Why do patients prefer hospital emergency visits? A nested multinomial logit analysis for patient-initiated contacts *

Jaume Puig-Junoy [a], Marc Saez [b] and Esther Martínez-García [a]

[a] *Universitat Pompeu Fabra, Department of Economics and Business, Research Centre for Health Economics, c/Ramon Trias Fargas, 25-27, E-08005 Barcelona, Spain*
E-mail: puig_jaume@econ.upf.es
[b] *Universitat de Girona, Department of Economics, and Research Centre for Health Economics at Universitat Pompeu Fabra, Spain*

Received December 1997; revised June 1998

This paper analyzes the nature of health care provider choice in the case of patient-initiated contacts, with special reference to a National Health Service setting, where monetary prices are zero and general practitioners act as gatekeepers to publicly financed specialized care. We focus our attention on the factors that may explain the continuously increasing use of hospital emergency visits as opposed to other provider alternatives. An extended version of a discrete choice model of demand for patient-initiated contacts is presented, allowing for individual and town residence size differences in perceived quality (preferences) between alternative providers and including travel and waiting time as non-monetary costs. Results of a nested multinomial logit model of provider choice are presented. Individual choice between alternatives considers, in a repeated nested structure, self-care, primary care, hospital and clinic emergency services. Welfare implications and income effects are analyzed by computing compensating variations, and by simulating the effects of user fees by levels of income. Results indicate that compensating variation per visit is higher than the direct marginal cost of emergency visits, and consequently, emergency visits do not appear as an inefficient alternative even for non-urgent conditions.

Keywords: health care demand, emergency visits, nested multinomial logit, compensating variation, time costs

1. Introduction

This paper analyzes the nature of health care provider choice that patients make from among a nested set of alternative providers, specifically restricting our attention to the first stage of the process, patient-initiated contacts. We seek to analyze the effects of individual and provider specific factors on the individual's choice. The impact of travel and waiting time and the perceived quality of each alternative provider are deemed of special interest from among the relevant potentially explanatory characteristics. Implications for public policy are considered.

Many empirical studies of demand for health care implicitly consider the patient as the only agent determining the demand for medical care, especially those in the tradition of Grossman's model. Nevertheless, many of them do not suitably separate the modelling of contact analysis and frequency analysis (see the discussion in [24]). Modelling patient contact decisions in a National Health Service (NHS) is a relevant issue for policy-making, in order to design incentive regulation tools for improving the economic efficiency of individual decisions.

Applied studies of demand for health care in developed countries where there are no monetary access prices have paid little attention to the extremely high and continuously increasing use of hospital emergency visits as an alternative choice to other health care providers (especially primary care). This fact is one of the most important distinctive characteristics of the Spanish health care system of recent years. Hospital emergency services are probably perceived by individuals as higher quality providers than primary care services and there are no access barriers, given the low satisfaction with primary care services reported by patients; moreover, the price the consumer pays at the time of purchase of medical care is the same in both cases (zero). This paper attempts to highlight the factors affecting behavioural decisions and why they probably deviate from social efficiency criteria, given the absence of incentives to consider social opportunity cost in individual choice decision.

Three empirical observations at the aggregate level illustrate and confirm the need to explain the behavioural changes in choice decisions in the last decade and their implications for the efficiency of the Spanish health care system. First, hospital emergency visits that resulted in immediate discharge were, in 1981, 64.4 per 1000 inhabitants; by 1991 this figure had increased to 296.8; that is, that type of hospital visit multiplied by 4.6 in ten years.

Second, the technologically sophisticated and input-demanding services of hospital emergency units are treating an increasing number of less complex and less severely ill patients. The probability of a hospital emergency visit resulting in immediate discharge (taken as a proxy of the average severity of patients treated), increased from 0.577

* Financial support from the Department of Health and Social Services of the autonomous government of Galicia, Spain, is gratefully acknowledged. E. Martínez also acknowledges financial support from DGICYT no. PB94-0848. We are grateful to Guillem López and Carles Murillos from CRES, and to two anonymous referees for helpful comments that have substantially improved the paper.

in 1981 to 0.841 in 1991. That is, there was an increase of 45.7% in ten years.[1]

Third, we can observe considerable regional variation in the choice of emergency hospital services from patient-initiated contacts, which calls for explanation. The per capita rate of hospital emergency visits resulting in immediate discharge ranged from 0.448 in Catalonia to 0.174 in Castilla-La Mancha in 1991. That is, the difference between the highest and the lowest regional use is higher than 2.6 times.

Micro-economic models of discrete choice random utility are appropriate for explaining individual choice from among a discrete number of alternatives, taking into consideration the characteristics of each alternative. By means of a nested multinomial logit model (NMLM) we analyze the elements that influence individuals' choice between the following provider alternatives in the Spanish health system: GP (public or private), emergency visits (hospital or clinic) and specialist. These provider alternatives differ in various characteristics, such as quality of care, intensity of technology, price and time spent, which will be analyzed below. Only patient-initiated contacts are considered, in order to reduce the effects of supply-induced demand; visits may be related to diagnosis and/or treatment.

Applied economics literature on the discrete nature of the decision to utilize a medical service (conditional probability of contact) has employed various model specifications. Specifications use dichotomous dependent variables: the negative binomial distributed hurdle model [24]; the probit model [19,29]; the multinomial logit model [22]. Specifications using a polychotomous dependent variable: the nested logit model [8,9,11]. Bolduc et al. [3] estimated three different discrete choice models of provider choice: a multinomial probit model, a multinomial independent probit model and a multinomial logit model. Conditional utility functions may be defined in the analysis for each alternative considered in the decision-making problem, and each presents a random component.

The statistical distribution of the random component determines whether the appropriate model is a probit, a logit or a nested model. If the vector of random components is independently drawn from a normal distribution, probit is the appropriate model. If it is independently drawn from an extreme value distribution, it is a logit model. Logit and probit models are based on the idea of a continuous threshold-crossing latent dependent variable with an observable counterpart. We restrict our attention to the NMLM, testing for non-correlation among the unobserved components of utility for alternatives within a nest (if there were correlation, the model would be reduced to the multinomial logit). A possible alternative statistical specification to the nested multinomial logit model (NMNL) could be a multinomial probit model (MP). Like the NMNL, the MP does not suffer from the independence of irrelevant alternative hypotheses. The MP does, however, involve the evaluation of a multi-fold normal integral (depending on the number of choices), making it extremely difficult to estimate using standard techniques, although there is Gauss code for the multinomial probit [3].

This paper contributes to the literature on health care demand in several ways. First, in the analysis of the elements that make individuals choose emergency services as what we believe to be a substitutive choice to primary care for non-severely ill patients. Secondly, in the use of the NMLM to explain contact decisions in a developed country. To date, literature on the NMLM of health care demand has been restricted to developing countries and has not accurately differentiated between patient- and physician-initiated contacts, which need to be modelled as two different stochastic processes. In addition, we introduce waiting time in the surgery as an explanatory variable of choice between alternatives. Moreover, we explore measures of the compensating variation associated with some hypothetical scenarios. Policy implications are obtained from the estimated income, and time elasticities enable the construction of hypotheses to explain the causes of the rapid increase in emergency services utilization and to predict the effects of different user fee scenarios.

This paper formulates an individual choice model for selecting a type of health care provider and applies it empirically to a cross section of about 2000 individuals. It finds that waiting time is important especially in the use of emergency services and that if user fees were to be introduced for health care provision there would be regressive effects.

The paper is organized as follows. The section below presents the discrete choice model of individual demand for health services and the empirical specification of the conditional utility function. Section 3 includes the features of the nested multinomial logit model. Section 4 describes the full and restricted alternative choice decision set and data, and includes the definition of the variables. Results are presented and discussed in section 5 and section 6. Section 7 concludes with some final remarks.

2. Analytical framework

In this paper we present an extended discrete choice model for the analysis of patient-initiated contact, along the lines of Gertler et al. [11] and Dor et al. [8]. Past studies analyzing health care have identified significant effects of time costs [1,5–7,25]. Our model considers the opportunity cost of travel time and also waiting time in the budget constraint in the same way as if they were monetary prices, as suggested by Acton [1]. Expected effectiveness and service quality of each alternative are modelled to depend on patient and provider characteristics.

[1] Hospital outpatient activity is consequently moving towards emergency services: in 1981 emergency visits resulting in immediate discharge accounted for 11.8% of total outpatient visits; in 1991 they accounted for 39.7%.

Assume that individual i in a given period faces J health care provider alternatives. For each alternative j, the individual's utility is given by the conditional utility function:

$$U_{i,j} = U(H_{i,j}, C_{i,j}), \tag{1}$$

where $H_{i,j}$ = expected health status of individual i after receiving care from provider j; $C_{i,j}$ = consumption of goods other than health care, when individual i chooses health care provider j.

A simple budget constraint is defined as

$$Y_i = C_{i,j} + \mathrm{TP}_{i,j}, \tag{2}$$

where Y_i = individual income, and $\mathrm{TP}_{i,j}$ is the total price of choosing provider j choice. The total price is formed by two components: monetary price and non-monetary price. Then,

$$Y_i = C_{i,j} + (P_j + T_{i,j}), \tag{3}$$

where P_j represents the monetary price of provider j (which is identical for all individuals; price discrimination is not allowed) and $T_{i,j}$ is the non-monetary price, which is measured as the opportunity cost of time devoted to travelling and waiting in the provider choice j. Let $\mathrm{TT}_{i,j}$ and $\mathrm{WT}_{i,j}$ represent travel time and waiting time associated with the choice of alternative j, and let w_i be the opportunity cost of time for individual i, then

$$T_{i,j} = w_i \bullet (\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j}). \tag{4}$$

Provider price affects the contact decision, as a different proportion of the individual's income remains available for consumption of other goods.

Expected health status after being treated by provider j is represented by two additive factors: the expected health status with alternative 0, $j = 0$ being the case of self-care in the absence of formal treatment by a health care provider; and the expected effectiveness of alternative j in relation to alternative $j = 0$. That is,

$$H_{i,j} = E_{i,j} + H_{i,0}, \tag{5}$$

where $E_{i,j}$ = expected effectiveness (or quality measure) of provider j, and $H_{i,0}$ = expected health status from the choice of provider 0. Then, expected effectiveness may be represented as a household production function which depends on patient and provider characteristics:

$$E_{i,j} = E(X_i, Z_j), \tag{6}$$

where X_i is a vector of individual patient characteristics whose effect varies between alternatives (effectiveness and service quality perceived by the individual), and Z_j is a vector of provider characteristics.

The conditional utility function may now be expressed by substituting (3), (4) and (5) into (1):

$$U_{i,j} = U(H_{i,0} + E_{i,j}, Y_i - P_j - w_i\mathrm{TT}_{i,j} - w_i\mathrm{WT}_{i,j}). \tag{7}$$

Then, U_i^* being the highest utility the individual may obtain, the unconditional utility maximization problem for individual i in period t takes the form

$$U_i^* = \max(U_{i,0}, U_{i,1}, U_{i,2}, \ldots, U_{i,J}). \tag{8}$$

2.1. Empirical specification

A linear utility function would be inconsistent with income-constrained utility maximizing behaviour [11]. We define a conditional utility function with a consumption second order term in order to avoid this problem. The coefficients on consumption terms are fixed for each individual, and independent of the provider alternative.[2] The conditional utility function is specified as follows:

$$\begin{aligned} U_{i,j} &= \alpha_0 H_{i,j} + \alpha_1 C_{i,j} + \alpha_2 C_{i,j}^2 \\ &= \alpha_0 H_{i,0} + \alpha_0 E_{i,j} + \alpha_1 Y_i \\ &\quad - \alpha_1\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big) + \alpha_2 Y_i^2 \\ &\quad + \alpha_2\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big)^2 \\ &\quad - 2\alpha_2 Y_i\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big) + \varepsilon_{i,j}, \end{aligned} \tag{9}$$

where $\varepsilon_{i,j}$ is a random taste shock uncorrelated between alternatives. For the self-care alternative ($j = 0$):

$$U_{i,0} = \alpha_0 H_{i,0} + \alpha_1 Y_i + \alpha_2 Y_i^2 + \varepsilon_{i,0}. \tag{10}$$

Notice that in (9), when $\alpha_1 \neq 0$ and $\alpha_2 = 0$, marginal utility of income is constant, and the probability of choosing provider j is not a function of Y. The term $\alpha_2 \cdot C_{i,j}^2$ incorporates income effects by allowing the marginal utility of income to be a function of the level of income.

The terms that appear in the conditional utility function for all alternatives may be ignored given that they cannot influence the choice of a consumer (given his income level). These terms are: $\alpha_0 \cdot H_{i,0}$, $\alpha_1 \cdot Y_i$ and $\alpha_2 \cdot Y_i^2$. Then, when the alternative chosen is formal care ($j \neq 0$), the empirical conditional utility function becomes

$$\begin{aligned} U_{i,j} &= \alpha_0 E_{i,j} - \alpha_1\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big) \\ &\quad + \alpha_2\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big)^2 \\ &\quad + 2\alpha_2 Y_i\big(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\big) + \varepsilon_{i,j}, \end{aligned} \tag{11}$$

and

$$U_{i,0} = \varepsilon_{i,j}.$$

[2] In our utility function specification, marginal utlity of income does not depend on the alternatives, but the consumption factor varies between alternatives. Some authors define a consumption coefficient which varies between alternatives (a recent example is Ellis et al. [9]). That is, it is assumed that $a_2 = 0$ and a_1 is replaced by $a_{ij} = 0, 1, \ldots, J$. In this last specification, marginal utility varies across alternatives; it implies that the marginal rate of substitution differs depending on the alternative chosen by the individual, i.e., it is equivalent to accepting that "*holding income, prices and health constant, the marginal rate of substitution varies by alternative*" [11, p. 73].

Expected provider effectiveness and quality of service (marginal utility of quality) is specified as the following household production function:

$$\alpha_0 E_{i,j} = \beta_{0,j} + \beta_1 Z_j + \beta_{2,j} X_i + \upsilon_{i,j}, \qquad (12)$$

where $\beta_{0,j}$ is a constant factor which represents individual expected effectiveness associated with provider j regardless of other specific individual and provider characteristics. β_1 is the coefficient on a vector of provider characteristics which affect their perceived effectiveness, the most important of these being the facilities, training of health care personnel and the input intensity of each alternative. β_{2j} is the coefficient on individual characteristics, which is allowed to vary between alternatives: perceived effectiveness may be different for each provider given individual characteristics such as age, sex, marital status, employment status, perceived health status, chronic illness, days of restricted activity, life style, human capital, etc. υ_j is a zero mean random disturbance variable with finite variance, and is uncorrelated across individuals. In the empirical specifications, coefficients β_0 and β_1 in the household production function (12) may be additionally allowed to vary with the size of the town of the patients in order to test the hypothesis that there are differences in preferences or perceived effectiveness of available alternative providers between patients living in small or big towns.

Substituting equation (12) into (11), the conditional utility function may be specified as

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j} + \upsilon_{i,j}, \qquad (13)$$

where the indirect utility function $V_{i,j}$ is given by

$$\begin{aligned} V_{i,j} = {} & \beta_{0,j} + \beta_1 Z_j + \beta_{2,j} X_i - \alpha_1\bigl(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\bigr) \\ & + \alpha_2\bigl(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\bigr)^2 \\ & - 2\alpha_2 Y_i\bigl(P_j + w_i(\mathrm{TT}_{i,j} + \mathrm{WT}_{i,j})\bigr). \qquad (14) \end{aligned}$$

In the self-care choice ($j = 0$), the indirect utility function is reduced to

$$U_{i,0} = V_{i,0} + \varepsilon_{i,0}, \quad \text{where } V_{i,0} = 0.$$

3. The nested multinomial logit model

The probability of the utility given for an alternative being greater than the utility from any other alternative could be seen as the demand function for that alternative. Consequently, a multinomial logit model of choice could be estimated. The problem is the assumption of independence of irrelevant alternatives that underlies the multinomial logit model [20]. This assumption states that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set.

If independence of irrelevant alternatives holds, the estimates obtained when applying the multinomial logit model to the full choice set, β_f, and those of a restricted set, β_r, should not be statistically different [14], that is,

$$H_0\colon\ \beta_r - \beta_f = 0,$$
$$(\beta_r - \beta_f)'(\mathrm{cov}_r - \mathrm{cov}_f)^{-1}(\beta_r - \beta_f)_{-}X^2_m, \qquad (15)$$

where m is the rank of the covariance matrix.

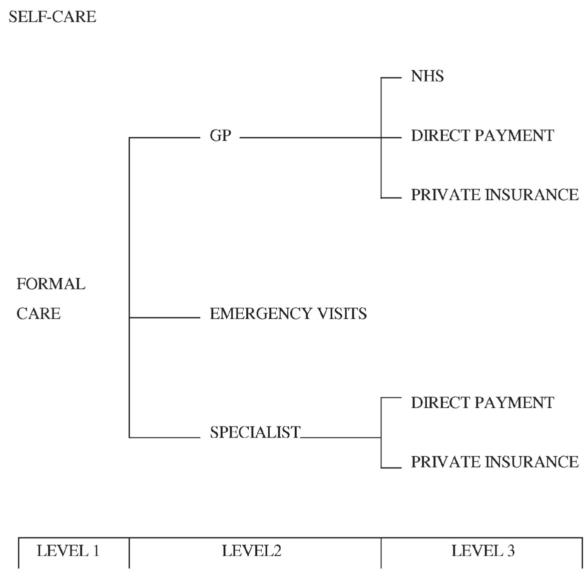

Figure 1. Choice decision set.

If the data do not support the assumption of independence of irrelevant alternatives, a nested multinomial logit (NMNL) should be estimated ([20], and also Dor et al. [8], Gertler et al. [11], Horowitz [15], Feldman et al. [10] and Ellis et al. [9]).

It is convenient to think of an NMNL as describing choices that are made sequentially according to a process that can be represented as a tree, such as the alternative choice decision set presented in figure 1.

First, individuals choose between self-care and formal care. The decision of self-care may or may not be accompanied by self-medication. When formal care is chosen, the individual faces three main alternative provider options: general practitioner (GP), emergency visits and specialized clinic services. In each provider decision, up to three different funding forms are possible according to the individual insurance scheme established previously to provider decision: National Health Service (public funding), private insurance (with or without copayment), and direct payment to provider.

Note that the individual decision does not sequentially imply a choice between alternative providers and funding service; in fact, not all provider alternatives are available when public funding is considered. In fact, the main choice being modelled is between the alternatives at level 3 of the tree. However, the choice process can be imagined to consist of first choosing an alternative at level 1 of the tree and then, conditional on this choice, an alternative at level 2; and finally, conditional on this choice, an alternative at level 3.

Let the utility of alternative k at level 3 be

$$U_k = V_k + \varepsilon_k, \tag{16}$$

where V_k denotes the deterministic component of utility and ε_k denotes the random component.

Let A_s denote the set of alternatives at level 3 that are connected by branches of the tree to alternatives at level 2 and at level 1.

The probability of alternative k at level 3 being chosen is

$$\mathrm{P}(k) = \mathrm{P}(k \mid A_s)\,\mathrm{P}(A_s), \tag{17}$$

where k is in C_s.

$$\mathrm{P}(k \mid A_s) = \mathrm{e}^{V_k/\rho_s}\left[\sum_{j\in A_s} \mathrm{e}^{V_j/\rho_s}\right]^{-1},$$

$$\mathrm{P}(A_s) = \mathrm{e}^{\rho_s I_s}\left[\sum_{r} \mathrm{e}^{\rho_j I_r}\right]^{-1},$$

$$I_s = \log \sum_{j\in A_s} \mathrm{e}^{\nu_j/\rho_s}.$$

I_s is called the inclusive price of alternative s at level 1. I_s indicates the average utility the patient can expect from alternatives within nest s. The key parameter is ρ_s. It could be interpreted as a measure of substitutability of alternatives across clusters. In order to guarantee the non-negativity of the density function that characterizes the NMNL, the parameter ρ_s should lie in the range [0,1] (the Daly–Zachary–McFadden condition [4]). Otherwise, the NMNL model may not be compatible with stochastic utility maximization (see Börsch-Supan [4] and Koning and Ridder [18]). When $\rho_s = 1$ the marginal effects do not depend on the location of alternatives, and therefore the NMNL reduces to a multinomial logit [20]. If, on the other hand, $\rho_s = 0$, each alternative should be regarded as a separate analytical unit or market [27]. If $\rho_s < 0$, the probabilities are inconsistent with utility maximization. Test of significance applied to the coefficient of the inclusive values can be used to test the independence of irrelevant alternatives property.

The NMNL could have been estimated by full-information maximum likelihood. Although this method is efficient it is also extremely burdensome computationally. As an alternative, we sequentially estimated the model in two stages with the inclusive values computed according to (17). The parameters that affect choice from among level S alternatives within level $S-1$ alternatives are estimated first. The idea is to apply standard logit estimation techniques to a data set in which each individual is assigned a choice set consisting of the level S alternatives contained within the level $S-1$ and level $S-1$ alternative the individual is observed to choose. Once the inclusive price of each alternative at the $S-1$ level was computed, we estimated the parameters that affect choice within level $S-2$ alternatives. This procedure is repeated up to the first level. This procedure yields consistent, asymptotically normal, although not asymptotically efficient, estimates of all the ρ_s coefficients and utility function parameters [15,20].

Besides the well-known log-likelihood ratio test of goodness-of-fit, two other specification tests were applied to the estimated models: Hausman and McFadden's test mentioned above and a specification test proposed by Horowitz [15]. The latter permits discrimination between any two nested models although their specifications are such that neither can be obtained as a parametric special case of the other. Nested logit models with different trees are examples of models that satisfy this requirement. A correctly specified model would have a larger log-likelihood than any other.

Consequently, if under the null model A is correctly specified,

$$\lim_{N\to\infty} \Pr\big[\big(\mathrm{L}^B - \mathrm{L}^A\big) > z\big] \leqslant F\big(-(2z)^{1/2}\big), \tag{18}$$

where L denotes the log-likelihood, B is another model, z is the standard normal variate, and F is the cumulative normal function.

In the empirical NMNL estimation of the model represented by equation (14), those parameters relating to variables that remain constant for all individuals (they do not vary between individuals given alternative j) cannot be estimated. This is the case of the vector β_1, whose effect is accumulated in the constant factor estimated for every provider alternative.

The estimated NMLM may be employed to calculate the *expected compensating* variation associated with various hypothetical scenarios. Individuals' maximum expected utility in period t is

$$V_i^* = V(X_i, Y_i, Z_j, P_j, T_{i,j}). \tag{19}$$

We define Z^0, P^0 and T^0 as actual vectors of alternative provider characteristics, provider prices and provider travel and waiting time, respectively, and Z^1, P^1 and T^1 are the same vectors after a specific policy is enacted. Then,

$$\begin{aligned} &V(X_i, Y_i, Z^0, P^0, T^0) \\ &\quad = V(X_i, Y_i + CV_i, Z^1, P^1, T^1), \end{aligned} \tag{20}$$

where CV_i in period t is the size of the budget change (positive or negative) which would restore the individual to the initial utility level from the after policy change level.

This framework makes it possible to calculate welfare effects (beside demand effects) resulting from changes in the actual available alternatives, such as variation in provider price or waiting time, or variations in provider facilities. It also permits evaluation of the welfare effects of more radical policies, e.g., a provider choice not being available. In the context of equation (11), compensating variation equals equivalent variation when $\alpha_2 = 0$.

In particular, and following Gertler et al. [11] and Kling and Thomson [17], the compensating variation could be computed as

$$CV = \mathrm{Log}\left[\sum_{s=1}^{S}\left(\sum_{j=1}^{J_s} \mathrm{e}^{V_{sj}^{1}/\rho_s}\right)^{\rho_s}\right] - \mathrm{Log}\left[\sum_{s=1}^{S}\left(\sum_{j=1}^{J_s} \mathrm{e}^{V_{sj}^{0}/\rho_s}\right)^{\rho_s}\right]. \qquad (21)$$

The measurement of welfare after a change in prices or income is based on the consumer's willingness to pay for the treatment modalities; welfare measurement therefore requires that patients are sufficiently informed about the relative effcetiveness of the treatment modalities (including self-care) so that their choice of treatment reflects what is best for them.

4. Data and variable definition

Our data were obtained from the results of the Spanish National Health Survey ("Encuesta Nacional de Salud") conducted in 1992. This survey includes a wide range of information on health conditions and health care utilization, as well as socio-economic data on non-institutionalized Spanish people. Because the sample is not representative for children, only individuals aged 16 or over were taken into account. Unfortunately, no data set available provides all the information required for the analysis, nor with the desirable characteristics. The data set used was one of the most adequate for our purposes, but it had some limitations that have to be taken into account when interpreting results, and imposed the need to introduce the following hypothesis and screening.

First, since we wanted to analyze the factors that influence first an individual's decision to seek care, and secondly what type of care provider to seek, both individuals who did not and those who did seek formal care were included in our sample. Second, our attention was focused basically on the decisions between emergency services provision and other types of provisions. Therefore, we excluded from the sample all contacts realised to obtain drug prescriptions (these are not provided by emergency room services). The same applies to preventive contacts, and hence all individuals who did not report any health problem within the two weeks previous to the interview were removed from the sample. Only treatment and/or diagnostic patient-initiated contacts were taken into account.

We also removed from the sample all those individuals who had an accident in the relevant period of analysis, since given the organization of health services in Spain, having an accident is a typical case in which an individual's choice of provider is more restricted: emergency visits are less a substitute for other alternatives.

The potential relevant set of choices has also to be restricted because of organizational characteristics of health services. The public (NHS) specialist alternative is irrelevant for the case of patient-initiated contacts, since GPs act as gatekeepers for those services. Also clinic emergency visits are not a usual and available alternative (and funding is primarily through the NHS). Hence, emergency services are restricted to hospital ones.

To choose the type of reported illness suitable to our analysis (once accidents are removed from the sample), we had to choose between different alternatives. Several indicators of health problems can be used. One which has been employed fairly often is having had to remain in bed for some days, as indicating the existence of some restriction in activity which may give cause to seek formal care. However, this measure is quite stringent for our analysis, since it can remove some causes of formal care contacts which may be relevant. Therefore, we chose another available indicator: having had any limitation in daily activity, which includes causes of restricted activity such as depression, diarrhoea, muscular pain, etc. Also, individuals having received more than one type of service during the period and those reporting chronic disorders have been excluded. After all these considerations, the sample consisted of 1959 individuals for whom the perceived health status was also known.

Notice our attention was on patient-initiated contacts, since the relevant issue of study was the choice of provider by individuals. Unfortunately, with the data set finally chosen, and once the screening had been applied as previously described to the initial data set, there were still difficulties to completely separate those individuals included in our data set who sought formal care in a patient-initiated contact from those referred by a professional. Therefore, some simplifying assumptions had to be made.

First, the decision to use services as emergency visits is always treated as a patient-initiated contact. Second, access to general practitioner's services is also considered to be the result of a patient-initiated decisions: this is clearly the case when the reported contact refers to the first visit of the clinical episode; otherwise, we assume that originally it was a patient-initiated contact. More problematic is whether to ascribe a contact with specialized clinic services to patient-initiated or referral decisions. Taking into consideration that in the NHS general practitioners act as gatekeepers for these services, NHS visits are excluded in this case; in contrast, we maintain privately financed contacts in the data set due to the direct access of patients to these services, in spite of the fact that some unknown number of them may reflect referred contacts. However, in this case, patient preferences constitute a crucial factor of choice, since there is, as already said, easy direct access for patients to this kind of providers.

4.1. The choice decision set

The choice decision set consists of a multiple nested decision set, in which the first decision is whether or not to seek formal care. Not to seek formal care is a possible alternative, and if not accounted for, the compensating variation estimation would be biased, since alternatives would be reduced to the formal care subset.[3] If formal care is chosen, then there are three nests between which the individual can choose: GP, emergency visits and specialist. The next subnests include alternative providers which differ basically in the financing mechanism of the provider: public provider (NHS) versus private provider (by direct payment or through private insurance schemes). The public provider is the NHS, which is financed mainly out of general taxation and which covers most of the population (98% in 1989).

4.2. Variable definition

The vector $\boldsymbol{X}$ of individual characteristics is given by the following variables (table 1 shows the labels and the description of the variables selected for the analysis): age, sex, smoking habits, physical exercise, town of residence size, education, perceived health status, and chronic disorders.

Age (a continuous variable) proxies the depreciation of health capital [13], as well as individuals' preferences towards health care. Another individual-related characteristic which affects health capital depreciation is sex (a categorical variable), and several studies have found that demand for health services differs according to sex (e.g., [2]). Also, the "level of physical stress" can affect the health capital rate of depreciation; we included the following variables to take account of it: being a smoker (yes, no), and "level of physical effort" (low, medium, high). Education may affect preferences, and the characteristics of each provider knowledge level (as well as own health time productivity). Four levels of education were distinguished: none or able to read and write, primary, secondary, further. Finally, the last individual characteristic to be included that can affect expected utility of providers was health status. Two measures of individual health status were taken into account: perceived health status (bad and very bad, fair, good and very good), and the number of chronic disorders.

Travel time to the provider and waiting time between arriving at the supplier office and being treated were obtained by direct response of individuals.[4] For those alternatives not chosen by the individual, we computed the average time of those who did choose them, controlling for possible differences due to the autonomous community of residence, patient income and the size of the town of residence (which may condition, among other factors, the availability of public transport and the distance between home and the provider's office). Following Gertler et al. [11], we estimated travel and waiting time using different sub-samples of individuals seeking care at each different provider.

[3] See E.R. Morey et al. [21].

[4] The time spent between booking an appointment in primary care services and the next available appointment can not be considered to be a factor influencing the choice of alternative providers, such as emergency room visits, because a physician visit in primary care can usually be obtained for the same day for the first contact.

Table 1
Variable description.

Variable	Description
$X_1(xx-yy)$	Age (from age xx to age yy)
X_2	Smoker (1 = yes; 0 = no)
$X_3(\cdot)$	Physical exercise (1 = low; 2 = medium; 3 = high)
X_4	Sex (0 = female; 1 = male)
$X_5(\cdot)$	Perceived health status (1 = poor; 2 = regular; 3 = good)
X_6	Number of chronic disorders
$X_7(\cdot)$	Education (1 = none; 2 = primary; 3 = intermediate; 4 = high)
$X_8(\cdot)$	Town of residence size (1 = <2000 inhab.; 2=2001–10000; 3=10001–50000; 4=50001–100000; 5=100001–400000; 6=400001–1000000; 7 = >1000000)
w	Time opportunity costs
TT	Travel time
WT	Waiting time
Y	Individual per hour income

Income is a non-observable variable; the survey only included social status (a categorical variable), proxied by employment status and level of studies. Since a categorical variable is not suitable in our model, income level was proxied by computing disposable income (after taxes) at each social status from the Spanish Family Expenditure Survey. Possible differences due to autonomous community and size of town of residence were taken into account. Family income was used, since it seems more relevant than individual income as a determinant of demand for health care.[5]

Prices payed by the individual at the point of consumption of formal care are as follows: all NHS alternatives are free of charge; all private insurance alternatives can also be considered free of charge at the moment of consumption, since there is evidence from other sources [12] that most of them are of this type;[6] for the remaining alternatives, market prices were taken as those average prices recommended by the physician's union. For those who sought care, price data were only available for the alternative they chose.

Finally, opportunity cost of time was calculated as income per hour, and taken from the contemporaneous Spanish Family Expenditure Survey. To take into account differences in the value of opportunity costs of time among individuals, income level was adjusted according to the employment status of individuals, distinguishing between those who work and those who do not, and whether they are retired, unemployed, students, or housewives. For individuals who were working, unit opportunity costs is taken

[5] It was not possible to compute equivalent income, since the information needed to do so was not compatible between the two sources of data.

[6] It is assumed that all private insurance packages held by the patients have no cost sharing provisions. In González [12] it is obtained that most private insurances are of the type of a restricted list of providers, which the insured can access by previous payment of a premium, and free of charge at the moment of consumption. Therefore, the problem of measurement error due to non-zero price at the moment of consumption is reduced.

to be per hour wage. For those not working, theory usually looks for the best alternative to leisure time, to put a value on it. For those who are non-voluntarily out of work (i.e., unemployed), time would be valued between zero and income which could be obtained if working. This income was assumed to be equal to that of those working, matched by gender, age and marital status. For those voluntarily not working (let us assume that this is the case for housewives and students), leisure would be valued as at least equal to the otherwise obtainable income. However, those who are not voluntarily working have more flexibility in scheduling care and fewer constraints on time, which may affect the demand for health services (see [25,28]). It can be considered that this time availability effect is allowed for by reducing opportunity costs for voluntary non-working individuals; as is quite common in these types of studies, we have considered opportunity cost of time to be one third of per hour wage.

Unfortunately, empirical measures of the vector of provider characteristics which affect their perceived effectiveness are not available. This is an unavoidable feature of the available data which represents a potential for omitted variable bias.

5. Results

The parameters of an NMLM estimated in the two-stage method described above are presented in tables 2–5 for the three decision levels previously specified. Note that at all the decision levels the log-likelihood ratio test of goodness-of-fit of the estimated model was statistically significant. Note that all standard errors were relatively high. This could be a consequence of the presence of a high level of multicolinearity in our estimated models. Therefore, since efficiency is not guaranteed in the estimation process, it is very likely that those coefficients with a t-ratio (or equivalently a Wald test) greater than one could actually be statistically significant [16].

The estimated values of ρ for each decision level are significantly less than one and significantly greater than

Table 2
Nested multinomial logit parameter estimates of level 1: formal care.

Variable	Coefficient	Standard error	T-statistic
Constant	8.4140	1.878	4.480
Y^{a}	0.081306	0.0325	2.495
Y squared[a]	0.000056	0.0000	2.303
X_1(30–48)	−0.03213	0.1412	−0.228
X_1(48–64)	0.06184	0.1565	0.395
X_1(> 64)	−0.05695	0.1767	−0.322
X_2	−0.20655	0.1167	−1.771
X_3(low)	−0.38009	0.2718	−1.398
X_3(medium)	−0.65258	0.3088	−2.113
X_4	−0.09851	0.1083	−0.909
X_5(bad)	0.12524	0.0710	1.764
X_5(regular)	0.07779	0.1154	0.674
X_6	0.06797	0.0556	1.221
X_7(none)	0.37777	0.2245	1.683
X_7(primary·)	0.22040	0.1942	1.135
X_7(intermediate)	0.02813	0.2004	0.140
X_8(2)	−0.19748	0.1993	−0.991
X_8(3)	0.02775	0.1895	0.146
X_8(4)	−0.49610	0.2580	−1.923
X_8(5)	−0.27396	0.1859	−1.474
X_8(6)	−0.35817	0.2644	−1.355
X_8(7)	−0.05514	0.2311	−0.239
Log-likelihood	−1250.464		
Restricted log-lik	−1331.998		
Chi squared	163.0679		
Significance level	0.0000		

[a] Coefficients are restricted to be equal among alternatives. These values are omitted in the following tables.

Table 3
Nested multinomial logit parameter estimates of level 2: GP, emergency visits and specialist.

Variable	GP	Emergency visits	Specialist
Constant	−2.9106	−36.774	32.819
	(−0.680)	(−0.103)	(2.671)
X_1(30–48)	−0.04041	0.46057	0.15989
	(−0.263)	(0.623)	(0.388)
X_1(48–64)	0.07023	−0.91073	−0.01483
	(0.418)	(−0.893)	(−0.029)
X_1(> 64)	−0.06704	−0.13326	0.51676
	(−0.353)	(−0.141)	(0.918)
X_2	−0.25196	−0.73835	0.41698
	(−1.989)	(−1.136)	(1.191)
X_3(low)	−0.40905	−0.84860	0.19076
	(−1.385)	(−0.872)	(0.178)
X_3(medium)	−0.71619	−13.462	0.44794
	(−2.132)	(−0.038)	(0.390)
X_4	−0.11257	1.1545	−0.36968
	(−0.962)	(1.976)	(−1.021)
X_5(bad)	0.23056	0.78904	0.36072
	(1.146)	(0.950)	(0.719)
X_5(regular)	0.11290	0.62525	0.00218
	(0.906)	(0.941)	(0.006)
X_6	0.05804	0.12209	−0.03934
	(0.975)	(0.430)	(−0.177)
X_7(none)	0.45203	13.525	−2.2111
	(1.853)	(0.038)	(−2.616)
X_7(primary·)	0.28140	12.909	−0.66967
	(1.320)	(0.036)	(−1.447)
X_7(intermediate)	0.01938	13.130	0.71855
	(0.088)	(0.037)	(−1.433)
X_8(2)	−0.40752	−12.101	0.23523
	(−1.863)	(−0.048)	(0.249)
X_8(3)	−0.17550	0.27939	0.61500
	(−0.843)	(0.247)	(0.710)
X_8(4)	−0.66870	0.46421	0.23429
	(−2.338)	(0.314)	(0.218)
X_8(5)	−0.50085	0.54470	0.85674
	(−2.423)	(0.499)	(1.042)
X_8(6)	−0.90661	2.4159	1.6274
	(−3.083)	(1.656)	(1.687)
X_8(7)	−0.43550	1.9907	0.36679
	(−1.634)	(1.354)	(0.384)
Log-likelihood	−1356.891		
Restricted log-lik	−1614.693		
Chi squared	515.6041		
Significance level	0.0000		

Note: parentheses indicate t-statistics.

Table 4
Nested multinomial logit parameter estimates of level 3: alternative choice between general practitioners.

Variable	NHS	Direct payment	Private insurance
Constant	16.930	0.74883	16.170
	(8.707)	(0.000)	(3.851)
X_1(30–48)	−0.12200	−1.2134	0.29443
	(−0.615)	(−1.615)	(0.489)
X_1(48–64)	0.26884	0.20089	0.19280
	(1.271)	(0.312)	(0.243)
X_1(> 64)	0.01371	0.50026	1.1777
	(0.057)	(0.683)	(1.413)
X_2	−0.24061	−0.32771	−0.45543
	(−1.512)	(−0.642)	(−0.855)
X_3(low)	−0.43047	18.450	−1.9643
	(−1.221)	(0.001)	(−1.474)
X_3(medium)	−0.63131	0.70286	−2.3982
	(−1.563)	(0.000)	(−1.410)
X_4	−0.01613	−0.06588	−2.0113
	(−0.110)	(−0.139)	(−3.231)
X_5(bad)	0.76080	−0.63935	−4.3433
	(2.296)	(−0.661)	(−4.178)
X_5(regular)	0.57980	−0.17410	−3.1094
	(1.961)	(−0.219)	(−4.815)
X_6	0.34695	0.27236	−3.0162
	(1.139)	(0.347)	(−3.896)
X_7(none)	0.16749	−0.43619	0.79052
	(0.832)	(−0.603)	(1.171)
X_7(primary·)	0.22152	−0.29688	−0.18490
	(1.424)	(−0.568)	(−0.328)
X_7(intermediate)	0.13917	−0.12957	−0.00280
	(1.941)	(−0.497)	(−0.010)
X_8(2)	−0.43540	0.17465	−1.6306
	(−1.578)	(0.152)	(−1.155)
X_8(3)	−0.08673	0.35645	−0.02164
	(−0.334)	(0.322)	(−0.018)
X_8(4)	−0.87030	−0.10575	−1.7592
	(−2.415)	(−0.078)	(−1.140)
X_8(5)	−0.48184	0.37497	0.13382
	(−1.871)	(0.335)	(0.114)
X_8(6)	−0.09826	0.86123	−1.6889
	(0.272)	(0.713)	(−1.111)
X_8(7)	−0.59214	−0.48743	−0.78922
	(−1.766)	(−0.357)	(−0.558)
Log-likelihood	−960.0286		
Restricted log-lik	−1534.451		
Chi squared	1148.844		
Significance level	0.000		

Note: parentheses indicate t-statistics.

Table 5
Nested multinomial logit parameter estimates of level 3: alternative choice between specialists.

Variable	Direct payment	Private insurance
Constant	0.23912	−10.921
	(0.001)	(−0.020)
X_1(30–48)	−0.11727	0.72046
	(−0.240)	(0.826)
X_1(48–64)	0.24805	0.04953
	(0.424)	(0.044)
X_1(> 64)	0.73816	0.91821
	(1.124)	(0.822)
X_2	0.71809	0.44572
	(1.766)	(0.661)
X_3(low)	−0.51304	11.820
	(−0.472)	(0.022)
X_3(medium)	−0.58689	13.093
	(−0.463)	(0.024)
X_4	−0.41720	−0.12181
	(−0.975)	(−0.177)
X_5(bad)	−0.41928	1.3544
	(−0.639)	(1.587)
X_5(regular)	−0.08111	0.59603
	(−0.190)	(0.777)
X_6	−0.26503	0.18806
	(−0.904)	(0.558)
X_7(none)	−1.6111	−12.962
	(−1.963)	(−0.058)
X_7(primary·)	−0.72129	−0.61591
	(−1.297)	(−0.659)
X_7(intermediate)	−0.27992	−1.5264
	(−0.508)	(−1.183)
X_8(2)	11.539	−0.01153
	(0.031)	(−0.010)
X_8(3)	12.639	−1.5401
	(0.034)	(−1.060)
X_8(4)	12.252	0.02318
	(0.033)	(0.016)
X_8(5)	12.985	−0.47454
	(0.034)	(−0.433)
X_8(6)	13.211	0.18273
	(0.035)	(0.122)
X_8(7)	12.225	−0.73964
	(0.032)	(−0.552)
Log-likelihood	−194.7302	
Restricted log-lik	−269.2772	
Chi squared	149.0940	
Significance level	0.0000	

Note: parentheses indicate t-statistics.

zero at the 1% level.[7] This confirms that the NMLM is consistent with the utility maximization hypothesis and the multinomial logit model may not be suitable in this case (a null hypothesis regarding independence of irrelevant alternatives is rejected). As expected, the results of the Hausman and McFadden tests suggest that the parameters of a full multinomial logit model and those of the restricted choice sets were statistically different.[8] Following Horowitz [15], we also tried other nested logit models with different trees. In all cases our original nested model had a larger log-likelihood than any other and, therefore, the former was preferred.[9]

[7] The obtained coefficients are as follows (standard errors in brackets):

Level 1	Formal care	0.28940 (0.0914)
Level 2	General practitioner	0.52752 (0.1890)
	Specialist	0.59258 (0.1147)

[8] Results of the Hausman and McFadden tests:

– Multinomial logit (full) versus GP (level 3.1)	54.5*.
– Multinomial logit (full) versus specialist (level 3.2)	35.7*.
– Multinomial logit (full) versus formal care (level 2)	45.8*.

(*) $p < 0.001$.

[9]

Log-likelihood estimated nested model	−2716.60.
Log-likelihood nested 2 levels	−2607.36.
Difference	109.24*.

(*) $p < 0.001$.

Table 6
ARC travel and waiting time elasticities by social status group.

Socio-economic status	Travel and waiting time range (hours)			
	0.00–0.25	0.25–0.50	0.50–0.75	0.75–1.0
General practitioners				
High	0.000000	0.000000	0.000000	0.000000
Medium	0.000000	0.000000	0.000000	0.000000
Low	0.000000	0.000000	−0.000003	−0.000155
Emergency visits				
High	0.000000	0.000000	−0.000133	−0.517657
Medium	−0.000003	−0.000751	−0.098267	−3.369974
Low	−2.946344	−3.533191	−4.034733	−5.196195
Clinic specialists				
High	0.000000	0.000000	−0.000031	−0.121004
Medium	0.000000	−0.000175	−0.022970	−0.787748
Low	−0.389474	−0.825389	−0.943138	−1.214560

The coefficients α_1 and α_2 on the individual income and individual income squared, respectively, are both positive and significantly different from zero ($p < 0.001$).[10] As previously defined, income variables refer to consumption other than health care after health care provider decision. This implies that the effect of travel and waiting time is reflected in the model via these terms.[11] Consumption varies between alternatives because travel and waiting time differ. Income and monetary and non-monetary costs are an important determinant of provider choice in the demand for medical care. The influence of the effect of these variables is explored by the analysis of time elasticities of the demand for general practitioners, emergency visits and specialists. Table 6 presents travel and waiting time elasticities calculated in the range of zero to two hours for each social status group.

The results in table 6 show differences in the time price elasticity for each social status group, holding income constant (by rows). At the same time, in this table we present the change in the time price elasticity as income rises, holding travel time constant, in order to better assess the influence of non-monetary price and income on the demand for medical care (travel and waiting time and income enter the demand functions in a highly nonlinear fashion). The arc travel and waiting time elasticities calculated are defined as the total percentage change in the demand for the alternative with respect to a change of one percent in total time cost. The elasticities are calculated for fifteen minutes to one hour. In the range of zero to one hour, general practitioner demand is very insensitive to travel and waiting time. That is, patients consider that up to one hour spent getting to general practitioner's services is not a reason to change their demand for medical care to an alternative provider. For most individuals, waiting time spent when general practitioner is the provider choice is greater than one hour. Our results show that demand for emergency services is vastly more elastic than demand for specialist services and for general practitioner's. Also, our results show that demand is much more sensitive to price for the lowest income group of patients than for the higher income group, which is in line with the pattern found by previous studies [8,11].

These results imply that the absence of fee payment in the access to emergency services does not preclude the existence of differences in the opportunity of access to those services. The elasticity of demand for low income patients is increasingly high as time increases. This trend was also observed by Dor et al. [8], who examined clinic and hospital ARC travel time elasticities by income quartile. Demand by low and middle income groups for emergency services is highly sensitive to travel and waiting time. As waiting and travel time decreases, the higher the demand increase is for these income groups. In fact, we observe that emergency services demand is very sensitive to time. Elasticities are higher for all income groups and for all time ranges for emergency services than for clinic specialist services.

These facts admit different interpretations, given the institutional context in which individual decisions are observed. We are inclined to consider that the great differences in time elasticities shown in table 6 between income groups not only reflect different individual responses to opportunity cost of time but also differences in the perceived quality or effectiveness of services. We may hypothesize that the data indicate the higher value that high income patients attach to hospital emergency services in comparison with general practitioners: they prefer to expend more time in accessing these hospital services, probably with a high subjectively attributed quality, than less time in accessing perceived less effective general practitioner. However, demand by high income groups for general practitioner services is not sensitive to time, which probably reflects the fact that the demand for this group is very low and/or the demand is in fact only sensitive when time is over one hour. When time cost for emergency services increases, individuals in the middle and low income groups decrease their demand very significantly, probably indicating a higher demand for general practitioner services. Results indicate that the effect of an increase in congestion costs of emergency services (higher waiting time imposed by an increase in utilization given emergency service capacity) may result in a greater utilization decrease by middle and lower income groups.

The estimated parameters of individual patient characteristics are for the most part consistent with expectations, given past literature and common sense. As was expected, health status (both perceived and the number of chronic disorders) plays an important role in individual decisions, both in terms of seeking formal care and the type of provider chosen. Even though the estimated parameters are not sig-

[10] It is assumed that income is an exogenous variable in the determination of health care demand.

[11] Average travelling time is greater for the NHS general practitioner option than for the emergency visit option: 0.23 hours for the general practitioner option. However, average waiting time in emergency visits is slightly less (0.55 hours) than the direct average waiting time for the general practitioner option (0.57 hours).

nificant in all cases, the general pattern observed is that worse health status increases the probability of contacting a formal care provider. Individuals with worse perceived health status are more likely to contact general practitioners under public financing, and less so private insurance and direct payment providers. Then the probability of using hospital emergency visits (and the quality attributed to the alternative) is greater for those individuals with regular or good perceived health status than for those with poor perceived health status and with more chronic disorders.

The probability of making use of formal care is significantly lower for males than for females, except for emergency visits. Differences in expected effectiveness perceived according to sex are also relevant between GP providers: the probability of private insurance financed contact is significantly lower for men. Age plays a minor role in determining the decision of which type of provider to choose. Even though the t-test for the estimates related to years of schooling are not significant in all cases, it can be said that the results obtained indicate that in general the lower the level of formal education, the higher the probability of seeking formal care, of making use of NHS GP services, and the lower the probability of choosing the specialist alternative versus emergency visits and GPs.

Town of residence size plays a significant role in the decision to choose between GPs and hospital emergency visits. In big cities expected effectiveness or quality of emergency visits is given a greater weight by individuals than in the case of general practitioners, once adjusted by other explanatory variables.

Health-related characteristics, such as being a smoker, or doing little physical exercise, do not increase the probability of seeking formal care but reduce it. The result obtained seems to be consistent with a Grossman style model in which individuals with a low demand for health would have a low demand for all health inputs.

To complement the analysis of results conducted above, we will now consider the results from the viewpoint of the level of decision (as opposed to that of the analysis of the effect of each explanatory variable). Health status (poor), having chronic disorders, and no formal schooling are the only significant variables and with positive effects on the probability of seeking formal care (some of them are clearly significant, others have t-test values greater than one). Being a smoker and doing little exercise, and in general, living in larger towns, are significant factors with negative effects on probability.

At decision level 2, it is worth noticing that being a smoker does seem to reduce the expected effectiveness (and, then, the probability of making use) of emergency departments, while males and those living in larger towns tend to expect a greater effectiveness of this provider alternative. Choosing the emergency department alternative does not seem to depend on the education level, age, having chronic disorders, or the physical exercise done.

6. Policy implications

Demand and welfare effects of user fees for hospital emergency visits. Cost sharing on emergency visits is a strategy that reduces the use of services, as observed in the Health Insurance Experiment [23]. However, more interestingly, in the Health Insurance Experiment cost sharing reduced the use of the emergency department more among patients with less severe diagnoses; that is, a selective effect on inappropriate visits was observed. Selby et al. [26] also report a decline of about 15 percent among members of an HMO when a small copayment for the use of emergency services was introduced. This study observed that the decline mostly affects patients with conditions considered likely not to present an emergency.

Demand and welfare effects of various scenarios for the access conditions of individuals to hospital emergency services financed by the NHS are considered. We explore the answer to two types of questions in order to evaluate economic factors influencing individual's decision to choose emergency visits.

Firstly, changes in demand and consumer welfare loss are observed when different user fees are imposed on the use of these services by individuals not having had an accident. A range of user fees from 1000 pesetas to 10000 is considered, which is equivalent to an increase in time (waiting and/or travel) monetary equivalent value. Our applied welfare analysis makes the assumption that the impact of financial user charges is equivalent to time costs. We focus our attention on the differential effects by socio-economic groups. Imposing user fees on emergency services may be interpreted as equivalent to consumers paying a two-part tariff composed of a fixed entry fee raised through general taxation and a marginal constant price when the consumer decides to use the service. Effects on consumer welfare of different user fee scenarios on patients who do not report an accident may be crucial in order to evaluate the optimal capacity of NHS emergency services when demand shows an accelerated growth with the monetary marginal price at the zero level. The marginal price may reflect the long-term marginal cost of the service and/or the marginal cost of congestion in terms of longer waiting time imposed on the other patients, given a defined level of capacity.

Secondly, we estimate the willingness to pay of each social status group of individuals, measured by means of the compensating variation. How much income could be taken away from the consumer so as to leave him or her indifferent when facing an emergency to the availability or unavailability of an alternative, in the light of his or her present tastes?

In table 7 we use the estimated demand function to simulate the effects of user fees on the demand and consumer welfare. Columns 2–5 in this table show the cumulative percentage change in demand in emergency services given different uniform fees being imposed at hospital facilities. Columns 6–9 show the percentage of welfare consumer loss expressed as percentage of individual per hour income.

Table 7
Demand and consumers' welfare effects of user fee simulations for hospital emergency services.

Socio-economic status	User fee simulation (Ptas.)							
	% Demand reduction				% Welfare loss[a]			
	1000	2000	5000	10000	1000	2000	5000	10000
High[b]	4.61	4.93	6.09	6.88	0.33	0.36	0.44	0.49
Medium	11.75	12.58	15.56	17.55	1.33	1.43	1.77	1.99
Low	25.30	27.10	33.50	37.80	8.07	8.58	10.69	11.96

[a] Consumer welfare loss expressed as percentage of individual per hour income.
[b] Wealthy and medium-high.

A user fee of 5000 pesetas, for example, generates a 6.09 percent reduction in demand of high income individuals, which is equivalent to a very low reduction in their welfare, in terms of individual per hour income. However, as observed previously in analyzing ARC time elasticities, the effects on the lower income groups are quite large and substantially higher than in the upper income ranges. Data in table 7 demonstrate that a fee of 5000 pesetas generates a reduction of 15.56% and of 33.50% in the demand of middle and low income individuals respectively. This reduction in demand produces a welfare loss equivalent to 10.69% of per hour income in lower income individuals. That is, not only is the reduction in total demand for emergency services concentrated in the lowest income groups, but the greatest relative welfare loss is borne by them. The simulations indicate that the introduction of user fees for emergency services are regressive in the sense that they generate a higher decrease in demand and welfare in poor individuals than in the case of rich individuals. This fact must not be regarded as an argument against the introduction of user fees, but a warning signal about the regressive effects they may generate. These regressive effects may be adequately compensated through the design of the tax system allowing for a tax reduction proportional to individual health expenditure in a progressive income tax.

We hypothesize a policy eliminating emergency services for non-urgent conditions (identified as those of individuals not reporting an accident and not resulting in a hospital admission), which is equivalent to raising the price of hospital emergency visits to infinity. The expected compensating variation is how much money you would have to give to the individual to make expected maximum utility after the policy equal to expected maximum utility in the original state. The estimated "per visit" compensating variation of this policy could be calculated for any individual as a function of exogenous variables, as in table 8.

This model predicts that the individuals in the sample would pay on average 4928 pesetas to avoid the elimination of the hospital emergency visit option for every visit. Estimated individual compensating variations vary from 4599 to 5533 pesetas, representing differences in characteristics of individuals. The estimated CV shows only small differences between individuals. In table 8 the mean CV for the five income quintiles confirms the homogeneity of individual willingness to pay. Also, a slightly increasing trend in CV is observed across higher levels of income.

Table 8
Compensating variation for emergency visits being unavailable.

Statistic	Compensating variation[a] (pesetas)
Mean	4928.3
Standard deviation	107.7
Range	4599.1–5533.3
First income quintile	4882.7 (102.0)
Second income quintile	4893.1 (110.0)
Third income quintile	4956.0 (101.9)
Fourth income quintile	4935.6 (112.7)
Fifth income quintile	4999.5 (94.9)

N=1959.
[a] Standard deviation in parentheses.

Our analysis suggests that CV for visits to emergency departments, including visits for minor medical problems, indicates that the value of an additional visit may clearly be higher than its direct marginal cost. Given the absence of reliable cost data, we may hypothesize that the regulated price paid for an emergency visit by the public sector represents the average cost. Williams [30] has observed that, for a sample of six community hospitals in Michigan, the marginal cost was 42 percent of the average cost. If this ratio holds for Spanish hospitals, the marginal cost of visits to emergency departments may be around 2300 pesetas.

Finally, own and cross time price elasticities are estimated separately for travel time and for waiting time of each alternative. The results are presented in table 9. Own travel time elasticities are statistically significant for the three alternatives considered: emergency visits, general practitioner visits and specialist visits. The travel time price elasticity of emergency visits is higher than that of general practitioners and specialists, indicating that the distance from the service is more important in this case, which is in agreement with common sense, given the presumed urgent demand. However, waiting time price elasticity of emergency visits is not significant, which may imply that patients consider that the time is very low in comparison with the complete waiting time of the alternatives, which involve consecutive visits to different health care services, and/or that this time is not important given the quality of the service. Waiting time price elasticity is higher for spe-

Table 9
Time price elasticities for each option.

Alternative	Travel time			Waiting time		
	Emergency visits	General practitioner visits	Specialist visits	Emergency visits	General practitioner visits	Specialist visits
Emergency visits	−0.618[a]	0.390[a]	0.072	−0.066	2.039[a]	0.084
General practtioner visits		−0.397[a]	0.493[a]		−0.394[a]	0.611[a]
Specialist visits			−0.288[a]			−0.537[a]

[a] $p < 0.05$.

cialist services than travel time price elasticity, and higher than for general practitioner visits. These results mean that if waiting time increases, then the demand for specialist visits (which are privately financed) decreases in a higher proportion than if general practitioner visits show the same increase in waiting time.

Cross time price elasticities yield an interesting result regarding the effect on emergency visit demand when the overall waiting time of general practitioner visits increases. Cross waiting time elasticity for emergency visits and general practitioner visits is 2.039. This result indicates that emergency service demand undergoes a considerable increase (decrease) in demand when primary services impose increasing (decreasing) cost on the patient in the form of prolonged waiting times. The sign of the cross time elasticities indicates that emergency services are substitutes for general practitioner and specialist visits.

7. Concluding remarks

We have derived a micro-economic model of discrete choice random utility to investigate individual choice among a discrete number of health care provider alternatives, taking into consideration the characteristics of each alternative, by means of a nested multinomial model. In addition to past literature on health care demand, we specifically analyzed patient-initiated contacts with an NMNL model introducing waiting time in the surgery as an explanatory variable of choice between alternatives.

The article has examined the elements that make individuals choose hospital emergency services as what we believe to be a substitute for primary care for non-severely ill patients. We focus our attention on the factors that may explain the continuously increasing use of hospital emergency visits as opposed to other provider alternatives. An extended version of a discrete choice model of demand for patient-initiated contacts is presented, allowing for individual and town of residence size differences in perceived quality (preferences) between alternative providers and including travel and waiting time as non-monetary costs. The results of a nested multinomial logit model of provider choice are presented. Individual choice between alternatives is considered, in a repeated nested structure: self-care, primary care, hospital and clinic emergency services.

The principal findings of this paper may be summarized in four main conclusions. First, our findings indicate that indirect access costs such as travel and waiting time play an important role in the health care provider choice when monetary prices are zero. The results confirm that the NMNL is consistent with the utility maximization hypothesis, the multinomial logit model being inadequate. The elasticity of demand for emergency services by low income patients is increasingly high as time increases and higher than for high income patients. Demand by low and middle income groups for emergency services is highly sensitive to travel and waiting time. We observe that demand for emergency services is very sensitive to time. Elasticities are higher for all income groups and all time ranges for emergency services than for clinic specialist services and general practitioner services.

Second, own and cross time price elasticities indicate that emergency visits are substitutes for general practitioner and specialist visits for patient-initiated contacts. We find that demand for emergency visits is highly elastic with respect to waiting time for general practitioner visits: a decrease of 10% in waiting time in the general practitioner alternative would produce a decrease of 20.4% in the demand for emergency visits.

Third, the simulations indicate that the introduction of user fees for emergency services may result in very regressive effects: the effects on the lower income groups are quite large and substantially higher than in the upper income ranges. These regressive effects may be adequately compensated through the design of a tax system allowing for a tax reduction proportional to individual health expenditure in a progressive income tax.

And fourth, individuals would pay on average approximately 5000 pesetas per emergency visit to avoid the loss of utility produced by a policy eliminating access to emergency services for non-urgent conditions. This result indicates that the compensating variation per visit is higher than the direct marginal cost of emergency visits (excluding congestion costs), and consequently, emergency visits do not appear to be an inefficient alternative even for non-urgent conditions.

This paper is motivated by the growth in the use of emergency visits and regional variation in the pattern of visits across Spain. In order to explore this issue we used a nested multinomial logit model to estimate patient initiated contacts with health care providers. However, lack of data make the individual choice model for selecting a type of health care provider more ambitious than the data: since there are no available variables reflecting characteristics of alternative providers, at the empirical stage this aspect of the utility function is just ignored. Also, quite strong assumptions had to be made with respect to time costs for those alternatives not chosen by the individual. However, these are quite common in the literature, and difficult to overcome.

References

[1] J.P. Acton, Non-monetary factors in the demand for medical services: Some empirical evidence, Journal of Political Economy (1975) 595–614.

[2] S. Birch, J. Eyles and B. Newbold, Equitable access to health care: methodological extensions to the analysis of physician utilization in Canada, Health Economics 2 (1993) 87–101.

[3] D. Bolduc, G. Lacroix and C. Muller, The choice of medical providers in rural Bénin: a comparison of discrete choice models, Journal of Health Economics 15 (1996) 477–498.

[4] A. Börsch-Supan, On the compatibility of nested logit models with utility maximization, Journal of Econometrics 43 (1990) 373–388.

[5] A.C. Cameron, P.K. Trivedi, F. Milne and J. Piggott, A microeconometric model of the demand for health care and health insurance in Australia, Review of Economic Studies 55(1) (1988) 85–106.

[6] S.D. Cauley, The time price of medical care, Review of Economics and Statistics 69(1) (1987) 59–66.

[7] A.D. Colle and M. Grossman, Determinants of paediatric care utilization, Journal of Human Resources 13 (1978) 115–158.

[8] A. Dor, P. Gertler and J. Van der Gaag, Non-price rationing and the choice of medical care providers in rural Cote d'Ivoire, Journal of Health Economics 6 (1987) 291–304.

[9] R.P. Ellis, K.D. McInnes and E.H. Stephenson, Inpatient and outpatient health care demand in Cairo, Egypt, Health Economics 3 (1994) 183–200.

[10] R. Feldman, M. Finch, B. Dowd and S. Cassou, The demand for employment-based health insurance plans, Journal of Human Resources 24(1) (1989) 115–143.

[11] P. Gertler, L. Locay and W. Sanderson, Are user fees regressive? The welfare implications of health care financing proposals in Peru, Journal of Econometrics 36 (1987) 67–88.

[12] Y. González, La demanda de seguros sanitarios, Revista de Economía Aplicada, III 8 (1995) 111–142.

[13] M. Grossman, On the concept of health capital and the demand for health, Journal of Political Economy 80 (1972) 223–255.

[14] J. Hausman and D. McFadden, A specification test for the multinomial logit model, Econometrica 52 (1984) 1219–1240.

[15] J.L. Horowitz, Specification tests of nested logit models, Environment and Planning A 19 (1987) 395–402.

[16] J. Johnston, *Métodos de Econometría* (Vicens Vives, Barcelona, 1987).

[17] C.L. Kling and C.J. Thomson, The implication of model specification for welfare estimation in nested logit models, American Journal of Agricultural Economics 78 (1996) 103–114.

[18] R.H. Koning and G. Ridder, On the compatibility of nested logit models with utility maximization, Journal of Econometrics 63 (1994) 389–396.

[19] W.G. Manning, J.P. Newhouse, N. Duan, E.B. Keeler, A. Leibowitz and M.S. Marquis, Health insurance and the demand for medical care: evidence from a randomized experiment, American Economic Review 77(3) (1987) 251–277.

[20] D. McFadden, Econometric models of probabilistic choice, in: *Structural Analysis of Discrete Data with Econometric Applications*, eds. C. Manski and D. McFadden (MIT Press, Cambridge, MA, 1981).

[21] E.R. Morey, R.D. Rowe and M. Watson, A repeated nested logit model of Atlantic salmon fishing, American Agricultural Economics 75 (1993) 578–592.

[22] G. Mwabu, M. Ainsworth and A. Nyamete, Quality of medical care and choice of medical treatment in Kenya, Journal of Human Resources 28(4) (1993) 838–862.

[23] K.F. O'Grady, W.G. Manning, J.P. Newhouse and R.H. Brook, The impact of cost sharing on emergency department use, New England Journal of Medicine 313 (1985) 484–490.

[24] W. Pohlmeier and V. Ulrich, An econometric model of the two-part decision-making process in the demand of health care, Journal of Human Resources 30(2) (1995) 339–361.

[25] J. Primoff and V. Hamilton, The time and monetary costs of outpatient care for children, American Economic Association, Papers and Proceedings, May 1995.

[26] J.V. Selby, B.H. Fireman and B.E. Swain, Effect of a copayment of the emergency department in a Health Maintenance Organization, New England Journal of Medicine 334 (1996) 635–641.

[27] M. Trajtenberg, The welfare analysis of product innovations, with an application to computed tomography scanners, Journal of Political Economy 97(2) (1989) 444–479.

[28] W.P.M. Van de Ven and J. Van der Gaag, Health as an observable: a MIMIC model of demand for health care, Journal of Health Economics 1 (1982) 157–183.

[29] G.J. Wedig, Health status and the demand for health, Journal of Health Economics 7(2) (1988) 151–163.

[30] R.M. Williams, The costs of visits to emergency departments, New England Journal of Medicine 334(10) (1996) 642–646.