NBER WORKING PAPER SERIES

ARE USER FEES REGRESSIVE? THE WELFARE IMPLICATIONS OF HEALTH CARE FINANCING PROPOSALS IN PERU

Paul Gertler

Luis Locay

Warren Sanderson

Working Paper No. 2299

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 1987

Please address all correspondence to Paul Gertler. The research reported here is part of the NBER's research program in Health Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #2299
June 1987

Are User Fees Regressive?
The Welfare Implications of Health Care Financing Proposals in Peru

ABSTRACT

In this paper, we derive a discrete choice model of the demand for medical care from a theoretical model that implies a natural interrelation between price and income. We show that, in the context of a discrete choice model, if health is a normal good, then the price elasticity of the demand for health care must decline as income rises. This implies that the models in previous discrete choice studies which restrict the price effect to be independent of income are misspecified.

The model is estimated using data from a 1984 Peruvian survey, and a parsimonious flexible functional form. Unlike previous studies, we find that price plays a significant role in the demand for health care, and that demand becomes more elastic as income falls, implying that user fees would reduce the access to care for the poor proportionally more than for the rich. Our simulations show that user fees can generate substantial revenues, but are accompanied by substantial reductions in aggregate consumer welfare, with the burden of the loss on the poor. These results demonstrate that undiscriminating user fees would be regressive both in terms of access and welfare.

Paul Gertler
Department of Health Policy and Management
Harvard School of Public Health
677 Huntington Avenue
Boston, MA 02115

Luis Locay
Department of Economics
State University of New York at Stony Brook
Stony Brook, NY 11794

Warren Sanderson
Department of Economics
State University of New York at Stony Brook
Stony Brook, NY 11794

I. INTRODUCTION

Many developing countries have created extensive publicly supported health care systems, access to which is at little or no cost.[1] The financial crisis of the 1980's has forced many of them to consider instituting user fees (i.e. charge individuals for access). Those in favor of user fees argue that they facilitate recovery of the cost of providing the service, and, if they are set at marginal cost, improve allocative efficiency.[2] The strongest argument against user fees is that they may be regressive in that they may not allow all income groups equal access to medical care because the poor may be more price sensitive than the rich. Even if everyone is equally price sensitive, user fess will be regressive if the welfare loss for the poor relative to income is larger than for the rich.

In the absence of user fees, equal access is still not assured. It has been well known since Acton (1975) that nonmonetary access costs such as travel time are important determinants of health care choices. The geographical distribution of services may make access more difficult for some groups. For example, locating facilities closer to the upper and middle classes discriminates against the poor. User fee proponents argue that revenues can be reinvested to reduce nonmonetary access costs, and consequently minimize consumers' welfare loss.

Since user fee proposals are so widespread and the potential welfare effects so large, it is important that some ex ante analysis be performed. This paper provides a methodology for such an ex ante analysis, and to our knowledge, the first estimates of expected revenues and welfare losses (measured as compensating variations) associated with one such proposal.

The analysis requires estimation of the demand for health care, from which the revenues and welfare changes of proposed user fees can be simulated. The magnitude of the revenue and welfare effects depend crucially on the price elasticity of demand. Previous studies in developing countries have found little if any impact of price on demand. These studies model the demand for health care as a discrete choice between alternative providers, with the price effect specified to be independent of income.[3] This assumption is extremely restrictive, since one would expect the wealthy to be less sensitive to price differences across providers than the poor. Indeed, we show that this specification is inconsistent with stable utility maximization, and that, if health is a normal good, the demand for health care must become more price elastic as income falls.

The discrete choice specification in this paper is derived from a theoretical model that implies a natural interaction between price and income in the demand functions, and those demand functions are estimated using a parsimonious flexible functional form that allows the data to determine the effect of income on price elasticities. The resulting model facilitates the study of the distributional impacts of user fees.

The empirical investigation considers the potential effects of user fees in urban Peru. The estimates show that price plays an important role in health care demand. Further, demand becomes more elastic as income falls indicating, as expected, that health is indeed a normal good. This implies that the introduction of health care user fees in Peru would reduce access proportionally more for the poor than the rich, and, in this sense, be regressive. Our simulations demonstrate that while user fees would generate substantial revenues, they would also generate substantial reductions in aggregate consumer welfare with the burden of the loss on the poor. The simulations also indicate that the welfare loss from the current spatial distribution of public health care services is roughly equal to the expected welfare loss from moderate user fees, and that the loss is fairly evenly distributed across income groups. Therefore, if the government imposed moderate user fees and used the revenues to solve the rationing problem, there would be little if any aggregate welfare loss, but there would be a redistribution of welfare from poor to rich.

II. BEHAVIORAL ASSUMPTIONS

The framework for this discussion is a static model in which utility depends on health and consumption of goods other than medical care. When an illness or accident is experienced, individuals must decide whether to seek medical care. The benefit from consuming medical care is an improvement in health, and the cost of medical care is a reduction in the consumption of other goods. Individuals not only have to decide whether to seek care, but also what type of care. They are faced with a set of alternative providers, each of which has a different potential impact (efficacy) on their health. This efficacy depends on providers' skills, individuals' characteristics (e.g. medical problems, general health status, and ability to implement the recommended treatment plan), and a random term that captures the notion that the efficacy of medical care is not deterministic. An individual's expectation of this impact can be viewed as the perceived quality of care.

In essence, individuals are faced with a discrete choice decision. A choice must be made between the various provider alternatives, including self-care. Each alternative offers a set package (quality) for a given price, where the price includes both monetary outlays and nonmonetary access costs such as travel and waiting time. Based on this information, their health statuses, types of medical problems, and incomes, individuals choose the alternatives that yield the greatest utilities.

We consider the short run utility maximization problem faced by an individual who has recently experienced an accident or illness. Let the utility, conditional on receiving care from provider j, be given by

$$U_j = U(H_j, C_j, T_j), \qquad (1)$$

where H_j is expected health status after receiving treatment from provider j, C_j is expenditures on consumption after paying provider j, and T_j is the nonmonetary cost of access to provider j.

The health care purchased from provider j is invested in health. The perceived quality (marginal product) of provider j's medical care is the expected improvement in health. Let H_0 be expected health status without professional medical care (i.e. self-treatment); then, the perceived quality of provider j's care is $Q_j = H_j/H_0$, which yields an expected health care production function of the form

$$H_j = Q_j \, H_0, \qquad (2)$$

where H_j is proportional to H_0. The quality parameter depends upon provider characteristics (e.g. training and facilities) and individual characteristics (e.g. type and severity of illness).

This production function takes on a rather simple form for the self-care alternative. Since H_j equals H_0, the proportionality factor is unity for the self-care alternative. In effect, this normalizes the health care production function so that the quality of a particular provider's care is measured relative to efficacy of self-care.

The level of consumption expenditure conditional on choosing provider j, C_j, is derived from the budget constraint. Let P_j be provider j's price and Y be income, then

$$C_j = Y - P_j, \tag{3}$$

with $C_j \geq 0$ required for feasibility.[4] Substitution of (3) into (1) yields

$$U_j = U(H_j, Y - P_j, T_j).$$

Income affects utility through the consumption term, and is assumed to be exoegnous.[5]

Now we are ready to specify the utility maximization problem. Suppose the individual has J+1 feasible alternatives (with the j=0 alternative being self-care). The unconditional utility maximization problem is

$$U^* = \max(U_0, U_1, \ldots, U_J), \tag{4}$$

where U^* is the highest utility the individual can attain.

If health is a normal good, then the demand for health increases with income. A necessary condition for normality is that as income rises, the marginal rate of substitution of consumption for health diminishes, holding health constant. This point is demonstrated in figure 1, where the continuous choice case with health being a normal good is pictured. As income rises the point of utility maximization moves out from the origin along the expansion path. Holding health constant at $\bar{H}$, we move to the right along the horizontal line as income rises, intersecting the indifference curves at points of flatter slopes, implying a diminishing marginal rate of substitution.

In a discrete choice world, normality implies that as income rises individuals are more likely to choose the "higher price/higher quality" options. Here as well, a necessary condition for normality is that as income rises, the marginal rate of substitution of consumption for health diminishes, holding health constant. This is demonstrated in figure 2, where the discrete choice case with health as a normal good is pictured. In figure 2, there is a choice between a "high price/high quality" option (P_h, Q_h), and a "low price/low quality" option (P_ℓ, Q_ℓ). At a low income level, say Y_ℓ, the choice is between points A and B; i.e. between a gain in health of ($H_h - H_\ell$) and a gain in consumption of ($P_h - P_\lambda$). At income Y_ℓ, the additional consumption is preferred to the additional health and the "low price/low quality" option B is chosen. The high income individual with income Y_h has a choice between points C and D. These points represent the same tradeoff between health and consumption as points A and B. As income rises the marginal rate of substitution of consumption for health falls along both horizontal lines H_h and H_ℓ. Eventually, at some income between Y_ℓ and Y_h, the gain in health is preferred to the gain in consumption. At income Y_h, the "high price/high quality" option C is chosen.

In a discrete choice world, if health is a normal good, a rise in income increases the likelihood that individuals purchase "higher price/higher quality" alternatives. Another way of looking at this is that an increase in price is less likely to dissuade richer individuals from choosing the "higher price/higher quality" alternatives. In a probabilistic sense, normality implies that richer individuals are less price elastic than poorer individuals.

III. EMPIRICAL SPECIFICATION

The solution to (4) yields a system of demand functions, whose forms are probabilities that the alternatives are chosen given that an individual experiences an accident or illness. The demand function for a given alternative is found by calculating the probability that this particular alternative yields the highest utility amongst all the alternatives. The functional form of the demand functions depends on the functional form of the utility function conditional upon choosing a particular provider and the distribution of the stochastic variables.

A. The Conditional Utility Function

It is customary to begin by considering a linear functional form for the conditional utility function in (1). Substitution of (3) into a linear utility function yields

$$U_j = \alpha_1 H_j + \alpha_2 (Y - P_j) + a_3 T_j + \epsilon_j \tag{5}$$

where ϵ_j is a random taste shock that is uncorrelated across alternatives. Notice that $\alpha_2 Y$ enters each alternative's utility function, implying that the influence of income on utility does not vary by alternative. Since only differences in utility matter, a linear utility function imposes the restriction that income has no effect on the choice of provider and that the marginal rate of substitution is constant. Therefore, this specification is inconsistent with health being a normal good.

A common method of trying to relax this restriction is to allow the coefficient on consumption to vary by alternative.[6] That specification violates the maximization of a stable utility function. It asserts that, holding income, prices, and health constant, the marginal rate of substitution varies by alternative.

A parsimonious parameterization that does not place second order restrictions on the marginal rate of substitution, does not violate the maximization of a stable utility function, and is linear in parameters, is the semi-translog, where health and access costs enter in log form and consumption enters in both log and log squared form.[7] Substitution of (2) and (3) into a semi-translog conditional utility function yields

$$U_j = \ln H_0 + \ln Q_j + \alpha_1 \ln(Y - P_j) + \alpha_2 \ln(Y - P_j)\ln(Y - P_j) + \alpha_3 \ln T_j + \epsilon_j \quad (6)$$

The quadratic term is necessary so that the specification does not impose normality and a diminishing marginal rate of substitution, but rather allows us to test for them.

B. Quality

In equation (6) neither $\ln H_0$ nor $\ln Q_j$ are observed. Since $\ln H_0$ appears in the utility function for all the choices and its value does not vary by alternative, it does not influence which alternative is preferred, and therefore can be ignored.

A more difficult issue arises because of the unobservability of $\ln Q_j$. To solve this problem we specify a quality (marginal product) function for each provider type. Specifically, let the expected quality from provider j be

$$\ln Q_j = \beta_{0j} + \beta_{1j} X + \beta_{2j} Z_j + \tau_j, \quad (7)$$

where X is a vector of the individual's characteristics (i.e. measures of health status, severity of illness and education), Z_j is a vector of characteristics of provider j, and τ_j is a random shock. The error term τ_j represents unobserved individual characteristics, such severity and complexity of illness, that may affect the providers' marginal productivities relative to self-care. Recall that quality is normalized relative to the self-care alternative, implying that $\ln Q_0 = 0$. The error term t_j may be correlated across the non-self-care alternatives.

The reduced form conditional utility function for alternative j is found by substituting (7) into (6). Specifically, for alternatives j=1,...,J,

$$U_j = V_j + \epsilon_j + \tau_j, \tag{8}$$

where

$$V_j = \beta_{0j} + \beta_{1j}X + \beta_{2j}Z_j + \alpha_1 \ln(Y - P_j) + \alpha_2 \ln(Y - P_j)\ln(Y - P_j) + \alpha_3 \ln T_j.$$

Note that the intercept and coefficients on the quality terms vary by alternative as do the values of consumption and access costs (but not their coefficients). Since $\ln Q_0 = 0$, $T_0 = 0$, and $P_0 = 0$, the reduced form conditional utility function for the self-care alternative becomes

$$U_0 = \alpha_1 \ln Y + \alpha_2 \ln Y \ln Y + \epsilon_0.$$

Note further that τ_0 does not exist as quality is normalized relative to the self-care alternative.

C. The Budget Constraint

Specification of the budget constraint requires determining the relevant budgeting period. Since the health care decision is discrete and made irregularly, consumers may be willing to borrow against future income. If capital markets are perfect and individuals (or families) can borrow without restriction, the relevant income constraint is the present value of income, or wealth. The other extreme assumption is that no resources outside each income period can be used. The actual period may be somewhere in between.

We let the data determine the appropriate budgeting period. Define y as permanent monthly income and r as the period discount rate, then the constraining income in (4) is ky, where the parameter k is a function of the length of the budgeting period and r. If budgeting is restricted to one period, then k is equal to 1. If the budgeting period is infinity (i.e. there is perfect borrowing and lending), then k is equal to 1/r.

The addition of k implies (8) is no longer linear in parameters. We linearize (8) using an approximation to the log of consumption. The log of consumption can be expressed as

$$\ln(ky - P_j) = \ln(ky) + \ln\left(1 - P_j/ky\right). \tag{9}$$

Since P_j/ky, the budget share of alternative j, is expected to be small, the second term in (9) can be approximated by $-P_j/ky$, which allows us to rewrite the log consumption and log consumption squared terms in (8) as

$$\alpha_1 \ln(ky) + \alpha_2 \ln(ky)\ln(ky) - \left((\alpha_1 + 2\alpha_2 \ln k)/k\right)(P_j/y) +$$

$$(a_2/k^2)(P_j/y)^2 - (2\alpha_2/k)(P_j/y)\ln y. \tag{10}$$

Notice that the first two terms in (10) are the same across all alternatives, including self-care. Since only differences in utility across alternatives matter, these terms have no effect on provider choice, and therefore, can be left out. Further, when k equals one, (10) reduces to

$$- \alpha_1(P_j/y) + \alpha_2(P_j/y)\left(P_j/y - 2\ln y\right). \qquad (11)$$

Since both (10) and (11) are linear in parameters, they provide us with an easy likelihood ratio test for k equal to one.

D. The Demand Functions and Welfare

The demand function for an alternative is the probability that its utility is greater than from any other alternatives. McFadden (1981) shows that, given reasonable distributional assumptions on ϵ_j and τ_j, these demands take on a nested multinomial logit (NMNL) form, where it is first decided whether to seek care, and then conditional on seeking care decide from which provider to seek care. The probability that provider j is chosen is

$$\Pi_j = \frac{\exp\left[\sigma \ln\left(\sum_{j=1}^{J} \exp(V_j)\right)\right]}{\exp(V_0) + \exp\left[\sigma \ln\left(\sum_{j=1}^{J} \exp(V_j)\right)\right]} \, \frac{\exp(V_j)}{\left(\sum_{j=1}^{J} \exp(V_j)\right)} \; .$$

and the probability of self-care is

$$\Pi_j = \frac{\exp(V_0)}{\exp(V_0) + \exp\left[\sigma \ln\left(\sum_{j=1}^{J} \exp(V_j)\right)\right]},$$

where the V_j's are given by (8) with (10) substituted for the log consumption terms. Also the $\alpha_1 \ln(ky)$ and $\alpha \ln(ky)\ln(ky)$ are excluded as they do not vary by alternative, which inplies that $V_0 = 0$. The parameter σ is one minus the correlation of the j=1,...,J utilities introduced by the τ_j's.

McFadden also shows that NMNL reduces to a multinomial logit (MNL) when σ is unity. The NMNL is more general than MNL in that it allows correlation between the utilities that share common attributes, and therefore does not suffer from the independence of irrelevant alternatives assumption.

The estimated demand functions can be used to project the impact of user fees on demand (and revenues), and the number of people who do not seek health care as a result of user fees. These demand functions also form the basis of our computation of the welfare costs of user fess, where the welfare costs are measured by compensating variations.[8] For example, consider changing the vector of provider prices from P^1 to P^2. Following Small and Rosen (1981), in the case of a nested multinomial logit, the amount of income the individual must be given to make him as well off at P^2 as at P^1 is

$$\Delta e = (1/\lambda)\left\{\ln\left[\exp(V_0^1) + \left(\textstyle\sum_{j=1}^{J} \exp(V_j^1)\right)^{\sigma}\right] - \ln\left[\exp(V_0^2) + \left(\textstyle\sum_{j=1}^{J} \exp(V_j^2)\right)^{\sigma}\right]\right\} \quad (12)$$

where V_j^1 and V_j^2 are evaluated at P^1 and P^2, respectively, and λ is the marginal utility of income.[9] The compensating variation for nonprice changes (such as travel time) can be similarly calculated.

IV. DATA AND INSTITUTIONAL ENVIRONMENT

The empirical work utilizes data from a 1984 Peruvian household survey, the Encuesta Nacional de Nutricion y Salud (ENNSA). The survey contains a rich set of socio-economic data, as well as morbidity and health care utilization information for a two week recall. Since this study analyzes contingent health care demand, we restricted our sample to those persons who reported having symptoms or an accident. The sample was taken from individuals living in the urban Sierra and Lima regions. Rural regions were excluded because reliable income data do not exist for them. A sample of 3412 individuals age 16 and above is the basis for this work. Descriptive statistics are presented in table 1.

Peru has a mix of public and private health care. The major provider of public health care is the Ministry of Health, which operates hospitals and clinics. The next largest provider of public health care is the Instituto Peruano de Seguridad Social (Social Security). It operates hospitals for its members, which are not available to non-members. In the analysis, Social Security hospitals are not viewed as a separate alternative, but rather are included in the public hospital alternative. A dummy variable indicating whether the individual was a Social Security member is included in the hospital equation to account for quality differences. The dominant private health care providers are physicians. Other types of private providers, such as traditional healers, and pharmacists were not numerically important, and were merged with the no consultation group to form our "self-care" alternative. The four alternatives are: (1) self-care, (2)public hospital, (3) public clinic, and (4) private doctor.

The arguments of the quality (marginal product) function are the initial state of health, the type of illness, human capital, and provider characteristics. Measures of health status prior to treatment are age and type of illness, which is measured by a set of dummy variables indicating whether the individual's medical problem was an accident or acute illness, digestive illness, respiratory illness, or other illness. The other illness dummy variable was excluded. The quality of providers is thought to vary by location. Hence, a set of regional dummy variables indicating if the individual lives in central Lima, the north and south cones of Lima, and the north, south, and central regions of the Sierra are included. The central Sierra dummy variable was excluded. In addition, the individual's education was included as a measure of human capital.

Income was measured as total family income in the month prior to the survey. Family income is the relevant concept here because family members are not provided or denied health care on the basis of their labor force statuses. This measure reduces the sensitivity of income to the illness of any particular family member.

Since income does not vary by alternative, we need variation in prices across alternatives to identify and estimate the coefficients on the log consumption and log consumption squared terms. In a discrete choice framework, identification requires variation across alternatives. Although variation across individuals is not necessary, it is desirable as it improves the estimation precision. In our data the public hospital and clinic prices do not vary by individual, but there is substantial cross-individual variation in private doctor prices as the data covers many different regions, were collected over a nine month period in which relative prices changed substantially.

Measuring prices posed a difficult problem. The model requires prices for each alternative, but these were not directly available. The ENNSA only collected price information for the provider from which the individual received care. For those who sought care, price data were only available for the alternative they chose, and for individuals who did not seek care there is no information.

The measurement problem was easily solved for hospitals and clinics, since they charged a user fee of 1,000 to 2,000 soles. In our sample, about 35% of hospital and clinic users reported paying nothing, about 50% reported paying 1,000 soles, and almost all the rest reported 2,000 soles. About half of the reported zero fees are from Social Security hospitals, which do not charge their members for services. The other half are probably a result of failure to collect the fees. Since these prices are minuscule relative to monthly family income (see table 1), we assumed individuals expected to pay 1,000 soles at Ministry of Health hospitals and clinics.

For private doctor prices, we used the available information to estimate hedonic price equations, and then imputed prices for all individuals. The equation specified price to be a function of age, illness, and market structure variables such as population and availability of health care services. Income was not used in order to avoid attributing higher prices to higher income individuals who may have purchased higher quality care. An additional problem was selectivity bias. The observed distribution of prices paid will not be representative of the *ex ante* distribution of prices because individuals are more likely to chose low price alternatives. We corrected for this selectivity bias by following an instrumental variables procedure used in Dubin and McFadden (1984).[10]

Finally, we measure nonmonetary access cost by travel time to the provider. The travel time data suffer from the same problems as the price data. In addition, travel time information was collected in discrete categories. Binary logit hedonic travel time equations (with selectivity bias correction) were used to estimate the probability of travelling more than an hour.

V. RESULTS

The parameters of a MNL and a NMNL were estimated by maximum likelihood. The NMNL nested the choice of provider within the choice of whether to seek care at all. The hypothesis that the NMNL is not different from the MNL was accepted at the .05 level, and the hypothesis that k equals unity was also accepted at the 0.05 level.[11] The estimated coefficients and associated t-statistics for the MNL with k equal to one are presented in table 2.

The coefficients on log consumption and log consumption squared are significant at the .1 and .01 levels, respectively. Price and income therefore play important roles in the demand for medical care. Since price and income enter in a highly nonlinear form it is difficult to assess their influence on demand just from looking at the coefficient values. For this reason, arc price elasticities for clinic, hospital and private doctor services were computed by sample income quintile and are presented in table 3. The price elasticities are negative over all prices and income groups, and demand is more elastic at lower incomes and at higher prices. The magnitude of the prices elasticities varies greatly by income. In the highest income quintile, demand appears to be completely inelastic, while demand in the lowest income quintile is much more sensitive to price.

We have assumed that income is exogenous. If, in fact, income is endogenous, there is a possibility of simultaneity bias. The bias is likely to have a downward impact on the estimated price and income effects, making them closer to zero. The effect we are interested in measuring is the causal impact of changes in income on health care demand. If health is a normal good, then that effect is positive. The simultaneity bias arises because an accident or illness may reduce income. The more severe and complex the illness or accident the greater the reduction in income. However, the more severely ill have greater medical need and are therefore more likely to seek medical care. This implies that the observed relation between income and demand will likely be biased towards zero. Since price enters our model as a reduction in consumption (Y - P), its effect is also likely to be biased towards zero. Therefore, our estimated price elasticities should be lower bounds on the true elasticities.

The coefficient on the probability of traveling more than an hour is negative and estimated with precision. This implies that increases in nonmonetary access costs reduce demand.

The estimated quality parameters are consistent with our expectations. The coefficients on age are positive and significant in the hospital and private doctor equations, and negative in the clinic equation. Hence, older individuals perceive private doctor and hospital care to be of higher quality than self-care and clinic care, and self-care to be of higher quality than clinic care. The coefficients on education are positive and significant in the private doctor and hospital equations, and negative and significant in the clinic equation. The coefficient estimates imply that education increases the expected productivity of private doctor care and hospital care relative to self-care, and reduces the expected productivity of clinic care relative to self-care.

The coefficients on the acute illness (emergencies) imply that hospitals and clinics have a comparative advantage in treating these problems over private doctor or self-care. Individuals with respiratory illnesses believe that they they have a comparative advantage in treating themselves. Finally, Social Security hospitals are perceived to provide higher quality than Ministry of Health hospitals, and there is perceived quality variation by region.

VI. USER FEE SIMULATIONS

In this section we use the estimated demand functions to simulate the effects of user fees. A uniform fee is imposed at public facilities (hospitals and clinics). We consider two levels of fees, 10 and 20 thousand soles. These are realistic fee levels; the average fee for a visit to a private doctor was about 20 thousand soles. Monthly demands, revenues, and compensating variations are calculated by summing the individual estimates over the sample and then extrapolating to obtain population projections. Revenues are calculated in April 1984 soles. The base for the extrapolation is the product of the regional population and the overall regional probability of having an illness. Two private markets scenarios are considered: (1) where private doctors do not adjust their prices in response to the changes in public user fees, and (2) where private doctors adjust their prices by the same amount. Further, these scenarios are analyzed under the assumption (1) that the resulting revenues are not reinvested in the health care system, and (2) that the revenues are used to reduce nonmonetary access costs.

A. User Fees Without Reinvestment

Columns 3, 4, and 5 of table 4 report the results of the aggregate user fee simulations under both scenarios. They report the cumulative percentage change in total demand, the increase in public (hospital plus clinic) revenues and the welfare loss due to the user fee increase.[12] The results show that the imposition of moderate user fees can generate substantial public revenues with small reductions in the total demand for health care, but, of course, with even larger losses in consumers' welfare. Under scenario (1), for example, a user fee of 10 thousand soles generates approximately an additional 6,386 million soles per month in public revenues accompanied by a 7.5 percent reduction in demand and a fall of 7,123 million soles in consumers' welfare. Under scenario (2), that fee generates approximately 6,516 million soles with a 12.5 percent reduction in demand and a fall of 12,460 million soles in consumer welfare.

Even though the aggregate change in total demand appears to be modest, the effects on the lower income groups are quite large and substantially higher than in the upper income ranges. This is demonstrated in table 5 which shows the percentage change in total demand accounted for by each income quintile, and the welfare loss as a fraction of income for each income quintile. On average, the lowest income quintile accounts for about 40 percent of the total decrease in the quantity of health care demand, while the highest income quintile accounts for only about 5 percent. Not only is the reduction in total demand concentrated in the lowest income groups, but the greatest welfare loss (relative to income) is also borne by them. The simulations show that the lowest income quintile suffers a reduction of welfare of between a 3 and 11 percent of income, whereas the highest income groups loses less than one half of one percent.

B. User Fees With Reinvestment

In this set of experiments we assume the government uses the revenues to reduce nonmonetary access costs. In our model nonmonetary access costs are measured by travel time. This simulation assumes that the revenues are used to reduce everyone's travel time to a public clinic and hospital to within one hour or less (i.e. to reduce the probability of traveling more than one hour to a public facility to zero). This is a fairly egalitarian change because our data show that the median travel time probabilities are similar across all income groups.

Columns 6, 7, and 8 of table 4 report the aggregate results for the user fee experiment with reinvestment. Under both scenarios, a user fee of 10 thousand soles and a reduction of travel time to less than an hour increases total consumers' welfare, but a user fee of 20 thousand soles reduces consumers' welfare. Therefore, at a user fee somewhere between 10 and 20 thousand soles, consumers in the aggregate are indifferent between the current (1984) user fees and the higher user fees with easier access. The missing component of this comparison is whether the revenues generated by this user fee would be sufficient to cover the costs of building and operating the additional facilities necessary to reduce travel time.

Even if revenues were sufficient, such a policy would redistribute welfare from poorer to richer. This is demonstrated in table 6 which presents the percent change in total demand within each income quintile, and consumers' welfare loss as a fraction of income. An increase in user fees with reinvestment would result in a substantial decrease in demand by the poor and a slight increase in demand by the rich. In addition there would be a relatively large welfare reduction for the poor and a slight rise in welfare for the rich.

VII. SUMMARY AND CONCLUSIONS

We have derived a discrete choice model of the demand for medical care from a theoretical model that implies a natural interrelation between price and income. We show that, in the context of a discrete choice model, if health is a normal good, then the price elasticity of the demand for health care must decline as income rises. This implies that the models in previous discrete choice studies that restrict the price effect to be independent of income are misspecified.

We estimated this model using data from a 1984 Peruvian survey, and a parsimonious flexible functional form. Unlike previous studies, we find that price plays a significant role in the demand for health care, and that demand becomes more elastic as income falls, implying that user fees would reduce the access to care for the poor proportionally more than for the rich. Our simulations show that user fees can generate substantial revenues, but are accompanied by substantial reductions in aggregate consumer welfare, with the burden of the loss on the poor. These results demonstrate that user fees would be regressive both in terms of access and welfare.

The simulations indicate that the welfare loss for some people having to travel more than an hour to a public health care facility is roughly equal to the expected welfare loss from moderate user fees, and the first loss is fairly evenly distributed across income groups. Hence, if the government imposed moderate user fees and used the revenues to solve this access problem, there would be little if any aggregate welfare loss, but there would be a redistribution of welfare from poor to rich. This result is what one would expect in an urban environment where services are fairly evenly distributed, and may not be applicable to rural areas.

We have found that the introduction of user fees in Peru have the potential for raising significant revenues for cost recovery by shifting the financial burden (and commensurate welfare loss) of the health care system from taxpayers to users. We also show that user fees are regressive both in terms of access and welfare. In essence, the health care financing dilemma for developing nations is that the improvement in allocative efficiency and cost recovery from user fees are accompanied by a redistribution of welfare from poorer to richer. A natural solution to this dilemma is to introduce user fee schedules that increase with ability to pay. This type of price discrimination may generate substantial revenues with minimum welfare loss, if administrative costs are contained.

FOOTNOTES

We gratefully acknowledge financial support from The Living Standards Unit of the World Bank and USAID, and note that the views expressed herein are those of the authors alone and not of the sponsoring organizations. We are also indebted to John Akin, Jim Brown, Angus Deaton, Avi Dor, Paul Glewwe, Charles Griffiths, Phil Musgrove, John Newman, Mead Over, Cesar Penaranda, T. Paul Schultz, Morton Stelcner, John Strauss, Pravin Trivedi, Jacques van der Gaag, Juan Fernando Vega, and the participants of seminars at Harvard, Johns Hopkins, SUNY at Stony Brook, Yale, and the World Bank for valuable comments.

[1] See de Feranti (1985) for a discussion of health care pricing methods in developing countries.

[2] Recently, the pros and cons of such proposals have been discussed in de Feranti (1985), and Jimenez (1986).

[3] Studies of the demand for health care in developing countries include Akin et. al. (1985 and 1986), Birdsall and Chuhan (1986), Heller (1983), and Mwabu (1987).

[4] The feasibility condition requires income to be at least as large as the price of the alternative. The constraining level of income depends on the length of time over which individuals are able to budget. For example, if capital markets are perfect, the budget period is the individual's lifetime and the constraining income the present value of lifetime income. On the other hand, if there are cash constraints, the budgeting period could be as short the interval in which the individual is paid. In section III.C, we propose a procedure which parameterizes the length of the budgeting period and allows it to be estimated.

[5] If, in fact, income is endogenous, there is a possibility of simultaneity bias. The simultaneity bias arises because an accident or illness may reduce income. We argue in section V that the bias is likely to have a downward impact on the estimated price and income effects, making them closer to zero. Hence, our estimated price elasticities should be lower bounds on the true elasticities.

[6] For example see Akin et al. (1985 and 1986), Mwabu (1986), and Birdsall and Chuhan (1986).

[7] An obvious extension to the semi-translog is to include interactions and squared terms for health and nonmonetary cost terms. The problem with this is that the health terms, as will be discussed in a moment, will be a function of variables whose coefficients necessarily vary by alternative. Hence, this extension would require a substantially larger parameter space. Since the major objective of this study is to analyze price elasticities, we require the most flexibility in the parameterization of the consumption term. In addition, this specification would violate the necessary conditions for the model to be consistent with utility maximization specified in Mcfadden (1981). This point is taken up further in footenote 9.

[8] See Deaton and Muelbauer (1980) for discussion of compensating variation and other welfare measures.

[9] In order for (12) to be exact, the marginal utility of income, λ, must be independent of alternative specific characteristics and price. See McFadden (1981) and Small and Rosen (1981) for more discussion on this point. Although λ is independent of quality, it is not independent of price. Specifically

$$\lambda = \left(\alpha_1 + \alpha_2 \ln(Y - P)\right)/(Y - P),$$

and

$$\partial\lambda/\partial P = \left(2\alpha_2(\ln(Y - P) - 1) - \alpha_1\right)/(Y - P)^2.$$

In most cases this term is likely to be small relative to λ, as the denominator is approximately income squared. Hence, λ is likely to be approximately constant across small differences is price. If indeed $\partial\lambda/\partial P$ is small, then each individual's average marginal utility of income over his/her alternatives is a good approximation of λ. Since this approximation is calculated for each individual, λ will vary greatly across individuals as there is substantial variation in income.

[10] A full description of the hedonic price and travel time methodologies and resulting estimates is provided in the Appendix.

[11] The estimated σ was 1.02 with a standard error of 0.86. The test statistic for the hypothesis that σ equals unity is 0.03 and is distributed student t. The critical value at the 0.05 level is 1.96. The test statistic for the null hypothesis that k=1 is 1.06 and is distributed $\chi^2(1)$. The corresponding critical value at the 0.05 level is 3.84. Our linearization of the log consumption term biases the estimate of k towards zero. However, the observed bias is minuscule when evaluated at the mean of the data.

[12] As discussed in footnote 9, the marginal utility of income, λ, is not constant across alternatives. Each individual's average over the three alternatives is a good approximation if the variation in λ across alternatives is small. In our simulations, the largest price difference across alternatives is 19 thousand soles. At the mean income level with a price of 1 thousand soles λ is 0.0111, and at a price of 20 thousand soles λ is 0.0115; a difference of 0.0004. This difference declines with income, implying that the goodness of the approximation increases with income. The approximation is poor only at very low levels of income.

APPENDIX

The hedonic private doctor price equation specifies the price of a single visit to be a function of the type of illness, age of the individual, and characteristics of the market. The market variables include the number of doctors, the number of hospital beds, the number of clinics, and the population of the district in which the individual lives. We correct for sample selection bias using a methodology derived in Dubin and McFadden (1982). This requires the estimation of a reduced form multinomial logit model of provider choice, from which a set of Dubin McFadden selection correction terms are constructed (predicted) for each individual. The predicted correction terms are included as regressors in the hedonic price regression. Separate models are estimated for Lima and the Sierra. The market variables are not included in the Lima regression as there is no variation. The estimated coefficients and t-statistics are presented in table A.

The hedonic travel time equations for private doctors, hospitals, and clinics specify the time it takes to travel to a provider to be a function of the market variables, the location of the individual, a dummy variable indicating whether the main road in the district is paved, and the Dubin McFadden selection correction terms. An additional problem arises because we only observe if the individual traveled more or less than an hour. The hedonic travel time equations were estimated as binary logits. Separate Lima and Sierra models were estimated for private doctors and hospitals, and, due to small sample sizes, a single pooled model was estimated for clinics. The estimated coefficients and t-statistics are also presented in table A.

REFERENCES

Acton, J. P., 1975, Nonmonetary factors in the demand for medical services: some empirical evidence, Journal of Political Economy 3, 595-614.

Akin, J. S., C. C. Griffiths, D. K. Guilkey, and B. M. Popkin, 1985, The demand for primary health care services in the third world (Roman and Allanhel, Totawa, New Jersey).

Akin, J. S., C. C. Griffiths, D. K. Guilkey, and B. M. Popkin, 1986, The demand for primary health care services in the Bicol region of the Philippines, Economic Development and Cultural Change 4, 755-782.

Birdsall, N. and P. Chuhan, 1986, Client choice of health care treatment in rural Mali, Mimeo, Department of Population and Nutrition, The World Bank, Washington, D.C.

Deaton, A. and J. Muellbauer, 1980, Economics and consumer behavior (Cambridge University Press, Cambridge, England).

de Ferranti, D., 1985, Paying for health services in developing countries: an overview, World Bank Staff Working Papers No. 721.

Dubin, J. and D. McFadden, 1984, An econometric analysis of residential electric appliance holdings and consumption, Econometrica 2, 345-362.

Gertler, P., L. Locay, and W. Sanderson, 1986, Are user fees regressive? The welfare implications of health care financing proposals in Peru, Stony Brook Working Paper No. 289.

Heller, P., 1983, A model of the demand for medical and health services in peninsular Malaysia, Social Science and Medicine, 267-284.

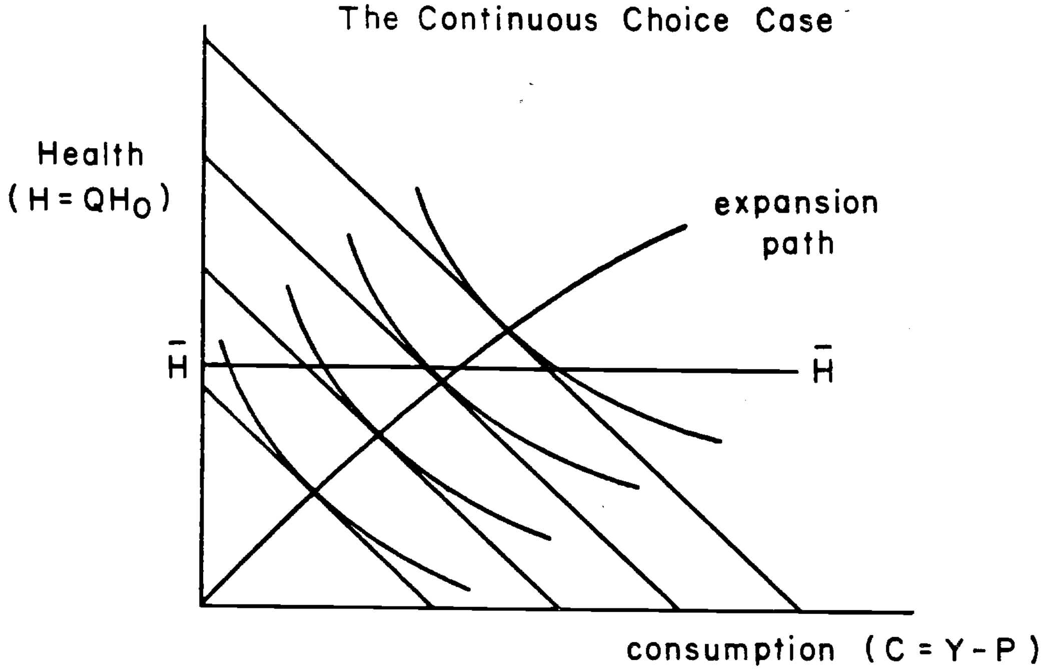
Jimenez, E., 1986, The public subsidization of education and health in developing countries: a review of equity and efficiency, The Research Observer, 111-129.

McFadden, D., 1981, Econometric models of probabilistic choice, in: C. Manski and D. McFadden, eds., Structural analysis of discrete data with econometric applications (MIT press, Cambridge, Mass.).

Mwabu, G., 1987, Conditional logit analysis of household choice of medical treatments in rural villages in Kenya, forthcoming in Social Science and Medicine.

Small, K. and H. Rosen, 1981, Applied welfare economics with discrete choice models, Econometrica 1, 105-130.

The Continuous Choice Case



The Discrete Choice Case

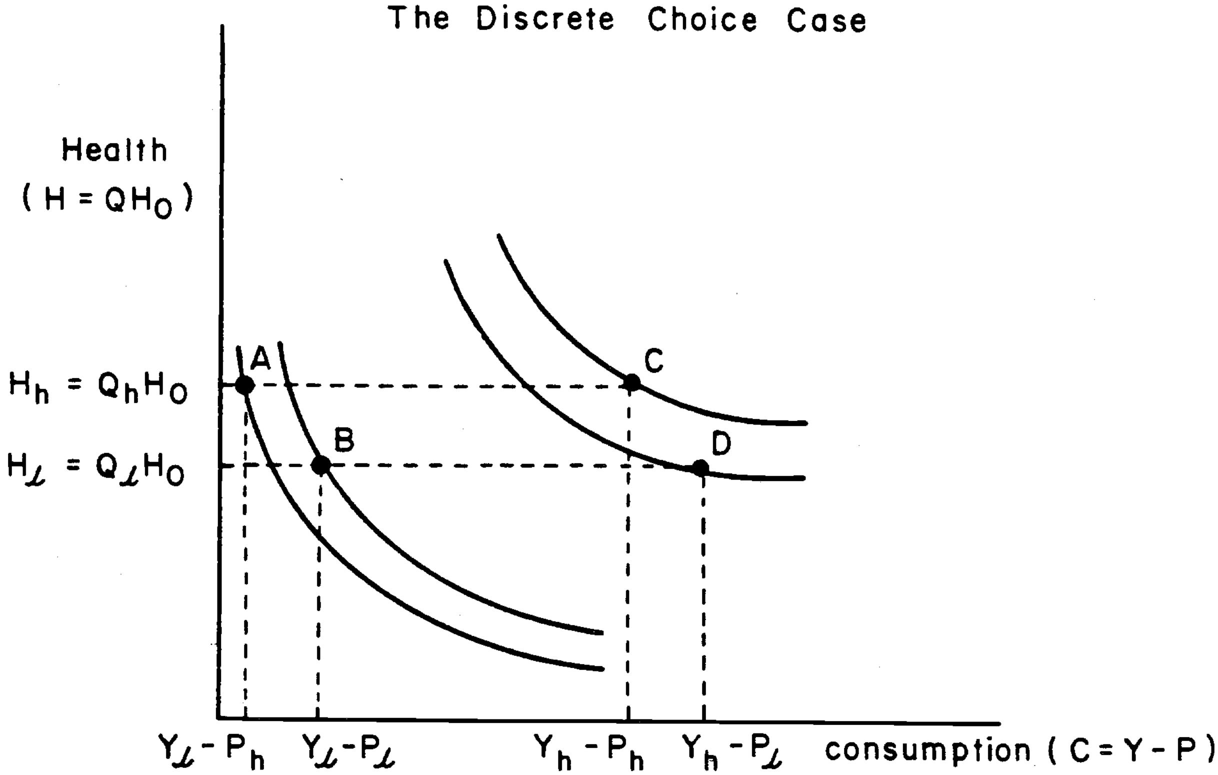


Table 1
Summary Statistics (N=3412)

Variable	Mean	Standard Deviation
Went to a public clinic (past 14 days)*	0.05	(0.22)
Went to a public hospital (past 14 days)*	0.11	(0.32)
Went to a private doctor (past 14 days)*	0.09	(0.29)
Age	39.18	(17.57)
Years of Education	7.73	(4.82)
Social Security*	0.15	(0.36)
Acute illness (past 14 days)*	0.05	(0.22)
Respiratory illness (past 14 days)*	0.15	(0.35)
Digestive illness (past 14 days)*	0.45	(0.50)
Resident of Lima*	0.37	(0.48)
Resident of South Cone*	0.10	(0.30)
Resident of North Cone*	0.22	(0.41)
Resident of South Sierra*	0.08	(0.27)
Resident of North Sierra*	0.15	(0.36)
Price of visit to private doctor**	19.01	(7.54)
Monthly income**	426.45	(1070.39)
Prob. travel time to clinic > 1 hour	0.01	(0.03)
Prob. travel time to hospital > 1 hour	0.13	(0.26)
Prob. travel time to private doctor > 1 hour	0.07	(0.14)

* Dummy variables (= 1 if answer is yes, = 0 otherwise)
** In 1,000's of April, 1984 soles.

Table 2
Multinomial Logit
Estimated Coefficients and t-Statistics

Variable	Hospital	Clinic	Private Doctor
Log Consumption*	-2.77 (1.81)	-2.77 (1.81)	-2.77 (1.81)
Log Consumption* Squared	0.62 (2.40)	0.62 (2.40)	0.62 (2.40)
Travel Time*	-2.05 (3.44)	-2.05 (3.44)	-2.05 (3.44)
Age	0.01 (4.35)	-0.01 (1.67)	0.01 (2.53)
Education	0.04 (2.77)	-0.05 (2.10)	0.05 (3.91)
Acute Illness	0.78 (3.87)	0.83 (2.77)	-0.29 (0.90)
Respiratory Illness	-0.64 (5.19)	-0.37 (2.13)	-0.74 (5.42)
Digestive Illness	0.09 (0.59)	0.32 (1.49)	-0.17 (0.95)
Lima	0.22 (1.11)	1.21 (2.71)	-0.10 (0.46)
South Cone	0.53 (1.62)	1.69 (3.58)	0.02 (0.08)
North Cone	0.36 (1.50)	1.31 (2.91)	-0.45 (1.94)
South Sierra	0.63 (2.33)	0.78 (1.51)	-0.05 (0.18)
North Sierra	-0.07 (0.26)	1.19 (2.52)	-0.00 (0.01)
Social Security	0.77 (5.55)	-	-
Constant	-2.72 (8.70)	-3.12 (5.71)	-1.99 (6.32)

* The coefficients are restricted to be equal across equations.

Table 3
Arc Price Elasticities by Income Quintile

User Fee Change*		Quintile 1 (lowest)	Quintile 2	Quintile 3	Quintile 4	Quintile 5 (highest)
Clinic	0-10	-.17	-.12	-.09	-.06	-.03
	10-20	-.62	-.42	-.23	-.15	-.09
	20-30	-1.43	-.58	-.38	-.26	-.14
Hospital	0-10	-.15	-.12	-.08	-.05	-.03
	10-20	-.57	-.34	-.23	-.15	-.09
	20-30	-1.52	-.56	-.39	-.26	-.13
Private Doctor	0-10	-.17	-.12	-.07	-.06	-.03
	10-20	-.53	-.35	-.21	-.14	-.08
	20-30	-1.36	-.60	-.35	-.25	-.12

* Reported in thousands of April, 1984 Soles

Table 4
User Fee Simulations - Aggregate Results

		No Revunue Reinvestment			With Revenue Reinvestment		
Scenario	User* Fee Change	Cum % Δ in Total Demand	Public** Revenue Increase	Welfare** Losses	Cum % Δ in Total Demand	Public** Revenue Increase	Welfare** Losses
No Private Doctor	1-10	-7.5	6,386	7,123	+0.5	7,006	-7,354
Price Response	1-20	-14.3	11,306	13,872	-7.3	13,686	569
Equal Pri. Doctor	1-10	-12.5	6,516	12,460	-4.4	7,756	-2,160
Price Response	1-20	-23.9	11,906	23,957	-16.6	14,126	10,407

* Reported in thousands of April, 1984 Soles.

** Reported in millions of April, 1984 Soles.

Table 5
User Fee Simulations - Distributional Results (No Revenue Reinvestment)
Percentage Change in Total Demand Accounted for by Each Income Quintile and
Consumers' Welfare Loss as a Percentage of Income by Income Quintile

Scenario	User* Fee Change	Quintile 1 (lowest)		Quintile 2		Quintile 3		Quintile 4		Quintile 5 (highest)	
		A*	B**	A	B	A	B	A	B	A	B
No Pri. Doctor Price Response	1-10	38.4	3.0	29.3	1.2	16.2	0.6	11.1	0.4	5.0	0.1
	1-20	37.6	6.2	26.5	2.3	17.5	1.2	12.1	0.7	6.3	0.2
Equal Pri. Doc. Price Response	1-10	39.2	6.1	25.3	1.9	16.9	1.0	12.0	0.6	6.6	0.2
	1-20	38.1	11.2	24.7	3.5	17.2	2.0	13.1	1.3	6.9	0.5

* A = Percentage Change in Total Demand Accounted for by Each Quintile.

** B = Consumers' Welfare Loss as a Percentage of Income by Quintile.

Table 6
User Fee Simulations - Distributional Results (With Revenue Reinvestment)
Percentage Change in Demand by Income Quintile and
Consumers' Welfare Loss as a Percentage of Income by Income Quintile

Scenario	User* Fee Change	Quintile 1 (lowest)		Quintile 2		Quintile 3		Quintile 4		Quintile 5 (highest)	
		A*	B**	A	B	A	B	A	B	A	B
No Pri. Doctor Price Response	1-10	-7.5	1.7	-2.9	0.4	2.3	-0.1	3.8	-0.3	6.1	-0.6
	1-20	-23.8	4.9	-14.1	1.7	-6.2	0.6	-0.8	0.1	3.8	-0.4
Equal Pri. Doc. Price Response	1-10	-18.7	4.5	-9.4	1.1	-2.8	0.3	1.0	-0.1	4.1	-0.4
	1-20	-44.8	10.0	-26.0	2.9	-13.8	1.3	-7.1	0.6	1.0	-0.2

* A = Percentage Change in Demand Within Each Quintile.

** B = Consumers' Welfare Loss as a Percentage of Income Within Each Quintile.

TABLE A – HEDONIC PRICE AND TRAVEL TIME REGRESSIONS

INDEPENDENT VARIABLE	PRIVATE DOCTOR PRICE LIMA	PRIVATE DOCTOR PRICE SIERRA	PRIVATE DOCTOR TRAVEL TIME LIMA	PRIVATE DOCTOR TRAVEL TIME SIERRA	HOSPITAL TRAVL TIME LIMA	HOSPITAL TRAVL TIME SIERRA	CLINIC TRAVEL TIME
CONSTANT	1.99 (3.50)	3.78 (6.51)	2.14 (1.79)	0.71 (0.39)	1.88 (2.28)	0.95 (0.49)	3.29 (2.05)
AGE	0.18 (0.94)	0.46 (2.48)	–	–	–	–	–
ACUTE ILLNESS	0.34 (0.73)	-0.54 (1.06)	–	–	–	–	–
RESPIRATORY ILLNESS	0.12 (1.02)	0.07 (0.50)	–	–	–	–	–
DIGESTIVE ILLNESS	-0.07 (0.41)	-0.23 (1.01)	–	–	–	–	–
NORTH CONE OF LIMA	-0.14 (0.87)	–	0.45 (0.97)	–	-0.34 (1.13)	–	1.99 (1.69)
SOUTH CONE OF LIMA	-0.14 (0.96)	–	-0.64 (1.52)	–	-0.91 (2.46)	–	-1.12 (1.47)
NORTH SIERRA	–	-0.72 (2.57)	–	-0.46 (0.72)	–	1.76 (1.31)	-0.33 (0.47)
SOUTH SIERRA	–	-0.25 (0.60)	–	-0.31 (0.29)	–	1.46 (1.37)	-0.16 (0.47)
# OF DOCTORS IN DISTRICT	–	-0.01 (1.25)	–	0.03 (1.68)	–	–	–
# OF HOSPITAL BEDS IN DIST.	–	-0.00 (3.40)	–	–	0.01 (1.98)	–	–
# OF CLINICS IN DISTRICT	–	-0.18 (3.66)	–	–	–	–	–
DISTRICT POPULATION	–	2.42 (1.46)	–	4.72 (2.11)	–	5.13 (2.51)	–
DISTRICT POP SQ'D	–	-2.72 (2.11)	–	-1.86 (1.45)	–	-2.88 (2.50)	–
GOOD ROAD DUMMY	–	–	–	1.30 (2.04)	–	0.66 (1.05)	0.86 (1.73)
HOSPITAL SELECTION TERM	-1.51 (1.99)	-0.64 (1.09)	2.25 (1.43)	1.91 (0.74)	–	–	3.80 (1.39)
CLINIC SELECTION TERM	2.21 (2.54)	1.99 (2.20)	-1.07 (0.63)	-2.21 (1.01)	-1.41 (1.09)	-5.13 (1.33)	–
PRIVATE DOC SELECTION TERM	–	–	–	–	2.07 (1.22)	-0.75 (0.22)	-4.57 (1.39)
SELF-CARE SELECTION TERM	-0.62 (1.04)	-0.80 (1.28)	-0.83 (0.57)	0.22 (0.12)	-0.36 (0.24)	5.42 (1.54)	0.88 (0.47)