PP} > 0$.

S is characterized as S = S(Td, P), where Td is time spent on defensive activities (Alberini et al. 1996). In other words, the time spent ill is modeled as a function of the time spent in defensive activities, and the exposure to contamination. The household's budget constraint can then be specified as:

$$I + wT = wL + X + wTd + wS(Td, P) + p_d Td \qquad \text{........(3)}$$

where I is total non-labour income; w is the wage rate; T is total time; p_d is out of pocket expense on defensive behaviour. Assuming that all individuals work for a positive amount of time, implies that $[T-L-Td - S(Td,P)] > 0$.

The household's decision-making problem can then be characterized as follows:

$\max\ \Delta = U(X,L,S(Td,P)) + \lambda\,[I + w(T - L - Td - S(Td,P)) - X - p_d Td]$ ....(4)

With T being the total time available and λ being the Lagrange multiplier for the income constraint, the first order conditions are derived as:

$\Delta_X = U_X - \lambda = 0$ ......(5)

$\Delta_L = U_L - w\lambda - \mu = 0$ ......(6)

$\Delta Td = U_S S_{Td} - w\lambda S_{Td} - w\lambda - \lambda p_d - \mu S_{Td} - \mu = 0$ .....(7)

$\Delta_\lambda = I + wT - wL - wS(Td,P) - X - wTd - p_d Td$ .....(8)

Therefore equations 5 and 6 represent the trade-off between labour and leisure while equation 7 yields;

$U_S S_{Td}/\lambda - wS_{Td} - p_d = w$ ......(9)

This optimality condition can be interpreted as follows:

$\lambda = U_X$ (from equation 1);

therefore condition (9) can be written as:

$UsS_{Td}/Ux - wS_{Td} - p_d = w$

The first term above, $U_S S_{Td}/U_X$ gives the marginal rate of substitution between dU/dTd and dU/dX i.e. it gives the (implied) gain in utility (in terms of X), from a unit increase in Td. The second term, $-wS_{Td}$ is positive, (assuming that w >0), since S_{Td} is negative by formulation. As time spent on defensive activities goes up sick time would reduce). Thus, $-wS_{Td}$ gives the gain due to reduced sick time, valued at the wage rate. Therefore, $(U_S S_{Td}/U_X - wS_{Td})$ gives the gross gain from an increase in Td. The third term, $-p_d$ implies the expense incurred for defensive activities. Therefore, the right hand side of the optimality condition gives the net gain from a unit increase in Td.

At the margin, the wage loss (w) corresponding to the unit increase in time spent on defensive activities, must equal the net gains (in terms of non-health consumption expenditure and reduced sick time) from the increase in time spent on defensive activities.

Methodology for Econometric Estimation

The utility maximization exercise leads to a set of first order conditions which would solve for an optimal Td* which would in turn be a function of (w, p_d, I, P).

$$Td^* = Td^*(w, p_d, I, P) \qquad \text{...(E1)}$$

On substituting Td* in S, we get the actual time spent ill as:

$$S = S(Td^*, P) \qquad \text{...(E2)}$$

The model would therefore estimate (E1 and E2) using household level data to study how individuals respond to a threat of contamination and illness.

The estimation procedure involves the estimation of a reduced form relationship between illness and the ambient pollution levels while controlling for other variables that affect health status. Such a health production function can be estimated using cross-section data on illness and defensive behaviour as per model specifications. In actual estimations of the health production function, one would also be controlling for other determinants of health status such as physical and socio-economic characteristics of age, sex, income and education. Such a model which allows for the joint determination of health and behaviour captures the effect of behaviour on health and is thereby more meaningful in analyzing real world situations.

(E1 and E2) can be estimated using household level data. One possible specification for both these equations is based on binary observed dependent variables. Suppose that households engage in defensive behaviour if the value taken by a random variable y_1^* is greater than zero. y_1^* would be determined by individual/household characteristics (including the wage rate, non-labour income, costs of defensive behaviour) and some risk factors known to the researcher such as the count of coliforms in the water, or other such proxies for contamination such as the mode of disposal of liquid wastes. These observable variables are summarized into a vector of regressors (x_1). Further, in estimating any such model, one has to allow for risk factors that are known to the household in determining defensive behaviour but remain unaccounted for among the variables elicited by the survey. Such factors may include, for example, fluctuations in the quality of water that are not captured in a one time testing exercise, previous experience of illness, etc. Summarizing these factors into a variable R*, the reduced form equation for defensive behaviour can be written as:

$$y_1^* = x_1\boldsymbol{b}_1 + \boldsymbol{g}_1 R^* + \boldsymbol{e} \qquad \text{...(E3)}$$

where e is a random error term. It is assumed that the coefficient g_1 is positive i.e. a higher value of R*, is associated with higher risk of diarrhoeal diseases and hence a higher level of defensive activities. Equation (E3) can be estimated using binary dependent variable techniques.

For the second reduced form equation relating illness with its associated determinants, we assume a binary specification is assumed where a random variable y_2^*, defined as

$y_2^* = x_2 b_2 + g_2 R^* + d y_1^* + h$ ......(E4)

takes on a value greater than zero if diarrhoeal illness is observed in a household. Here x_2 is also a set of individual characteristics and sources of risk (contamination of drinking water) for diarrhoeal disease that are available to the researcher. It is to be noted that there are no *a priori* restrictions imposed on x_1 and x_2 and these could comprise of certain common or distinct regressors. Again, unobservable risk factors R* are included among the determinants of illness with g_2 being positive implying that a higher risk of contamination leads to a higher likelihood of contracting the disease. Diarrhoea is controlled by defensive behaviour y_1^*, so that the coefficient d should be negative. Equation (E4) is also estimated using binary response techniques. The error terms e and h are assumed to be independent of each other. Since the risk factors are unobservable, these will be absorbed into the error terms $v_1 = g_1 R^* + e$ and $v_2 = g_2 R^* + h$.

A probit regression of observed defensive behaviour on the selected regressors yields consistent estimates, provided it is assumed that x_1 is independent of the error v_1. However, a probit regression of diarrhoeal illness on individual characteristics and defensive behaviour would yield inconsistent estimates because the hidden risk factors would introduce a correlation between defensive behaviour (one of the regressors here) and the error term v_2 in the illness equation. Therefore, the correct procedure to be adopted is as follows. It is to be noted that E3 is already expressed in reduced form as it contains only exogenous regressors. On substituting E3 into E4, a second reduced form equation is obtained where defensive behaviour is eliminated from the regressors and diarrhoeal disease depends only on individual or household characteristics and unobservable risk:

$y_2^* = x_2 b_2 + x_1(d b_1) + [(d g_1 + g_2) R^* + (d e + h)]$ .....(E5)

The error term of E5, (in brackets), is correlated with the error term of the first equation, $v_1 = g_1R^* + e$. The covariance between the error terms of the reduced form equations, E3 and E5, is equal to $(dg_1 + g_2)\ g_1V(R^*) + ds_2e$, and is in general nonzero, implying that the probability of becoming ill is not independent of that of engaging in defensive behaviour. The quantity $(dg_1 + g_2)$ in particular can be interpreted as the net effect on illness of a change in the unobservable risk i.e. after the individual's defensive actions. As a consequence of the non-independence of the probabilities of defensive behaviour and illness, we can estimate E3 and E5 jointly as a (simultaneous equation) bivariate probit model. Such an estimation would imply the assumption that v_1 and v_2 are (jointly) normally distributed, and tends to behave reasonably well provided that the true distribution belongs to the exponential family of distributions. If that is the case then the estimates for the coefficients would be consistent and "robust" standard errors can be produced from the information matrix (I) and the matrix of cross products of the first derivatives of the log likelihood function (F), (i.e. from the matrix: $I^{-1}F\,I^{-1}$).

It is also to be noted that the parameters cannot all be identified separately. As with standard probit regressions the bivariate probit estimation exercise would lead to estimates of the ratios $b_1=b_1/s_1$ and $b_2=b_2/s_2$, where s_1 and s_2 denote the standard deviations of the reduced form error terms. Here s_1 and s_2 cannot be identified, nor can the two g and d. To sum up, the econometric model has two equations of the form:

Model: $z_{i1} = \beta_1' x_{i1} + \varepsilon_{i1}$, $y_{i1} = 1$ if $z_{i1} > 0$

$y_{i1} = 0$, otherwise

$z_{i2} = \beta_2' x_{i2} + \varepsilon_{i2}$, $y_{i2} = 1$ if $z_{i2} > 0$

$y_{i2} = 0$, otherwise

$(\varepsilon_{i1}, \varepsilon_{i2}) \sim$ bivariate normal $[\ 0,0,1,1,\rho\ ]$

The parameters for this model would be estimated using a complete sample on (y_1, y_2, x_1, x_2).

Sampling Frame for the Study

Cholera as an indicator of the diarrhoeal disease burden in a community

At the outset the present study recognizes that the complexity of transmission routes of the diarrhoea-causing pathogen has to be taken into account in assessing the impact of any single intervention to control the disease (Briscoe 1984). Nevertheless, the case for interventions focusing on improvements in drinking water has withstood the test of time and empirical analysis (Esrey et al 1985, Esrey et al 1991, Anderson and Cavendish 1992). Diarrhoea is defined as the passage of loose, liquid or watery stools. By definition, these liquid stools are passed more than 3 times a day. Diarrhoea lasting for more than 3 weeks is classified as chronic diarrhoea. The WHO/UNICEF (WHO 1985) have also defined "acute diarrhoea" as an attack of sudden onset, which usually lasts 3 to 7 days, but may last up to 10-14 days. While a wide assortment of organisms cause acute diarrhoea, the occurrence of cholera signifies serious levels of contamination of water as the infective dose to cause clinical disease by V*ibrio cholerae* is highest among all other pathogens causing diarrhoeal diseases. In view of this, the sampling procedure for the field survey adopted in the present study, was based on data on cholera cases occurring in Delhi. The data on cholera cases served as a reliable indicator since as a routine, microbiological diagnosis of diarrhoea cases is not done in Delhi except for cholera and, the cholera cases are thus the best documented among the diarrhoea cases.

Sampling technique – selection of households

The sampling technique was based on a two-stage sampling procedure. At the first stage, keeping in mind the epidemiological considerations for such a study, all the 14 localities in Delhi that had been reporting more than 5 cases of cholera annually, consecutively over the three years, 1996 to 1998 were selected. The occurrence of more than 5 cases of cholera in a locality is taken as a standard benchmark for determining the vulnerability of an area to waterborne diseases by epidemiologists at the municipal health department. The survey was conducted during the summer months of 1999.

At the second stage, 600 households were to be selected from these 14 localities. Based on the data on number of cases for these three years, the three-year period prevalence was first calculated for these 14 localities. Period prevalence is defined as the frequency of all current cases (old and new) existing during a defined period of time expressed in relation to a defined population. In terms of a mathematical formula it can be stated as:

*period prevalence rate = [(number of existing cases (old and new) of a specified disease during a given period of time interval) / (estimated mid-interval population at risk)] * 100 .*

Thus, the sum of the cases over these three years was divided by the mid-interval population to arrive at the period prevalence. Each period prevalence was converted into a proportionate prevalence by dividing it by the sum of the prevalence ratios. A set of weights was generated which took into account the disease load for each locality. The total number of households in each of these 14 localities was subsequently weighted by the corresponding weights. A simple proportional rule was thereafter applied to these weighted household numbers in order to distribute the sample requirement of 600 households across the 14 selected localities. On further investigation, these 14 localities were found to be spread across 23 specific colonies. In all such cases where a locality was spread out over one or more colonies, the designated number of households to be sampled from each locality were divided among the constituent colonies in the same ratio as that of the total number of households in the component colonies. The actual selection of households within each colony was done by random sampling.

Technique for laboratory testing of water samples

An important part of the field study involved the testing of water samples from the surveyed colonies. Water samples were collected from each colony and laboratory testing of these water samples was conducted at a government-accredited laboratory. All the samples were tested for faecal coliform counts. Among the available bacteriological tests for water quality surveillance the presumptive coliform test was the most relevant one for purposes of the present study. This test is based on the most probable number (MPN) of coliform organisms in 100 ml of water. The test is carried out by inoculating measured quantities of the sample water – 0.1, 1.0, 10 and 50ml into tubes of McConkey’s lactose bile salt broth with bromcresol purple as indicator. The tubes are incubated for 48 hours. From the number of tubes showing acid and gas, an estimate of the mpn of coliform organisms in 100 ml of the sample can be obtained from standard tables. This result is known as the “presumptive coliform count”, the presumption being each tube showing fermentation, contains coliform organisms. *Based on the guidelines for drinking water quality recommended by the WHO in 1996, the ISO (IS 10500) prescribes that the coliform count of a drinking water sample should be less than or equal to 10 per 100 ml of water.* The WHO recommendation for drinking water quality lays down the primary standards as prescribed limits that must never be exceeded in water meant for drinking purposes. The “coliform organisms” include all aerobic and facultative anaerobic gram-negative, non-sporing, motile and non-motile rods capable of fermenting lactose at 35^o - 37^o c in less than 48 hours. The coliform group includes both faecal and non-faecal organisms. As a working rule it is assumed that all coliform organisms are of faecal origin unless non-faecal origin can be proved. There are several reasons why coliform organisms are chosen as indicators of faecal contamination. An average person excretes an average of 200-400 billion coliforms (which are present in great abundance in the intestine) daily. These organisms are foreign to potable water and hence their presence is a reasonably sure indication of faecal contamination. Further, they are easily detected by culture methods. The coliform bacilli survive in the water much longer than the pathogenic organisms and most importantly the coliform bacilli have greater resistance to the forces of natural purification than the water-borne pathogens. Thus, the presence of other intestinal pathogens is a valid assumption in a water sample in which coliforms have been detected.

Selection of water samples

For purposes of sampling a total of 90 water samples were collected i.e. 15% of the 603 households that were surveyed. This would be sufficient to ensure that epidemiologically sensitive results were obtained for deriving conclusions about the exposure to contamination through the water that was being accessed by these 600 households. It is important to note that the term water source is defined in this study with reference to the means by which water is being accessed. The means of access play the most important role in influencing as well as indicating the quality of water being consumed by a particular household. Means of access that have similar characteristics and implications for water contamination, are clubbed together and labeled as a particular source. Thus, for example, shallow handpumps would be one source, as distinguished from deepbore tubewells which would be considered to be another distinct source. In all there were nine such sources defined in the present study. The distribution of the 90 water samples was done in a manner ensuring that from each colony, at least one of each type of water source with similar characteristics was tested. For colonies where community level water sources (e.g. public hydrants, community handpumps) were being accessed, all the specific points that were being accessed by the surveyed households were sampled. For those with individual water sources, be it handpumps, tubewells or illegal connections, 15% of the total number of households surveyed were randomly selected and their water samples tested in each colony. For colonies with individual piped connections, while the same 15% rule applied, in addition water samples were also taken from the entry points of the water mains to the colony.

Survey instrument

The primary data for the analysis was gathered through a survey conducted as a personal interview between the surveyor and the head of the household. The survey questionnaire had 13 modules covering various aspects of socio-economic and demographic features of the household, apart from detailed sections relating to water and illness aspects.

Preliminary Data Analysis

Table 1 gives the distribution of the primary source of drinking water by income quintiles. While most of the water sources mentioned in these tables are self-explanatory, code 9 deserves special mention. This code is used in all those cases where there is a private arrangement for buying water from a neighbour's deepbore tubewell. The piped sources constitute the bulk of the water sources being accessed for drinking purposes. Predictably, the reliance on individual piped supplies increases while that on groundwater options decreases as one moves up the income brackets. It would be interesting to see how this data compares with the data on the number of cases being reported by households. Examination of the data reveals that out of 264 households that reported cases, 163 households used piped water supplies (either individual connections or public hydrants) from the municipality. It becomes clear that there is a need for analyzing issues relating to both groundwater as well as piped water.

The data on illness can also be classified by income quintiles, identical to the ones used for classifying water sources. Table 2 reveals that there is no straightforward relationship between households reporting cases and the income quintiles to which these belong. Thus although there is some pattern in the water sources being used by different income groups, this income factor does not get reflected on the incidence of diarrhoeal diseases in the same manner.

It is also of interest to study the data on frequency of supply interruptions for those with piped water supply sources. Out of 603 households surveyed, the majority (355 households) is dependent on piped water. Interruptions in piped water supply have been identified as a crucial variable in explaining diarrhoeal illness in earlier studies (Alberini et al. 1996). The data reveals that about 70% of households using either individual piped connections or public hydrants are subject to interruptions in the water supply.

The data on defensive behaviour reveals that only 97 households adopted purification measures for their drinking water. While 26 households boil water, 35 use filtration techniques and 36 use chlorine tablets. It becomes evident that purification measures are being primarily adopted by those using piped water supplies.

Results from laboratory analysis of water samples

Table 3 reveals that out of the 90 samples tested for coliform counts, 53% had coliform counts that were unacceptable for safe drinking water. While the proportion of samples failing to meet the required coliform standards for drinking water varies across the different sources of water, an overwhelming 93% of water samples taken from shallow handpumps failed the test. All the samples taken from deep bore tubewells also failed the test. In the case of legal piped water supply, for both individual connections and public hydrants, the picture is reversed with failure rates ranging from 31.6 to 37.5%. The evidence from the raw data therefore, points overwhelmingly in favour of piped water supplies as being less contaminated than ground water supplies. For most of the colonies, especially the slum clusters, public hydrants turn out to be a better option than handpumps or tubewells, which is of course an expected outcome. What is of more concern however, is that while groundwater has come to be accepted as a potentially risky source with regard to contamination, several public hydrants too contain coliform counts that are above the acceptable limits. The implications for diarrhoeal and other waterborne illnesses for areas where large numbers are dependent on public hydrants is very serious.

Incidence of illness

It is to be recalled that the definition of diarrhoeal illness includes a reporting of at least three loose stools per day, along with any other symptoms of gastro-intestinal discomfort such as vomiting, nausea, etc. for adults. In the case of children, the mothers reporting of illness is accepted. The incidence rates for diarrhoeal illness among the population surveyed, both cluster-wise and overall, for adults, male and female children were calculated. All these incidence rates are very high as shown in Table 4. However, it must be understood that the present study covered the most vulnerable colonies during the most vulnerable season from the point of view of diarrhoeal illnesses. Hence, it is more pertinent to compare these results with studies that were conducted under similar circumstances (e.g. Bhatnagar and Dosajh 1986, Bhandari 1992, Mehrotra 1988). In the present study, 19.95% of the children below 5 years had suffered from one or more episodes of diarrhoea over the previous two weeks at the time of the survey. This is directly comparable to the 24.6% figure obtained by the Bhatnagar and Dosajh study (1986) on diarrhoeal disease morbidity in children below 5 years in urban slums in Delhi. The average duration of illness was 3.9 days per episode for the latter study while, in the present study the mean duration was 3.2 days per episode. Bhandari's study (1992) of a slum in Delhi also provided evidence of high disease burdens in infancy with overall incidence rates that soared during the summer and monsoon months. Mehrotra's (1988) incidence rates of 79/1000 for Sundernagri (a resettlement colony) for the months July to August are directly comparable to the results of the present study. In fact it is worth reiterating that incidence rates 12 years later are comparable to those obtained in 1988, which was an epidemic year.

Results from the Econometric Estimation

Table 5 presents the summary statistics on the regressors used in the estimation. The dependent variable for equation 1 is diarrhoeal illness – it takes a value of 1 if the household reports at least one case of diarrhoeal illness, otherwise 0. For equation 2, the dependent variable is purification of water - it takes a value of 1 in all those instances where the household reports that it undertakes measures to purify water before drinking the same, otherwise it takes a value of 0.

For purposes of estimation, the present study clubs the different sources into two broad categories, piped (legal) and non-piped sources. This classification leads to a crucial variable for the model. Within the piped category, both individual connections and public hydrants are included. It may be recalled that while 72% of the laboratory tests conducted on water samples from the non-piped category failed, in the piped category 34% of the tested water samples failed to meet the required coliform count.

The monthly per capita income has been used in logarithmic form in the estimation, and is self-explanatory. The educational level of the household is a categorical variable, which has been reclassified as five dummies, each representing a certain level of educational attainment. Dummy 1(the omitted category) takes a value of 1 if the person is illiterate, dummy 2 takes a value of 1 if the person is literate without a formal education, dummy 3 takes a value of 1 if the person has completed his primary education, dummy 4 is for those who have completed their secondary education, and dummy 5 denotes a higher degree than dummy 4. Timings seeks to capture the effects of intermittent supply of drinking water. This dummy has been defined interactively for all those who have a piped supply as a source of drinking water. The storage variable seeks to capture contamination of drinking water due to faulty storage practices. Thus, it acts as a control for the estimation exercise, seeking to differentiate source contamination from secondary contamination, which could lead to the occurrence of diarrhoeal illness.[2]

The estimation also includes several important indicators of both individual and neighbourhood sanitary conditions and the risk to illness resulting there from. Whether or not the household has access to a (sanitary) latrine, its mode of disposal of solid wastes and, the presence (or absence) of garbage dumps in the immediate neighbourhood are all perceived to be risk factors for diarrhoeal illness, arising out of an unsanitary environment. The absence of sewers and a foul smell in the water are two indicators of a household’s perception of risk that are used in the purification equation.

---

[2] It maybe noted that washing of hands after defecation, or before the preparation and serving/eating of food has traditionally been a standard control while analyzing diarrhoeal illness. However, our primary data located 100% positive response to the question concerning this aspect of personal/family hygiene. Consequently this variable could not be used to throw further light on the determinants of diarrhoeal illness.

Finally, the households analyzed in this study can be grouped in clusters according to their geographical location. This is done by using five location specific dummies. Briefly, these are North-North Delhi, North Central Delhi, North West Delhi, East Delhi and South East Delhi.

Table 6 gives the detailed results from the specification that was finally chosen for further analysis. The results are discussed in brief below. To start with the model specification has a good explanatory power as revealed by the Wald chi2 statistic. It is to be noted at the outset, that this estimation is a probit exercise and hence the variable coefficients cannot be interpreted directly in terms of their magnitudes. However, the signs and significance levels do have a story to tell. The likelihood ratio test of $\rho=0$, shows that the null hypothesis is rejected, implying that the specification is appropriate; and it is appropriate to model these equations jointly as a seemingly unrelated bivariate probit.

In equation 1, diarrhoeal illness has been regressed on a set of explanatory variables. The source of drinking water is significant and positively related to diarrhoeal illness indicating that households who are dependent on non-piped sources have a higher probability of reporting diarrhoeal illness. It can therefore be inferred from this result that those consuming drinking water from non-piped sources such as, shallow handpumps and Mark II handpumps, are most vulnerable to suffer from diarrhoeal illness. This is in keeping with the hypothesis underlining the need to provide *piped* water supplies to these infrastructurally deficient areas. The variable is however not a significant determinant of the decision to undertake defensive activity.

Certain other risk factors theoretically important for explaining diarrhoeal disease are found to play a much less significant role. Thus not having access to sanitary latrines does not emerge as a significant explanatory variable for diarrhoeal illness. On the other hand, educational attainment of the head of the household also does not play a role among the determinants of diarrhoeal illness in the estimation. Among other factors, irregular removal of garbage from dumps within colonies and non-availability of municipal disposal facilities for household solid waste, play a significant role in explaining the incidence of illness. The results also confirm the hypothesis that interruptions in water supply are a significant determinant of diarrhoeal illness relative to households who do not experience such interruptions in piped supply.

Per capita income is significant as a determinant of illness. Expectedly and in keeping with the findings of earlier studies (Bhandari 1992, Panda 1996), it is found to be negatively related to the probability of illness. It possibly captures the effects of related socio-economic variables, which have not been quantified otherwise. Better-off households are in a better position with regard to buying information, and have a greater capacity to adopt alternative choices that affect diarrhoea, e.g. better hygiene, better living conditions, etc. The storage variable is a significantly positive determinant of diarrhoeal illness. The immediate interpretation seems to be that households who store water are at an increased risk of illness as compared to those who do not store drinking water. This finding is of importance in the related context of the debate on quantity versus quality of water. Among the location dummies, it is found that households in North–North Delhi and those situated in East Delhi (Shahdara North) are more prone to diarrhoeal illness geographically as compared to the households located in South East Delhi.

As far as equation 2, the adoption of defensive practices is concerned, both per capita income and the educational level of the head of the household play a significant and positive role in determining defensive behaviour. Not having access to a latrine and the presence of a foul smell in the water act as significant proxy risk factors for possible contamination of the water and induce defensive behaviour. It is more difficult to explain the result with regard to sewers. The results indicate that not having a sewer connection has a negative impact on defensive practices. What is more probable is that this variable captures certain unquantified aspects of the socio-economic levels of the households. This in turn leads to a confounding effect on the sewer variable. The timings variable is also significantly and positively related to defensive behaviour. This is an important finding since it seems to indicate that among households with piped supplies, those who experience interruption in the supply are more likely to undertake defensive measures to purify water.

The location dummies indicate that households located in North Central and North West Delhi are more likely to adopt defensive practices as compared to those in South East Delhi. These results are in keeping with the current understanding of the theory of diarrhoeal diseases and the associated risk factors, which could lead to defensive activity.[3]

Monetisation of the Health Damages

The model estimated above can now be used for deriving estimates for the predicted probability of observing illness in each household which has been surveyed in the study. The univariate (marginal) predicted probability of success in the outcome, which in this case is defined as the probability of observing diarrhoeal illness in a household, is estimated for the sample. The average value obtained for this predicted probability is 0.44.

In this section, this probability value is used along with other statistical measures to arrive at the monetised cost of illness for the entire population under discussion. This cost of illness is an objective assessment of the damage caused by water contamination. The damage value of contaminated water is monetised in order to obtain the total monetised value of the opportunity cost of illness. These costs can alternatively be interpreted as an indicator of the (minimum) willingness to accept compensation for the surveyed population. The cost of illness comprises of two components, the costs of treatment and the wage-loss arising from absence from work due to ill health.

The cost of treatment

For arriving at a representative household's treatment costs, the data on the individuals sampled in this survey was first divided into three age categories. Children were defined as the age group below 15, adults as age 15 to 60, and those above 60 were classified as the elderly. Such an age classification enables a look at each of these age groups separately, in terms of the probability of falling ill and the average treatment expenses

---

[3] In an effort to disentangle the effects of income from water sources, alternative specifications dropping the water source dummy were experimented with. However, the results obtained do not differ significantly.

incurred by each age group. Further, it serves to identify the adult and employable population.

The probability of an individual being ill, in a representative household can be calculated for each age category. If the probability of a household being affected is λ, the probability of being a child from an affected household is $\mu_{c,}$ and the probability of being ill if the individual concerned is a child from an affected household is $\varphi_{c,}$ then the probability of observing illness in a child in a representative household is defined as $\lambda \mu_c \varphi_{c.}$

Similarly, the probability of illness can be calculated for an adult and for an elderly for a representative household, as $\lambda \mu_a \varphi_a$ and $\lambda \mu_e \varphi_{e,}$ where the subscripts a and e denote adults and elderly, respectively. The average cost of treatment is also calculated for each age category, and can be denoted by c_c, c_a and c_e, where the subscripts denote child, adult and elderly respectively. Given the average size of the family (s), the cost of treatment for a representative household (c_1) is derived as:

$$c_1 = s \times \lambda (\mu_c \varphi_c c_c + \lambda \mu_a \varphi_a c_a + \lambda \mu_e \varphi_e c_e)$$

It is worth repeating that c_1 is the total cost of illness for a representative household for a period of 15 days, since 15 days was the recall period used in the survey. This c_1 can be alternatively interpreted as the (minimum) offer price for a representative household, for avoiding illness.

In defining children as below 15, the study adopts the International Labour Organization's mandate on defining child labour for most sectors as those who are aged less than 15 and working (1981). In the present dataset, no child below 15 was reported as being employed. Adults are defined as those between 15 to 60 years of age; the upper limit of 60 years is based on the retirement age currently adopted by the Government of India. The households sampled can also be classified into two groups. Those households that have reported at least one illness are hereafter referred to as the affected households, while the rest are the unaffected ones. There are a total of 266 households who are affected by this classification. The individuals surveyed in these affected households can similarly be classified by their age into adults, children and the elderly.

In all there were 1422 individuals in these affected households, comprising 544 children, 824 adults and 36 elderly. On the same lines, the data from the affected households is used to identify the numbers for those who are ill among these 1422 individuals. While there were 150 children who had reported illness episodes, 168 adults and 9 of the elderly reported at least one illness episode during the preceding 15 days, at the time of the survey. In all, there were 327 individuals who had been affected by at least one episode of illness. Table 7 gives the detailed results for the computations on treatment costs.

The wage loss from illness

The preceding section derives the cost of treatment for a representative household. From an economics perspective, the opportunity cost of an illness episode would comprise of the costs of treatment as well as the implied wage loss, arising from mandays lost due to ill health. The wage loss is computed for only the adult population in the present study since as discussed earlier, child labour (i.e. employment below 15 years of age) does not feature in this dataset.

The probability of an adult being ill has already been derived in the previous section as $\lambda \mu_a \varphi_{a.}$ This probability of illness is multiplied by the average man days lost, the average wage rate over all working adults, the rate of employment for the sample, and finally the average size of the household (s), in order to arrive at the total wage loss (w_1) for a period of fifteen days, for the representative household. Table 8 presents the computations for the wage loss.

Cost of illness for the population

To arrive at the total social value of the damages from contaminated water, the damages have to be computed for the entire population under discussion. The first step involved would be to derive the annual cost of illness, combining both the wage loss and costs of treatment for a household. A simple sum of these two components leads to a total cost of Rs 71.43 as the average cost for a representative household over 15 days during the peak period for diarrhoeal illness. This is an important point since the survey was conducted during the months of May and June, when diarrhoeal illness is at its peak.

Thus, while deriving the annual cost of illness, the variation in the incidence of illness over the year has to be kept in mind.

Therefore, a set of weights was derived by which the cost of illness obtained for the survey period, could be scaled according to the distribution of cases across the 12 months. The average number of cases reported from the colonies covered in the survey over the three years from 1996 to 1998 was calculated for each month. The resulting figures for May and June were 33.33 and 44.67, respectively. An average of these two months' cases [(33.33 + 44.67)/2 = 39] was computed. Thus, each month's average number of cases was divided by this scaling factor of 39, in order to convert the average cases to a weighting factor for the cost of illness. Therefore, the cost of illness during 15 days in each month is obtained by multiplying the cost for the peak period by the corresponding weight for that month. The monthly cost is derived by doubling these figures. Table 9 presents the results on the total cost of illness.

Conclusion

The present study can be viewed at one level as an empirical investigation on the effects of both engineering /infrastructural variables and individual behaviour on diarrhoeal disease in low-income pockets of Delhi. This empirical analysis would provide policy inputs, for planners so as to make their intervention strategies more effective in terms of reducing the burden of diarrhoeal illness in a cost-effective manner.

Overall the results seem to point towards the importance of overall infrastructural variables such as garbage removal facilities, water sources, interruptions in water supply, which may not always be within the control of the individual household. Community and government involvement is called for, the latter more so in the context of the low income households which have been studied here. Thus even variables like mother's educational level are not found to be significant explanatory variables linked with the occurrence of diarrhoeal illness. This result differs from most other studies for rural India (Panda 1996).

The results obtained indicated the importance of safe water supplies from the point of view of both the household and the civic authorities who are interested in interventions to control diarrhoeal illness. Comparison of the costs of illness with the costs that would be incurred in supplying an adequate quantum of water to these very same households could for example, be crucial inputs into policy making for the water sector. Again, the costs of illness could be used for judging the willingness-to-accept compensation for the affected households. Such a criterion for valuation would be more acceptable particularly for the low-income, urban context as compared to alternative exercises of benefit revelation such as direct willingness-to-pay methods of elicitation. The willingness-to-pay measures are conditioned on the income distribution within a community. This makes the efficiency criterion a poor and inadequate one for judging the benefits of a water supply project when one is considering the desirability of a project in terms of its social value as much as its pure economic value. In the context of the present study of primarily low-income households in a developing economy, who experience inadequate access to a basic infrastructural service, this point is a crucial one.

Table 1: Source of Drinking Water by (Household) Income Quintiles

5 Quintiles of Income						
Source	*1*	*2*	*3*	*4*	*5*	*Total*
1	19	12	16	13	7	67
2	10	6	18	11	5	50
3	0	0	0	0	2	2
4	6	2	13	13	17	51
5	30	19	42	38	61	190
6	75	32	39	12	15	173
7	13	9	15	4	4	45
8	6	3	4	1	4	18
9	0	0	0	2	5	7
Total	59	83	147	94	120	603

Codes for source of drinking water: 1 - shallow handpump, 2 - tanker (municipal supply), 3 - community tank, 4 - deepbore tubewell, 5 - individual piped connection (legal), 6 - public hydrant, 7 - individual piped connection (illegal), 8 - Mark II (deep bore) handpumps, 9 - piped water from private deep bore tubewell.
Income ranges (in Rs per month) ~ 1: 1000 - 2500, 2: 2600 – 3250, 3: 3300 – 4500, 4:4600 – 6000, 5: 6400 – 25000.

Table 2: Distribution of Households Reporting Diarrhoea Cases (by Income Categories)

Number of Cases	*Income Quintiles*					*Total*
	1	*2*	*3*	*4*	*5*	
1	53	21	55	25	46	200
2	11	15	5	11	9	51
3	0	1	2	2	3	8
4	0	0	0	2	1	3
5	0	0	1	0	1	2
Total	64	37	63	40	60	264

Income ranges(in Rs per month) ~ 1: 1000 - 2500, 2: 2600 – 3250, 3: 3300 – 4500, 4:4600 – 6000, 5: 6400 – 25000.

Table 3: Source-wise Distribution of Water Samples

Source	*Total samples tested (%)*	*Proportion of unacceptable samples (%)*
1	22	93
2	6	0
3	0	-
4	6	50
5	23	37.5
6	28	31.6
7	8	60
8	6	100
9	1	100
Total	100	53

Codes for source of drinking water: 1 - shallow handpump, 2 - tanker (municipal supply), 3 - community tank, 4 - deepbore tubewell, 5 - individual piped connection (legal), 6 - public hydrant, 7 - individual piped connection (illegal), 8 - Mark II (deep bore) handpumps, 9 - piped water from private deepbore well

Note: The total number of water samples tested is 90.

Table 4: Illness Incidence Rates for the Sample Households

Definition	*Incidence rate*
1.For entire sample (number of cases/ total population)	113.10
2. Male children (total reported cases of male children ill/total number of male children)	128.55
3. Female child (total reported cases of female children ill/total number of female children)	135.36

Notes: A child is defined as one below 15 years of age.

Definition of incidence rate (IR) = (total number of cases/ total population surveyed) * 1000

Table 5: Summary Statistics on Explanatory Variables

VARIABLE	MEAN VALUE
Source (=1 if non-piped water)	0.51
Sanitary latrine (=1 if sanitary latrine is accessed)	0.64
Waste (=1 if no municipal collection of household solid waste)	0.98
Timings (=1 if intermittent water supply for piped source)	0.52
Per capita income(rupees per month)	987.13
Storage (=1 if no separate storage of drinking water)	0.11
Dump (=1 if local garbage dump with irregular removal present)	0.29
Educational attainment of head of household (on a 5 point scale)	3.06
Access to latrine (=1 if there is access to a latrine)	0.08
Sewer(=1 if there is no sewer facility)	0.81
Foul Smell (=1 if water is foul smelling)	0.50

Table 6: Bivariate Probit Results

Equation 1: dependent variable = diarrhoeal illness

Independent Variable	*Coefficient Value*	*T – statistic*
Source	0.74	6.25*
Sanitary latrine	-0.10	-0.87
Mode of disposal of solid waste	1.12	1.68**
Timings	0.28	2.37*
(Log) Per capita income	-0.23	-2.19*
Storage	-0.31	-1.75**
Dump	0.25	2.07*
Education of Head_2	0.19	1.02
Education of Head_3	0.03	0.15
Education of Head_4	-0.08	-0.47
Education of Head_5	0.10	0.54
Location dummy_1	0.56	3.47*
Location dummy_2	0.27	1.41
Location dummy_3	-0.05	-0.27
Location dummy_4	0.60	1.95**
Constant	-0.51	-0.50

Equation 2: dependent variable = purification of water

Independent Variable	*Coefficient Value*	*T – statistic*
Source	-0.11	-0.77
Timings	0.32	2.07*
(Log) Per capita income	0.28	2.12*
Education of Head_2	1.08	3.17*
Education of Head_3	0.61	1.69**
Education of Head_4	0.99	3.03*
Education of Head_5	1.12	3.29*
Access to latrine	0.47	1.81**
Sewer	-0.44	-2.66*
Foul smell	0.41	2.73*
Location dummy_1	-0.04	-0.22
Location dummy_2	-0.42	-1.85**
Location dummy_3	-0.82	-3.35*
Location dummy_4	-0.44	-1.13
Constant	-3.58	-3.49*

Notes: Number of observations used is 603. * denotes that the t-statistic is acceptable at the 95% level of confidence and ** denotes acceptance at the 90% level of confidence.

Table 7: Treatment Costs of Illness

Variable name	*Value*
λ	0.44
μ_c	554/1422 = 0.39
φ_c	150/554 = 0.27
probability of a child being ill = $\lambda\ \mu_c\varphi_c$	0.046
μ_a	824/1422 = 0.58
φ_a	168/824 = 0.203
probability of an adult being ill = $\lambda\ \mu_a\varphi_a$	0.052
μ_e	36/1422 = 0.025
φ_e	9/36 = 0.25
probability of an elderly being ill = $\lambda\ \mu_e\varphi_e$	0.003
average cost of treatment for a child (c_c)	83.33 (Rs)
average cost of treatment for an adult (c_a)	52.96 (Rs)
average cost of treatment for an elderly (c_e)	70.00 (Rs)
average size of family (s)	5.34
cost for a representative household: $c_1 = s \times \lambda\ (\mu_c\ \varphi_c\ c_c + \mu_a\ \varphi_a\ c_a + \mu_e\ \varphi_e\ c_e)$	36.26 (Rs)

Table 8: Wage Losses due to Illness

Variable name	*Value*
probability of an adult being ill = $\lambda\ \mu_a\varphi_a$	0.052
average size of household (s)	5.34
average man days lost (m)	2.52
rate of employment (rt)	0.50
average daily wage-rate (w)	100.53 (Rs)
wage-loss for a representative household: $w_1 = \lambda\ \mu_a\varphi_a \times s \times m \times rt \times w$	35.17 (Rs)

Table 9: Total Cost of Illness for the Population

Month	Sum of cases 1996-98	Average cases	Weighting factor	15 days cost of illness (Rs)	Monthly cost of illness (Rs)
January	0	0	0	0	0
February	0	0	0	0	0
March	2	0.67	0.02	1.43	2.86
April	20	6.67	0.17	12.14	24.29
May	100	33.33	0.85	60.72	121.43
June	134	44.67	1.15	82.14	164.29
July	268	89.33	2.29	163.57	327.15
August	218	72.67	1.86	132.86	265.72
September	103	34.33	0.88	62.86	125.72
October	41	13.67	0.35	25.00	50.00
November	9	3.00	0.08	5.71	11.43
December	1	0.33	0.01	0.71	1.43

Cost of illness for a representative household: (Rs) 1094.31
Cost of illness for 603 sampled households: (Rs) 659868.93
Total annual cost of illness for population (150748 households): (Rs) 1649.65 lakhs

References

Alberini, A., Eskeland, G.S., Krupnick, A. and McGranahan, G. (1996), "Determinants of Diarrhoeal Disease in Jakarta", *Water Resources Research* 32(7): 2259 - 69.

Anderson, D. and Cavendish, W. (1992), "Efficiency and Substitution in Pollution Abatement: Three Case Studies", World Bank Discussion Papers no. 186, World Bank, Washington D.C.

Barah, B.C., Sipahimalani, V. and Dhar, P. (1998), "Urban Water Supply and Sanitation", in U. Sankar and O.P. Mathur, eds., *Economic Instruments for Environment Sustainability*, New Delhi: National Institute of Public Finance and Policy.

Bhandari, N. (1992), "The household management of diarrhoea in its social context : study of a Delhi slum", Unpublished Ph.d. thesis, CSMCH, Jawaharlal Nehru University, New Delhi.

Bhatnagar, S. and Dosajh, U. (1986), "Diarrhoeal disease morbidity in children below 5 years in urban slums in Delhi", National Institute of Health and Family Welfare, New Delhi.

Briscoe, J. (1984), "Intervention studies and the definition of dominant transmission routes", *American Journal of Epidemiology* 120: 449-55.

Esrey, S., Feecham R.G. and Hughes, J.M. (1985), "Interventions for the control of diarrhoeal diseases among young children: improving water supplies and excreta disposal facilities", *World Health Organisation Bulletin no. 63(4*): 757-72.

Esrey, S.A., Potash J.B., Roberts L. and Shiff C. (1991), "Effects of improved water supply and sanitation on ascariasis, diarrhoea, dracunculiasis, hookworm infection, scchistosomiasus and trachoma", *World Health Organisation Bulletin no.*69(5): 609-21.

Mehrotra, R P. (1988), "The gastroenteritis outbreak and its control in Sunder Nagri: An assessment at community level", Unpublished report, CSMCH, Jawaharlal Nehru University, New Delhi.

Panda, P.K. (1996), "The effects of safe drinking water and sanitation on diarrhoeal diseases among children in rural Orissa", in K. Pushpangadan, ed, *Proceedings National Seminar on Rural Water Supply and Sanitation*, Centre for Development Studies: Trivandrum.

WHO (1985), *Treatment and Prevention of Acute Diarrhoea: Guidelines for the Trainers of Health Workers*, Geneva: World Health Organisation.