A LOCATION MODEL APPLIED TO HEALTH CARE PLANNING

Roberto F. Iunes
Takemi Program in International Health
Harvard University
1995/1996

I would like to thank the W.K. Kellogg Foundation for its support during the Program.
# TABLE OF CONTENTS

I. INTRODUCTION ........................................................................................................................................2

II. THE FULL ACCESS CASE: HEALTH CARE PLANNING BASED ON NEED ....................6
   II.1 THE FIRST-BEST SCENARIO: FULL INFORMATION ..........................................................7
       II.1.1 The Basic Model .............................................................................................................7
       II.1.2 The Health Care Pyramid Introduced .........................................................................12
   II.2 THE SECOND-BEST SCENARIO: IMPERFECT INFORMATION ......................................19
       II.2.1 One-Level Misinformation .........................................................................................21
       II.2.2 The Referral System as a Mean to Minimize the Costs of Imperfect Information ....25
       II.2.3 Examining the Results ............................................................................................31
       II.2.4 Two-Level Misinformation .......................................................................................44
       II.2.5 Summary of Results and Conclusions ....................................................................50

III. CHOICE CONSIDERED: PLANNING BASED ON DEMAND ..............................................55
   III.1 THE BASIC DEMAND MODEL ....................................................................................56
       III.1.1 Some Equity Considerations ....................................................................................68

IV. A FINAL DISCUSSION AND POLICY IMPLICATIONS ..................................................71
I. INTRODUCTION

In less developed countries (LDCs), the lack of medical care facilities to serve the population leads to long lines, a large number of unattended patients, and unreported cases. Particularly in rural areas health facilities tend to be too dispersed and too distant for many consumers. In this sense a physical expansion of the health care system becomes not only desirable “per se”, i.e. for the increase in the number of sources of treatment, but also for its impact on reducing transportation costs to these services, in many cases the binding constraint to access the system.

Therefore, health care planners of developing countries aiming at the physical expansion of the medical care system try to maximize society’s welfare while facing a trade-off between the reduction in transportation costs to consumers and an increase in expenditures on the health care budget.

There have been two distinct ways of approaching the planning strategy: one is based on the patient’s individual structure of preferences and budgetary constraints; the other disregards completely these factors to focus on some technically defined medical needs of the population. The economists’ arguments in favor of the first, the demand approach, are based on the fact that consumers allocate their time and monetary endowments according to their own individual structure of preferences, which means that any conflicting distribution of resources is likely to be inefficient. These arguments, however, have been very pragmatically disregarded by most health professionals — who generally are the health planners of developing countries — because the need concept can be translated into norms that are not only very easy to use, but are also easily sustained on
ethical grounds in front of the public opinion. Moreover, they argue that the demand approach is inequitable.\(^1\)

There is a growing trend among economists, however, not to regard these two approaches as completely antagonist; moving away from the strong criticism of the past, economists are beginning to recognize the need concept as an element in the policy and decision-making process, particularly in developing countries. It is in this context that this paper develops its models. It analyzes the welfare implications of possible alternative health care systems if the concept of need is indeed the underlying framework upon which health care planners try to organize the health sector in developing countries.

This work also formally demonstrates some of the differences that emerge between the need and demand approaches, including equity, by contrasting the results of a simple demand-based health care system with those obtained from the need framework. The models developed in this paper were inspired by the framework of spatial competition developed by Hotelling (1929) and later expanded by Spence (1976), and in particular by Salop (1979).

The next section outlines the main features of the framework of analysis and develop the model of planning based on need. Subsection II.1 describes a world with perfect information: a first-best scenario. In the basic model of Subsection II.1.1 only one kind of treatment is considered, which we call outpatient care. A new type of medical service, called hospital care, is introduced in Subsection II.1.2. These two levels of services describe what is usually defined as a health care pyramid. The first-best scenario

\(^{1}\) For an analysis of these issues see Iunes (1996).
of Subsection II.1 assures that patients will always seek the right facility; therefore, these services will work independently from one another, as the similarities in the results of the two sub-sections clearly show.

Imperfect information is a classic topic in the health economics literature. Under imperfect agency it could lead to induced demand and conflict of interests. In the context of this paper, imperfect information means that patients are not always able to identify the proper type of treatment that they need, and, therefore, they may look for assistance in the wrong service. Subsection II.2 analyzes the implications of this second-best scenario. Its first topic examines the possibility of patients overestimating their health problem and therefore unnecessarily seeking care in more complex and expensive services.

As a way of minimizing some of the extra costs that exist in an imperfect setting, decision makers frequently establish a structure of medical care in which patients that require attention at higher levels of the pyramid have to be referred from a lower-level facility. Subsection II.2.2 describes this system. Subsection II.2.3 compares the outcomes developed in the previous sub-sections. It shows the conditions that would make a referral system an effective option from an economic standpoint.

Subsection II.2.4 examines the possibility that patients may not only overestimate their health problem, as done in II.2.1, but considers simultaneously, the fact that some individuals may not be fully aware of the seriousness of their condition. Patients may not only go to hospitals when they could be treated in clinics, but they also may seek primary care with serious problems, which implies that they will have to be referred to the more complex facilities. Subsection II.2.5 summarizes all the main results developed and presents the conclusions for the needs-based model.
Planning models based on the demand for health services take into consideration a patient’s willingness to pay for health care, that is, the total amount of money a person is willing to spend, in price plus transportation costs, to get medical care. Section III explicitly considers the issue of access to medical care through a price-elastic demand and the constraints imposed by a government health budget: Sub-section III.1 presents a simple demand-based model of health care planning, while Subsection III.2 analyses some of the equity implications of this framework.

Section IV concludes this paper.
II. THE FULL ACCESS CASE: HEALTH CARE PLANNING BASED ON NEED

The purpose of this section is to use the model of spatial competition to describe the main characteristics of health care systems that result from planning based on the concept of need. Examples of such systems can be found in many western-European countries, in particular the British National Health System (NHS), and also in many developing nations. These are government-financed health care systems in which the relevant information is the medical need of the patient, with economic variables such as prices usually excluded from the model.

This type of system “allows selective access according to the effectiveness of health care in improving health (‘need’). It seeks to improve the health of the population at large through a tax-financed system free at the point of service. It allows public ownership of the means of production subject to central control of budgets; it allows some physical direction of resources; and it allows the use of countervailing monopsony power to influence the rewards of the suppliers” (Culyer, Maynard and Williams, 1981 p. 134. Italics added). Under the set of hypotheses that constitute the need approach, health planners use the information provided by the epidemiological data to identify the main medical needs of the population and design a health care system that is able to satisfy such needs — or in other words, to define the “optimal” number of services and their distribution as to assure that each patient in need of medical can be taken care of. If the services are actually going to be used or not is not taken into consideration. In such context what matters is that each individual has the opportunity of seeking care open to him or her. It is in this sense that the system is regarded as providing full access.
II.1 The First-Best Scenario: Full Information

II.1.1 The Basic Model

The health care system described here is that of a single market served by government-owned firms delivering homogeneous care. In this subsection only one form of medical service, denominated as outpatient or ambulatory care, is considered. These services are provided in health centers or clinics which differ from one another only in their distance from the consumer.

In order to avoid the boundary problems found in the line model, where the end-firms don’t have competitors or consumers on one side, it is assumed that the $n$ equally spaced and identical health care centers are located around a circle of unit circumference. Based on the epidemiological evidence available, health care planners estimate that there are $J$ consumers that need medical attention and would have to be served in order to assure that the system does provide full access. These patients are evenly distributed around the circle.

In summary, the main assumptions that characterize this basic model are as follows:

Assumption 1: the space is defined by a circle of unit circumference;
Assumption 2: the health system provides only one type of service, called primary care;
Assumption 3: there are $n$ identical health care centers or clinics evenly distributed around the circle (therefore the distance between clinics is equal to $1/n$);

---

2 Primary care, ambulatory care and outpatient care are used as synonyms.
Assumption 4: the health care facilities provide only outpatient care;

Assumption 5: there are $J$ consumers, uniformly distributed around the entire circle, in need of medical care;

Assumption 6: the services are government-owned.

While the second assumption will be relaxed in the next subsection, the other five premises constitute general characteristics of the model and therefore will be kept throughout the analysis. Figure 1, below, displays the case of six clinics (i.e. $n = 6$). In the figure, “C” indicates the presence of a clinic.

![Figure 1](image.png)

Potentially, each individual may face two types of costs in order to be able to consume medical care: the monetary price to be paid (if any) and the transportation costs incurred to get to the center and return home. Accordingly, the consumer’s total expenditure is described by $e$:

\[
e = p + 2\alpha
\]
Here $p$ is the monetary price charged, $c$ is the unit cost of transportation and $x$ is the distance the consumer is from the center. As discussed above, the framework of planning based on need assumes zero prices. Finally, it should be noted that a person will seek medical care if $e \leq b$, with $b$ representing the monetary equivalent to the benefit that the patient will derive from the treatment.

The cost function for the $k^{th}$ health care center is expressed by equation (2):\(^4\)

\[
\Gamma_k = f + mq_k
\]

In the expression above, $f$ represents fixed or investment costs and $mq$ variable costs. With $q$ describing the quantity of health care services produced in that clinic, the marginal variable cost is constant and equal to $m$.

Since the distance between centers is $l/n$ (see Figure 2 below), the maximum that a person will have to travel to get to a center is $l/2n$. With the distribution of the population assumed to be uniform, the minimum travel distance is zero, and the average distance for all consumers is $l/4n$. The total transportation cost for this economy, considering round-trips to the clinics, is therefore given by:

\[
d \left( \frac{1}{4n} \right)^2 = \frac{Jc}{2n}
\]

It can be easily seen from the figure that an individual located over $L$ has to travel $l/2n$ to get to either clinic C2 or C3 (note that since all clinics are identical, the person

\(^3\) Which, as will be shown later, essentially provides results that are similar to the usual framework of spatial competition in which demand is price-elastic (e.g. Salop, 1979).

\(^4\) The cost function expressed in equation (2) follows the general structure found in the spatial economics literature (see, for instance, Greenhut et al., 1987).
would be indifferent between C2 or C3). A person living over C1, however, has a travel distance equal to zero to get ambulatory care.

![Diagram](image)

Figure 2

The discussion above shows that there are three types of costs imposed on society by this health care system: (i) transportation costs; (ii) operating (or medical) costs; and (iii) investment or fixed costs, which are summarized by the total social cost displayed in equation (3):

\[ TSC = \frac{Jc}{2n} + Jm + rf \]  

(3)

It should be noted that the model incorporates the health planner’s assumption that the system is to be conceived as to allow full access to all patients that need medical care.

In this context, the health planning objective is to define the number of clinics necessary to maximize the welfare of society. Since the return (benefit) obtained from outpatient care is equal to \( b \), the objective function becomes simply:

\[ \max. \quad W = TS - TSC = Jb - \frac{Jc}{2n} - Jm - rf \]  

(4)
One of the most criticized features of the need approach is the fact that the welfare maximization process is defined without constraints, in this framework tradeoffs are nonexistent or not considered during the planning process, which implies that the conception of need falls under an “on-off” type of rationale (see Williams, 1992).

Accordingly, the socially optimal number of health care centers can be directly derived from the first order condition of welfare maximization:

\[
\frac{\partial W}{\partial n} = -f + \frac{Jc}{2n^2} = 0
\]

Which defines \(n^*\), the (optimal) number of outpatient facilities necessary to assure care to all those that need medical attention:

\[
n^* = \sqrt{\frac{Jc}{2f}}
\]

The results expressed in (6) are intuitive: the optimal number of health care centers \((n^*)\) should increase with the number of cases and with transportation costs, and should be reduced if investment costs increase.

Even though this basic model is simpler than Salop’s (op. cit.), for it contains no prices, the result expressed in (6) is the same he derived for the monopoly market (see his expression (17) p. 147). This is because Salop assumes an inelastic demand for the differentiated product (conditional on a purchase). Since the concept of need implies, by definition, an inelastic demand for health care,\(^5\) even with non-zero prices the planning process based on need would still remain unaffected. This “robustness” of the model is in

\(^5\) See Iunes (op. cit.).
fact consistent with two attributes of prices that allow them to be excluded from a planning process based on need: (i) they are unrelated to health needs, and (ii) they affect all patients in the same way. The first topic is self-evident, but even though it may be a necessary condition for exclusion from this context of planning, it is not a sufficient one: transportation costs, which also bear no relation whatsoever to health needs, are present in the planning process due to the fact that individuals have different costs of transportation. This means that transportation costs affect equity, the core element of the needs framework. If, however, all consumers face the same monetary price for medical care, the relative slopes of their budget constraints would remain unchanged (attribute (ii) above), which, according to some definitions of access (e.g. Le Grand, 1982), (horizontal) equity would not be affected.

Finally, it must be seen that all clinics will assist the same number of patients. The market area for each center is given by \(2(1/2n) = 1/n\): each clinic serves both sides of half the arc that separate any two facilities, which means that the market share of each health care unit is equal to \(J/n\).

II.1.2 The Health Care Pyramid Introduced

Until now it has been assumed that outpatient or ambulatory care was the only form of medical treatment available. In this subsection the possibility of accessing another type of care will be explicitly considered.

A medical care system is frequently described by a multilevel “pyramid.” The base of the pyramid is constituted by the numerous health clinics delivering low-cost primary care and receiving the vast majority of the medical cases. Moving upward,
towards the top of the structure, the degree of specialization and sophistication of care increases, and the number of institutions providing services diminishes. The top of the pyramid is generally composed of the very specialized and complex tertiary or quaternary hospitals.

For the sake of clarity, the model considers only two levels of care. The base of this simplified pyramid comprises the outpatient care centers presented in the previous subsection. The top of the structure is formed by another type of facility, called hospitals, delivering more complex inpatient services.\(^6\)

The assumptions presented below define the main characteristics of the model:

Assumption 1: the unit circumference circle defines a space which will be labeled “country”;

Assumption 2: the health system provides two types of services, called primary care and secondary or inpatient care;\(^7\)

Assumption 3: there are \(n\) health care centers or clinics evenly distributed around the circle (thus the distance between clinics is \(1/n\));

Assumption 4: the clinics provide only outpatient care;

Assumption 5: there are \(N\) hospitals evenly distributed around the circle (the distance between hospitals is therefore equal to \(1/N\));

Assumption 6: the hospitals can provide inpatient and outpatient care;

---

\(^6\) It must be noted that this simple two-level model is able to capture the main characteristics of a health care pyramid.

\(^7\) Secondary, inpatient or hospital care are terms used interchangeably in this work.
**Assumption 7**: in this stylized country there is a place, the “capital” that defines the initial segment for both hospitals and clinics, which implies that the capital always has an ambulatory clinic and an impatient care facility;

**Assumption 8**: the capital is located along the unit circle at an angle of zero radians, i.e. over the horizontal diameter (see Figures 3 and 4);

**Assumption 9**: there are \( J \) consumers, uniformly distributed around the circle, in need of medical care;

**Assumption 10**: all services are government owned.

Assumptions 7 and 8 are necessary to allow a general notation. Assumption 2 has been modified from the previous subsection to allow for the introduction of the health care pyramid. Assumptions 5 and 6 describe the new type of care. The other assumptions remain from the basic model.

Because inpatient care requires more specialized and expensive inputs, the marginal variable cost of providing this service, \( M \), is greater than \( m \), which, as indicated in the previous model, is the marginal variable cost of the health care centers. Similarly, the hospital’s fixed cost is defined by \( F > f \), the latter indicating the ambulatory clinic fixed cost. The total cost function for the \( k^{th} \) inpatient facility is therefore described by: \( F + MQ_k \) (see equation (2) above).

One of the basic characteristics of the health care pyramid is derived directly from the difference in cost observed between the two kinds of medical care services considered: the number of hospitals is less than the number of outpatient clinics, i.e. \( N < n \). In order to allow for a general notation, it is postulated that \( n/N \) is always an integer. Therefore:
Assumption 11: the ratio of the number of health care centers over the number of hospitals \((n/N)\) is always an integer greater than one.

Initially it is assumed that all patients are fully aware of their health status, i.e. they know what type of treatment they will need and therefore will (correctly) self refer to the specific facility.\(^8\) Thus:

Assumption 12: consumers have perfect knowledge of their health condition, therefore they always seek the adequate facility.

It must be noted that the fact that \(n/N\) is an integer (Assumption 11) implies that hospitals are never isolated, or, in other words, every location that has a hospital also has a clinic (see Figures 3 and 4). This means that the perfect knowledge assumption (Assumption 12) implicitly presumes that patients in need of primary care, and living closer to locations where hospitals and clinics are available, will seek care in the clinics rather than in hospitals.

---

\(^8\) This perfect information assumption will be relaxed in the next subsection, and the need for a referral structure is discussed.
Figure 3 and Figure 4, above, show the distribution of hospitals and clinics for the ratio of clinics over hospitals \((n/N)\) as an odd or an even integer. Figure 3 displays the distribution of hospitals and clinics for an even integer \((n = 8, N = 4)\), and Figure 4 shows the odd integer case for \(n = 9, N = 3\). The location of the country’s capital is also indicated in both cases. In the figures, “C” indicates the locations in which a health clinic is present and “H” those that have a hospital.

From the total patient population of size \(J\), only \(\beta J\) \((0 < \beta < 1)\) individuals present a health problem serious enough to require the more intensive type of care supplied by the hospitals. The remaining \((1-\beta)J\) patients can be adequately served at the primary care units.\(^9\) It is assumed that both types of patients are homogeneously distributed around the circle:\(^{10}\)

\textbf{Assumption 13}: all types of patients are uniformly distributed around the country.

\(^9\) Health planners usually refer to the fact that only about 20% of all cases require hospitalization, which means a \(\beta=0.2\).

\(^{10}\) For notational consistency, variables in capital letters refer to hospitals. The only exception to that is the size of the total patient population: \(J\). The inpatient population is \(\beta J\), as described above.
Since the total number of clinics is different from the number of hospitals, it must be clear that, on average, patients will have to travel different distances depending on the type of health care facility that they need to go to. Because there are more primary care clinics than hospitals, the (travel) distance necessary to get to these facilities will be, in general, less than that to the hospitals: the maximum distance to a hospital is $\frac{1}{2}N$ for those patients that live midway between two hospitals. Because the minimum distance is zero, the average round-trip to a hospital is $\frac{1}{2}N$. As seen in the previous subsection, the average round-trip to a center is $\frac{1}{2}n$.\textsuperscript{11}

The total social cost of this health care system for the country is given by:

\begin{equation}
TSC = \beta \frac{J}{2N} C + \beta J M + \text{NF} + (1 - \beta) \frac{J}{2n} C + (1 - \beta) J m + r f
\end{equation}

The first three elements in (7) display, respectively, the transportation costs, medical costs and fixed costs relative to the (potential) use of the hospitals, while the last three exhibit the same cost items that result from the planned utilization of the clinics.

As discussed above, the medical cases that require hospitalization are regarded as being more serious than those treatable on an outpatient basis. If $B$ is defined as the monetary equivalent of the benefit that a patient would get from a hospital treatment, it follows that $B > b$, which indicates that a person would be willing to travel longer distances to get inpatient care than primary care. This result is intuitive and can be easily shown using an expression like (1) and the fact that $b \geq e$ and $B \geq E$ (where $E$ is the expenditure a patient would incur to get to the closest hospital): in the limit $b = cx_m$, and

\textsuperscript{11} Because $n > N$, it follows that $\frac{1}{2}n < \frac{1}{2}N$. It is also important to note that, for general notation, it is required that $\beta J / N$ and $(1- \beta) J / n$ be even integers.
therefore \( x_m = b/c \), where \( x_m \) is the maximum distance a person is willing to travel to get outpatient care; similarly \( B = cX_m \) and \( X_m = B/c \). Since \( B > b \), \( X_m > x_m \).

The country’s welfare function associated with (7) can be written as:

\[
W = \beta JB + (1-\beta)Jb - \beta JM - (1-\beta)Jm - NF - nf - \frac{\beta Jc}{2N} - \frac{(1-\beta)Jc}{2n}
\]

The first order conditions that determine the welfare maximizing number of hospitals and clinics are simply:

\[
\frac{\partial W}{\partial N} = -F + \beta \frac{Jc}{2N^2} = 0
\]

\[
\frac{\partial W}{\partial n} = -f + (1-\beta) \frac{Jc}{2n^2} = 0
\]

The socially optimal number of facilities, i.e. the number of clinics and hospitals that guarantee that the health needs of the population are satisfied, is expressed as:

\[
r^* = \sqrt{\frac{(1-\beta)Jc}{2f}}
\]

\[
N^* = \sqrt{\frac{Jc}{2F}}
\]

The similarities between the results stated in (9) and (10), and the outcome of the single-service model shown by expression (6), are evident. The optimal number of hospitals and clinics varies inversely with the respective fixed costs, and directly with their share of the total patient population and transportation costs.
Each hospital will have a market share equal to $\beta J / N$. Similarly, each clinic is responsible for serving $\left[(1 - \beta)J\right] / n$ patients.

The optimal ratio of outpatient to inpatient facilities is defined by:

$$\frac{n^*}{N^*} = \frac{\sqrt{F\left(\frac{1}{\beta} - 1\right)}}{f(\beta)}$$

The relative number of facilities is only a function of their fixed costs and of the proportion of patients requiring hospitalization. “Ceteris paribus,” small betas would define a larger $n^*/N^*$ ratio, because fewer patients would need inpatient care.

II.2 The Second-Best Scenario: Imperfect Information

The previous subsection described a first-best world in which patients know what type of medical care they need and therefore go directly to the correct type of medical facility. If, however, health care planners believe that the population in general does not possess the information or knowledge necessary to determine the adequate kind of medical care that they need, the adoption of a health system such as the one described above would lead to the misuse of services. In this scenario, the greatest reason for concern is the possibility of wasting resources with high shadow prices, such as the inputs primarily used in hospitals: specialized and skilled personnel and sophisticated equipment.

The discussion that follows describes the conditions of imperfect information and how they affect the health care system. The next subsection presents the solution usually proposed to minimize the costs imposed by a second-best scenario.

There are three main reasons why a person may look for care in a hospital when he or she can be helped in a primary care clinic:
a) the patient overestimates his or her illness and therefore unknowingly seeks hospital care;

b) a hospital might be more “attractive,” in the sense that it is believed by some patients to provide better care for the same type of problem; and

c) the hospital is closer.

Since in this work $n/N$ is always an integer — as seen before, this implies that every location that has a hospital also has a clinic — case (c) does not apply.$^{12}$ Case (b) can be easily thought of as a special case of (a). In this sense, it is possible to say that most instances in which hospital care is mistakenly used, could be classified under (a). This scenario will be termed one-level misinformation, for it only considers the possibility of those requiring primary care being wrong.

It is obviously possible also to have patients that require inpatient care misinterpreting their health problem and therefore seeking the wrong type of help, i.e. looking for assistance in the clinics. The situations in which both types of error occur are defined as two-level misinformation. Note, however, that a patient would not rationally seek care in a clinic if he or she knew that his or her condition required hospitalization — even if living relatively much closer to a clinic than a hospital — since this patient would have to be transferred from the clinic to an inpatient care facility anyway and would, therefore, incur greater transportation costs. The discussion of the two-level misinformation will be left to subsection II.2.4.

---

$^{12}$ For most cases, since $n/N$ is likely to large, the assumption of integer ratios implies a relatively small rounding.
II.2.1 One-Level Misinformation

In this sense, it is assumed, for the time being, that health problems that require hospitalization are serious enough for consumers with such ailments to correctly identify the type of care needed. Thus Assumption 12, presented in the previous subsection, becomes:

Assumption 12': some of the patients in need of outpatient care mistakenly seek care in hospitals. All those patients that require inpatient care are able to correctly identify the seriousness of their condition and therefore seek assistance in hospitals.

If that is the case, the inpatient population defined by $\beta J$ will direct themselves to the right institutions. However, from the $(1-\beta)J$ patients that only need ambulatory care, $\alpha(1-\beta)J$ individuals, $(0<\alpha<1)$, overestimate their conditions and seek care in hospitals. It has to be noted, though, that the benefits these patients get from the treatment in the hospital is the same they would get from the care at a clinic: $b$. On the other hand, these patients that misjudge their condition and mistakenly seek care in a hospital, will impose on society the higher marginal cost of treatment $M$, instead of $m$, the marginal cost of a treatment in a primary care clinic. The similarity between this scenario and moral hazard must be noted: in both cases patients are using services that provide benefits that are smaller than their (marginal) cost of production to society.

---

13 In order to capture the costs imposed by the inappropriate utilization of resources it is assumed that the marginal variable cost of a treatment is only a function of the inputs used and independent of the severity of the illness. In this sense, any hospital treatment costs $M$, and any ambulatory service costs $m$. 
As a result, the transportation schedule observed in this second-best scenario is constituted by the following elements:

a) the \((1-\alpha)(1-\beta)J\) patients that correctly seek primary care travel an average of \(1/2n\) to go to the clinic and return home;

b) the \(\beta J\) cases that need inpatient care will travel an average of \(1/2N\);

c) those \(\alpha(1-\beta)J\) individuals that overestimate their health problem will also travel an average of \(1/2N\), even though their average travel distance should have been only \(1/2n\).\(^{14}\)

The social costs encountered in such a second-best world would amount to:\(^{15}\)

\[
\text{TSC}_i = (1-\alpha)(1-\beta) \frac{Jc}{2n} + [\beta + \alpha(1-\beta)] \frac{Jc}{2N} + (1-\alpha)(1-\beta)Jm \\
+ [\beta + \alpha(1-\beta)]JM + rf + NF
\]

The expression shows that there are \((1-\alpha)(1-\beta)J\) patients seeking the primary care services offered by the clinics, and therefore traveling an average distance of \(1/2n\), and \(\beta + \alpha(1-\beta)\) patients traveling on average \(1/2N\) to get medical assistance in hospitals.

It can be readily seen from above that, as expected, an imperfect information world would lead to excess costs. With respect to the misuse of personnel and equipment this excess cost amount to \(\alpha(1-\beta)J(M-m)\), i.e. the additional burden, measured as the differences in medical costs \((M - m)\), imposed on society by the individuals that used the more expensive hospital inputs instead of the resources available at the clinics. As noted above, on average these patients also travel longer distances, since hospitals are more

\(^{14}\) General notation requires \(\alpha(1-\beta)J/N\) and \((1-\alpha)(1-\beta)J/n\) to also be even integers (see note 11).

\(^{15}\) The subscript “i” denotes the imperfect knowledge scenario.
sparsely distributed. The resulting excess cost of transportation is equal to $\alpha(1-\beta)J_c[(1/2N) - (1/2n)]$.

The expression below displays the welfare function that health planners would have to maximize in a world characterized by one-level misinformation:

$$W_i = \beta JB + (1-\beta) Jb - \left[\beta + \alpha(1-\beta)\right] \frac{J_c}{2N} - (1-\alpha)(1-\beta) \frac{J_c}{2n}$$

$$- \left[\beta + \alpha(1-\beta)\right] Jm - (1-\alpha)(1-\beta) Jm - NF - rf$$

Note that total benefits are the same as in (8), since these are determined only by the medical condition of the patient and therefore, independent of the location in which the service was provided.

The first order conditions for welfare maximization lead to the following optimal number of clinics and hospitals:

$$n_i^* = \sqrt{\frac{(1-\alpha)(1-\beta) J_c}{2f}}$$

$$N_i^* = \sqrt{\left[\beta + \alpha(1-\beta)\right] \frac{J_c}{2F}}$$

A direct comparison between expressions (13) and (14) with the first-best results of (9) and (10) show that if the scenario, upon which a health care system designed to secure full access is established, is characterized by the presence of one-level misinformation, welfare maximization will lead to relatively more hospitals and fewer clinics than would be observed in a first-best world. That is: $N_i^* > N^*$ and $n_i^* < n^*$. The greater number of hospitals becomes necessary in order to attend these individuals that unknowingly (and inappropriately) seek assistance in these facilities.
These differences can also be seen from the fact that the optimal ratio of clinics to hospitals under the imperfect knowledge scenario is clearly smaller than the one prevailing in the model of the previous subsection:

\[ \frac{n^*}{N_i^*} = \sqrt{\frac{F\left(\frac{1}{\beta + \alpha(1-\beta)}\right) - 1}{f(\beta)}} < \sqrt{\frac{F\left(\frac{1}{\beta} - 1\right)}{f(\beta)}} = \frac{n^*}{N^*} \]

An important implication that can be drawn from (13) and (14) is that the scenario of imperfect information can be explosive for developing countries. Since it is likely that the number of patients requiring inpatient care is a small fraction of the total patient population (i.e. \( \beta \) is small), even a relatively small proportion of patients wrongly seeking care in hospitals (i.e. a small \( \alpha \)), would mean a substantially greater number of hospitals than would have been observed under perfect knowledge. If, for instance \( \alpha = \beta = 0.2 \), then the number of hospitals resulting from (14) is, “ceteris paribus” 34% greater than the first-best results of (10), while the reduction in the number of clinics (shown by the difference between (13) and (9)), is of only 11%. These numbers mean that, compared to a first-best scenario, a second-best world would require from these countries substantial investments in expensive hospitals while allowing for only small “savings” from the fewer number of clinics.\(^\text{16,17}\)

It must be clear that imperfect knowledge does impose an important burden on the health system. Moreover, many developing countries are not likely to have the necessary resources to provide the adequate number of facilities (particularly hospitals). If that is the case, then the

\(^\text{16}\) Not only because their fixed costs are smaller, but also because the reduction in the number of clinics is not substantial.

\(^\text{17}\) If \( \alpha = 0.3 \), instead of 0.2, the number of hospitals under imperfect information would have to be almost 50% greater.
case, health planners realize that full access will be denied and lines will form (mainly at hospitals). It is, therefore, evident that alternatives must be designed.

II.2.2 The Referral System as a Mean to Minimize the Costs of Imperfect Information

If the reality of a country is indeed characterized by a second-best world with imperfect information, alternatives that minimize the costs it imposes must be found, particularly in developing countries where resources are more scarce. Without price mechanisms to induce the desired utilization pattern, the solution usually proposed by health care planners is the organization of a referral strategy. The option for a referral structure can be seen as a government intervention in the economy that reduces consumers’ sovereignty, and is based on the diagnosis that market imperfection is brought by lack of perfect knowledge. Accordingly, consumers are not allowed to act freely in the market and choose the type of health service according to their own perception of their health status.

In this setting, the primary care clinics operate as the system’s gate keepers: they become responsible for screening, i.e. examining, all patients and referring the more complex cases for hospital admission. This policy is founded on the rationale that it would be less costly for the country to have the less expensive, less specialized, and less scarce resources available at the clinics overused. Moreover, it is argued that if there is two-level misinformation, some transfer of patients from one facility to another is already occurring than in the first-best world, if full access is to guaranteed.
(see section II.2.4), without being formally established and organized as a referral structure.

In this sense, the referral framework thus assures that the number of hospitalizations remains restricted to the $BJ$ cases technically judged as needed of such interventions. However, this outcome is obtained at the expense of providing a number of ambulatory visits that is equal to the entire patient population, $J$.\(^{18}\)

There are two alternative dynamics for the referral process, either the patient is referred directly from the clinic to the hospital, or, if the country does not have the necessary coordinating capacity, the hospital admission is scheduled for another date and the person is sent home. If the dynamics of the system is such that patients are transferred directly from the clinic to the nearest hospital, the average distance traveled from one facility to another varies depending on the proportion of clinics to hospitals the $n/N$ ratio:

a) if this ratio is an even integer, the average distance from a center to the closest hospital is given by $1/4N$;

b) if $n/N$ is an odd integer, this average distance is given by the expression:

$$\frac{1}{4}\left[\frac{1}{N} - \frac{N}{n^2}\right].$$\(^{19}\)

It is easy to see that, as the proportion of clinics to hospitals increases, $N/n^2$ quickly tends to zero and the value of the last expression tends to $1/4N$. Since it is likely that the $n/N$ ratio is large, we will assume, for all cases, that $1/4N$ is the average referral travel distance.

---

\(^{18}\) It must be noted that any referral system will show the same number of ambulatory visits and hospitalizations, therefore the same cost structure, independently of the fact that the lack of perfect knowledge that triggered its implementation is characterized by one or two-level misinformation.

\(^{19}\) The proof of these results are shown in the appendix.
Thus, in the case of direct referrals from the clinics to the hospitals, it can be seen that:

a) the \((1-\beta)J\) patients that only need primary care travel the average distance of \(1/2n\) to go to the clinic and return home;

b) these patients impose on society the cost of an ambulatory treatment \(m\);

c) the remaining \(\beta J\) patients that actually need hospital care were first examined at the primary care clinics, thus using resources from both types of facilities. Their (marginal) medical costs amount to \(m + M\);\(^{20}\)

d) these hospitalized patients travel, on average, \(1/4n\) to get to the clinic, plus the average referral distance \(1/4N\) discussed above, plus \(1/4N\) to return home from the hospital.\(^{21}\)

As a result, the system’s total social cost amounts to:\(^{22}\)

\[
TSC_r = (1 - \beta) \frac{Jc}{2n} + \beta \frac{Jc}{4} \left( \frac{1}{n} + \frac{2}{N} \right) + Jm + \beta JM + nf + NF
\]

Note that, by definition, the cost structure of the referral system is independent of the proportion of patients that overestimate their health problem (i.e. \(\alpha\)).

Even though all patients pass through an ambulatory visit, those that need inpatient care do not derive any benefit from the encounter: for such patients the stop at the clinic is a simple screening. This means that the structure of benefits is not affected by

\(^{20}\) As pointed out before, the mere utilization of a service amounts to a social burden equal to its marginal cost (see note 13).

\(^{21}\) A general notation requires that: \((1-\beta)J/n\), \(\beta J/n\) and \(\beta J/N\) be even integers. Which also implies that \(n/N\) will be an even integer.

\(^{22}\) The subscript “\(r\)” indicates the referral system.
the referral strategy, which is the same as saying that the presence of a health need is a
necessary condition for medical care to generate any benefit (there are no placebo effects).
Therefore, the level of benefits for the \((1-\beta)J\) patients requiring outpatient care is equal to
\(b\), while for the \(\beta J\) hospitalized cases it is still equal to \(B\).

The appropriate welfare function to be maximized is then equal to:

\[
W_r = (1 - \beta)Jb + \beta JB - (1 - \beta) \frac{Jc}{2n} - \beta \frac{Jc}{4} \left( \frac{1}{n} + \frac{2}{N} \right) \\
- Jm - \beta JM - nf - NF
\]

The first-order conditions for \(n\) and \(N\) are:

\[
\frac{\partial W}{\partial n} = (1 - \beta) \frac{Jc}{2n^2} + \beta \frac{Jc}{4n^2} - f = 0
\]

\[
\frac{\partial W}{\partial N} = \beta \frac{Jc}{2N^2} - F = 0
\]

Which provide the optimal number of clinics and hospitals under a referral structure:

\[
n^*_r = \sqrt{\frac{2 - \beta}{4f}} \frac{Jc}{2n^2} = \sqrt{\left(1 - \frac{\beta}{2}\right) \frac{Jc}{2f}}
\]

\[
N^*_r = \sqrt{\frac{\beta}{2F}} \frac{Jc}{2N^2}
\]

The result displayed in (18) shows that the referral system does attain its objective of
reducing the burden on hospital expansion: in fact, the optimal number of hospitals under
referral is the same as the one that would be observed under a first-best world (see
expression (10)). The optimal number of clinics (shown by (17)), on the other hand, has to
be greater than the perfect knowledge scenario (expression (9)), if these facilities are to be able to handle the increased workload.\textsuperscript{23}

As described above, the referral structure may be such that requires hospitalizations to be scheduled for another date. This logistic implies that:

a) all patients have to go to a clinic to be examined;

b) those that only need outpatient care are treated and return home (average round trip equal to $1/2n$);

c) those that need inpatient care will have their hospitalization scheduled by the clinic to a future date, and sent home (thus an average round trip also equal to $1/2n$);

d) the $\beta J$ patients that are to be hospitalized will have to cover, at each specific date, the distance to these facilities and the return trip home (average round trip equal to $1/2N$).

In this sense, the social costs imposed by this system are defined by:\textsuperscript{24}

\begin{equation}
TSC_{sr} = \frac{Jc}{2n} + \beta \cdot \frac{Jc}{2N} + Jm + \beta JM + r f + NF
\end{equation}

The welfare function is described by:

\begin{equation}
W_{sr} = (1 - \beta) Jb + \beta JB - \frac{Jc}{2n} - \beta \frac{Jc}{2N} - Jm - \beta JM - rf - NF
\end{equation}

\textsuperscript{23} A more detailed comparison of the results will be left to the next sub-section.

\textsuperscript{24} The subscript "sr" indicates the system of scheduled referrals.
Expressions (21) and (22), below, display the resulting optimal number of clinics and hospitals under a referral system with scheduled inpatient admission:

\[
(21) \quad n_{sr}^* = \sqrt{\frac{Jc}{2f}}
\]

\[
(22) \quad N_{sr}^* = \sqrt{\frac{Jc}{2F}}
\]

The last result shows that the number of hospitals remains unchanged from the system of direct referral. However, as it is clear from (17) and (21), the optimal number of clinics under scheduled referrals have to be greater than in a system that transfers patients directly from the ambulatory unit to the hospital (i.e. \(n_{sr}^* > n_r^*\)), in order to compensate the fact that patients requiring hospital care have to return home after the preliminary examination at the clinic.

In fact, the intuitive result that total transportation costs are smaller when the system is able to refer patients directly from the clinic to the hospital — instead of having to schedule future hospital admissions — is easy to verify. Formally it is required that:

\[
(1 - \beta) \frac{Jc}{2n_t^*} + \beta \frac{Jc}{4n_t^*} + \beta \frac{Jc}{2N_r^*} < \frac{Jc}{2n_{sr}^*} + \beta \frac{Jc}{2N_{sr}^*}
\]

The left-hand side displays transportation costs with direct referral, and the right-hand side these costs with scheduled referrals (see (15) and (19)). Since \(N_r^* = N_{sr}^*\), the expression simplifies to:

\[
\frac{Jc}{2n_t^*} - \beta \frac{Jc}{4n_t^*} < \frac{Jc}{2n_{sr}^*} \Rightarrow \frac{2 - \beta}{n_t^*} < \frac{2}{n_{sr}^*}
\]
Substituting (17) and (21) and squaring both sides results in:

\[(2 - \beta)4 < 8 \quad \text{q.e.d.}\]

This result, associated with the fact that, by definition, medical costs must the same under both referral arrangements, and that “ceteris paribus” fixed costs would be smaller under direct referral than under the scheduled scheme (c.f. expressions (15) and (19)),\(^{25}\) means that the former arrangement will impose a lower overall cost to society, that is: \(TSC_{sr} > TSC_r\). Furthermore, since the health benefits provided by any needs-based system must be the same, it follows that \(W_r > W_{sr}\). In this sense, if a referral structure is to be established, it is to be recommended that, whenever possible, it should be organized as such that the referral of patients is made directly from the clinics to the upper level facilities.

**II.2.3 Examining the Results**

This subsection summarizes and compares the results developed so far. It also provides an answer to the following problem faced by health care planners of developing countries: assuming that the second-best scenario is a better approximation to the real world, when should a referral system be implemented? Or in other words: what are the conditions that make the referral model an appropriate alternative for a country’s health care system?

Table 1, below, presents the results for the optimal number of clinics and hospitals that have been derived throughout this section. The first row shows the optimal

\(^{25}\) This results from the fact that \(N_r = N_{sr}\) and \(n_r < n_{sr}\), and total fixed costs are equal to \(n_rf + N_rF\) and \(n_{sr}f + N_{sr}F\).
number of clinics and hospitals when patients are able to correctly seek the appropriate type of care. The second row displays the outcomes derived under conditions of imperfect knowledge: when some patients overestimate the seriousness of their conditions or do not “trust” the care provided at the outpatient facilities (one-level misinformation), thus seeking care in hospitals when they could be treated just as well in the less complex and less expensive primary care clinics. The last two rows in Table 1 display the outcomes when imperfect information exists but a referral strategy is implemented. The table shows, in the third row the results when the referral process is established directly from one facility to the other (direct referral), and in the last row the case when hospitalization appointments have to be set for future dates (scheduled referral).

Table 1
SUMMARY OF RESULTS

<table>
<thead>
<tr>
<th>SCENARIO</th>
<th>NUMBER OF CLINICS</th>
<th>NUMBER OF HOSPITALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERFECT KNOWLEDGE</td>
<td>$n^* = \sqrt{\frac{(1-\beta)Jc}{2f}}$</td>
<td>$N^* = \sqrt{\frac{\beta Jc}{2F}}$</td>
</tr>
<tr>
<td>No Referral (One-level misinformation)</td>
<td>$n_i^* = \sqrt{(1-\alpha)(1-\beta)\frac{Jc}{2f}}$</td>
<td>$N_i^* = \sqrt{[\beta + \alpha(1-\beta)]\frac{Jc}{2F}}$</td>
</tr>
<tr>
<td>SECOND-BEST</td>
<td>$n_d^* = \sqrt{\frac{(1-\beta)Jc}{2f}}$</td>
<td>$N_d^* = \sqrt{\frac{\beta Jc}{2F}}$</td>
</tr>
<tr>
<td>Scheduled Referral</td>
<td>$n_s^* = \sqrt{\frac{Jc}{2f}}$</td>
<td>$N_s^* = \sqrt{\frac{\beta Jc}{2F}}$</td>
</tr>
</tbody>
</table>

Note: Expression numbers are shown in parenthesis.

It can be seen from the table that:
a) a one-level misinformation world without referral will have more hospitals and less clinics than a first-best scenario, on account of the fact that some patients in need of primary care will be seeking care in hospitals. Thus: \( N_i^* > N^* \) and \( n_i^* < n^* \).

b) health care systems with a referral structure implemented will always provide the same number of hospitals that would have been observed in a first-best world, independently on how the referral process is done, thus achieving its objective of minimizing the utilization of the more expensive and scarce resources encountered in these facilities. In this sense: \( N_{sr}^* = N_r^* = N^* \).

c) it is clear then that the introduction of a referral system imposes an extra burden only on the primary facilities, that now have to screen all patients. In this sense, health care systems that enforce the referral of patients from the lower levels of the pyramid will show more clinics than the first-best scenario, i.e.: \( n_r^* > n^* \) and \( n_{sr}^* > n^* \).

d) \( n_{sr}^* \) is in fact maximum: is the same optimal solution provided by a system that has no inpatient services (see expression (6)). The greater number of clinics is necessary to compensate the extra transportation costs imposed by a system with scheduled referrals;

e) reflecting its relatively lower transportation costs, a system with direct referrals requires an optimal number of clinics \( (n_r^*) \) that is midway between \( n_{sr}^* \) and \( n^* \), thus: \( n_{sr}^* > n_r^* > n^* \).

---

26 The first-best scenario is the reference against which the alternative scenarios are measured.
f) in summary, the previous items unequivocally show that: \( n_{sr}^* > n_r^* > n^* > n_i^* \) and \( N_i^* > N_{sr}^* = N_r^* = N^* \).

These results show that if the first best scenario is used for health care planning but in reality consumers do not possess the level of knowledge or information necessary to correctly assess their health condition — sometimes overestimating the importance of their ailment — the country’s health care system will end up with excess capacity in the primary care clinics and excess demand in the hospitals.

In this sense, if the behavior of the population can be better predicted assuming a second-best world, the key issue for health policy-makers is to know the conditions that determine the appropriateness of a referral system within the structure of health care services. In order to assess the circumstances that would make such intervention desirable, the social costs of the referral structure must be compared with the one-level misinformation scenario without referral. For the former to be the desirable option it is required that \( TSC_{sr} < TSC_i \) and/or \( TSC_r < TSC_i \).\(^{27}\)

Accordingly, from expressions (15) and (11) \( TSC_r < TSC_i \) is written as:

\[
(1 - \beta) \left( \frac{JC}{2n_i^*} + \frac{JC}{4n_i^*} + \frac{2}{N_r^*} \right) + JM + \beta JM + f_{n_i^*} + FN_r^* <
\]

(23)

\[
(1 - \alpha)(1 - \beta) \left[ \frac{JC}{2n_i^*} + \left( \beta + \alpha(1 - \beta) \right) \frac{JC}{2N_i^*} + JM + \frac{f_{n_i^*} + FN_i^*}{2} \right]
\]

\(^{27}\) It has already been shown that the level of benefits remains the same with or without referral. Therefore, if \( TSC_r < TSC_i \) (\( TSC_{sr} < TSC_i \)), then \( W_r > W_i \) (\( W_{sr} > W_i \)). Moreover, it has been shown in the previous sub-section that \( TSC_{sr} > TSC_i \). Thus, if \( TSC_{sr} < TSC_i \), it follows directly that \( TSC_r < TSC_i \). However, for completeness both cases will be presented.
And $TSC_{sr} < TSC_i$ (see (19) and (11)) is expressed as:

$$
\frac{Jc}{2n_r} + \beta \frac{Jc}{2N_r} + JM + \beta JM + f_{n_r} + F_{N_r} < \\
(1 - \alpha)(1 - \beta) \frac{Jc}{2n_i} + \left[ \beta + \alpha (1 - \beta) \right] \frac{Jc}{2N_i} + \\
(1 - \alpha)(1 - \beta) JM + \left[ \beta + \alpha (1 - \beta) \right] JM + f_{n_i} + F_{N_i}
$$

(24)

Since there are too many parameters in the last two expressions for any direct comparisons between the alternative scenarios, total costs ($TSC$) will be divided in social medical and non-medical costs ($SMC$ and $SNMC$, respectively). Let us consider expression (24) as an example: its inequality would be unambiguously satisfied if $SMC_{sr} < SMC_i$ and $SNMC_{sr} \leq SNMC_i$ (or $SMC_{sr} \leq SMC_i$ and $SNMC_{sr} < SNMC_i$). In this case the four strict inequalities that satisfy (23) and (24) are:

i. $SMC_r < SMC_i$:

(25) $JM + \beta JM < (1 - \alpha)(1 - \beta) JM + \left[ \beta + \alpha (1 - \beta) \right] JM$

ii. $SNMC_r < SNMC_i$:

(26) $\frac{Jc}{2n_r} + \beta \left( \frac{1}{n_i} + \frac{2}{N_r} \right) + f_{n_r} + F_{N_r} < \\
(1 - \alpha)(1 - \beta) \frac{Jc}{2n_i} + \left[ \beta + \alpha (1 - \beta) \right] \frac{Jc}{2N_i} + f_{n_i} + F_{N_i}$

iii. As noted earlier, the medical costs under the two alternative referral strategies are necessarily the same, thus $SMC_{sr} < SMC_i$ is as in (i):

28 It is, of course, possible to imagine a case in which $SMC_{sr} > SMC_i$ and $SNMC_{sr} < SNMC_i$ (or $SMC_{sr} < SMC_i$ and $SNMC_{sr} > SNMC_i$) and still have $TSC_{sr} < TSC_i$. However, the analysis that follows imposes the stronger double constraint described in the text.

29 Obviously the same type of restrictions would have to be satisfied for $TSC_r < TSC_i$. 
The solution for (i) and (iii) are easy to obtain. With a few simplifications and rearrangements the inequalities in (25) and (27) will be satisfied if:

\[
\frac{M}{m} > 1 + \frac{\beta}{\alpha(1 - \beta)} \quad \text{and} \quad \frac{m}{M} < \frac{\alpha(1 - \beta)}{[\beta + \alpha(1 - \beta)]}
\]

According to the last result, the referral strategy is more likely to generate less medical costs to society than a non-interventionist approach if:

a) ambulatory care is relatively expensive;

b) the marginal cost of producing hospital care is relative small;

c) the relation between the cost of providing primary care over the cost of an inpatient admission is large;

d) those individuals that misjudge their health condition represent a large proportion of all patients that seek care in hospitals. That is, as the proportion of patients that unnecessarily seek care in hospitals increases, the proportion in which the marginal cost of producing inpatient care must exceed the cost of ambulatory care diminishes (see also the first configuration in (29)).

The cells in Table 2 display, for several possible magnitudes of \( \alpha \) and \( \beta \), the values of the \( M/m \) ratio that equate \( SMC_r \) and \( SMC_i \) (therefore also equate \( SMC_{sr} \) and
SMC_i). In other words, the table presents the ratios of marginal costs that make the medical cost of a system with any type of referral equal to one that does not have the strategy implemented. In this sense, the ratio of marginal treatment costs (M/m), has to be greater than the values shown in the table in order to satisfy the strict inequality of expression (25).

Table 2
M/m Values that Return SMC_r = SMC_i or SMC_s = SMC_i

<table>
<thead>
<tr>
<th>α</th>
<th>β</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.111</td>
<td>3.5</td>
<td>5.286</td>
<td>7.667</td>
<td>11</td>
<td>16</td>
<td>24.333</td>
<td>41</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>1.556</td>
<td>2.25</td>
<td>3.143</td>
<td>4.333</td>
<td>6</td>
<td>8.5</td>
<td>12.667</td>
<td>21</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.37</td>
<td>1.833</td>
<td>2.429</td>
<td>3.222</td>
<td>4.333</td>
<td>6</td>
<td>8.778</td>
<td>14.333</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>1.278</td>
<td>1.625</td>
<td>2.071</td>
<td>2.667</td>
<td>3.5</td>
<td>4.75</td>
<td>6.833</td>
<td>14.333</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.222</td>
<td>1.5</td>
<td>1.857</td>
<td>2.333</td>
<td>3</td>
<td>4</td>
<td>5.667</td>
<td>9</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1.185</td>
<td>1.417</td>
<td>1.714</td>
<td>2.111</td>
<td>2.667</td>
<td>3.5</td>
<td>4.889</td>
<td>7.667</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>1.159</td>
<td>1.357</td>
<td>1.612</td>
<td>1.952</td>
<td>2.429</td>
<td>3.143</td>
<td>4.333</td>
<td>6.714</td>
<td>13.857</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.139</td>
<td>1.313</td>
<td>1.536</td>
<td>1.833</td>
<td>2.25</td>
<td>2.875</td>
<td>3.917</td>
<td>6</td>
<td>12.25</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.123</td>
<td>1.278</td>
<td>1.476</td>
<td>1.741</td>
<td>2.111</td>
<td>2.667</td>
<td>3.593</td>
<td>5.444</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen from the table, with small α’s and large β’s the referral system would be desirable only if hospital services are relatively costly. In cases like these — significant proportions of patients requiring hospitalization but a relatively small segment overestimating their health problem — only highly expensive hospital inputs would justify the extra costs of having a relatively large proportion of the patients examined at the clinics and not deriving any actual benefit from the medical encounter, since they cannot be helped by the technology available in these facilities. Conversely, small β’s and large

\[\text{Table 2}

M/m Values that Return SMC_r = SMC_i or SMC_s = SMC_i

<table>
<thead>
<tr>
<th>α</th>
<th>β</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>2.111</td>
<td>3.5</td>
<td>5.286</td>
<td>7.667</td>
<td>11</td>
<td>16</td>
<td>24.333</td>
<td>41</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>1.556</td>
<td>2.25</td>
<td>3.143</td>
<td>4.333</td>
<td>6</td>
<td>8.5</td>
<td>12.667</td>
<td>21</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.37</td>
<td>1.833</td>
<td>2.429</td>
<td>3.222</td>
<td>4.333</td>
<td>6</td>
<td>8.778</td>
<td>14.333</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>1.278</td>
<td>1.625</td>
<td>2.071</td>
<td>2.667</td>
<td>3.5</td>
<td>4.75</td>
<td>6.833</td>
<td>14.333</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.222</td>
<td>1.5</td>
<td>1.857</td>
<td>2.333</td>
<td>3</td>
<td>4</td>
<td>5.667</td>
<td>9</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>1.185</td>
<td>1.417</td>
<td>1.714</td>
<td>2.111</td>
<td>2.667</td>
<td>3.5</td>
<td>4.889</td>
<td>7.667</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>1.159</td>
<td>1.357</td>
<td>1.612</td>
<td>1.952</td>
<td>2.429</td>
<td>3.143</td>
<td>4.333</td>
<td>6.714</td>
<td>13.857</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>1.139</td>
<td>1.313</td>
<td>1.536</td>
<td>1.833</td>
<td>2.25</td>
<td>2.875</td>
<td>3.917</td>
<td>6</td>
<td>12.25</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td>1.123</td>
<td>1.278</td>
<td>1.476</td>
<td>1.741</td>
<td>2.111</td>
<td>2.667</td>
<td>3.593</td>
<td>5.444</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

As can be seen from the table, with small α’s and large β’s the referral system would be desirable only if hospital services are relatively costly. In cases like these — significant proportions of patients requiring hospitalization but a relatively small segment overestimating their health problem — only highly expensive hospital inputs would justify the extra costs of having a relatively large proportion of the patients examined at the clinics and not deriving any actual benefit from the medical encounter, since they cannot be helped by the technology available in these facilities. Conversely, small β’s and large

30 Since it is easier to visualize, the M/m ratio, rather than its inverse (the m/M ratio), is presented in the table.
\(\alpha\)'s describe a society with few people actually requiring inpatient care, but with a large proportion of patients believing that they need to be hospitalized (or not trusting the care offered at the health clinics), and therefore seeking help in these inpatient facilities. This is, undoubtedly, the setting in which a referral strategy would have the greatest impact in terms of reducing unnecessary medical expenditures. In a scenario like that, the referral structure would be feasible even if hospital services are not notably costlier than the primary care services. Consider, for instance, the commonly used figure of 20% of medical cases requiring hospitalization (the column of \(\beta = 0.2\) in Table 2), it can be seen from the table that the referral system is likely to be a viable strategy even when the marginal treatment costs in hospitals are not substantially greater than those observed in the primary care clinics.

The fact that the value of the \(M/m\) ratio necessary to satisfy the inequality of expression (29) decreases as \(\alpha\) increases, and augments with larger values of \(\beta\), are formally shown by:

\[
\frac{\partial (M/ m)}{\partial \alpha} = \frac{-\beta}{\alpha^2(1-\beta)} < 0
\]

\[
\frac{\partial (M/ m)}{\partial \beta} = \frac{1}{\alpha(1-\beta)^2} > 0
\]

These trends can be visualized in Figure 5, which plots the values presented in Table 2:
Since the second derivatives of $M/m$ are positive for both $\beta$ and $\alpha$, larger and larger values of $\alpha$ will make the referral strategy more rapidly viable.\textsuperscript{31} On the other hand, as the proportion of patients requiring hospitalization increases, the referral structure becomes ever more difficult to be supported. In other words, as the proportion of patients requiring hospitalization expands, the referral strategy would only be justified if the cost of a treatment at a hospital facility, relative to that at a clinic, not only increases but increases at higher rates (i.e. ambulatory care must be increasingly cheaper).

The solutions to the expressions involving the so called non-medical costs (inequalities (26) and (28)) , however, do not provide results that are as easy to interpret. Substituting the values of $n_{r,*}$ and $N_{r,*}$, and $n_{sr,*}$ and $N_{sr,*}$ into (26) and (28), respectively, and using $n_{i,*}$ and $N_{i,*}$ in both cases, the following expressions are obtained:

\textsuperscript{31} The second derivative of $M/m$ with respect to $\alpha$ is $(2\beta)/[\alpha^2(1-\beta)^2]$, and with respect to $\beta$ is $2/[\alpha(1-\beta)^2]$. 
\[
\begin{align*}
\text{(30)} \quad \frac{F}{f} & > \left( \frac{\sqrt{2-\beta} - \sqrt{2(1-\alpha)(1-\beta)}}{\sqrt{2[\sqrt{\beta + \alpha(1-\beta)}]}} \right)^2 \text{ for } \text{SNMC}_r < \text{SNMC}_s \\
\text{(31)} \quad \frac{F}{f} & > \left( \frac{1 - \sqrt{(1-\alpha)(1-\beta)}}{\sqrt{\beta + \alpha(1-\beta)} - \sqrt{\beta}} \right)^2 \text{ for } \text{SNMC}_sr < \text{SNMC}_s 
\end{align*}
\]

Since it has been shown earlier that \(SMC_r = SMC_{sr}\) and \(TSC_r < TSC_{sr}\), for any given value of \(\alpha\) and \(\beta\), the ratios of fixed costs \((F/f)\) necessary to make the referral strategies feasible must be greater under the scheduled scheme than with the alternative of direct referral, i.e. the right-hand side of expression (30) must be smaller than the right-hand side of (31). In fact, it is easy to see that this is the case considering that the former can be written as:

\[
\left( \frac{1}{\sqrt{\frac{1}{2}}} \frac{1}{\sqrt{\beta + \alpha(1-\beta) - \sqrt{\beta}}} - \frac{\sqrt{(1-\alpha)(1-\beta)}}{\sqrt{\beta + \alpha(1-\beta)} - \sqrt{\beta}} \right)^2
\]

Which is clearly smaller than the right-hand side of (31), since \(1-(\beta/2)^{1/2} < 1\).

Similarly to what was done above, Table 3 displays, for different values of \(\alpha\) and \(\beta\), the magnitude of the \(F/f\) ratio necessary to implement, in a world with one-level misinformation, a strategy of direct referrals that will impose on society the same level of non-medical costs that would prevail if the system were to allow the patients to decide by themselves which services to use. In this sense, if, for a given \(\alpha\) and \(\beta\), the observed ratio of fixed costs is greater than the one shown in the table, the direct referral structure will be a less costly alternative and should be offered to the population.
Table 3
F/f Values that Return SNMC_r = SNMC_i

<table>
<thead>
<tr>
<th>α</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.389</td>
<td>1.494</td>
<td>4.484</td>
<td>12.152</td>
<td>31.99</td>
<td>86.79</td>
<td>261.01</td>
<td>994.10</td>
<td>7,060</td>
</tr>
<tr>
<td>0.2</td>
<td>0.351</td>
<td>0.947</td>
<td>2.256</td>
<td>5.171</td>
<td>11.977</td>
<td>29.342</td>
<td>81.128</td>
<td>287.62</td>
<td>1,915</td>
</tr>
<tr>
<td>0.3</td>
<td>0.384</td>
<td>0.859</td>
<td>1.779</td>
<td>3.64</td>
<td>7.665</td>
<td>17.301</td>
<td>44.506</td>
<td>147.83</td>
<td>925.40</td>
</tr>
<tr>
<td>0.4</td>
<td>0.439</td>
<td>0.873</td>
<td>1.64</td>
<td>3.088</td>
<td>6.036</td>
<td>12.732</td>
<td>30.752</td>
<td>96.201</td>
<td>567.30</td>
</tr>
<tr>
<td>0.5</td>
<td>0.51</td>
<td>0.933</td>
<td>1.633</td>
<td>2.883</td>
<td>5.305</td>
<td>10.562</td>
<td>24.123</td>
<td>71.373</td>
<td>397.07</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
<td>1.03</td>
<td>1.705</td>
<td>2.857</td>
<td>4.999</td>
<td>9.467</td>
<td>20.554</td>
<td>57.698</td>
<td>303.05</td>
</tr>
<tr>
<td>0.7</td>
<td>0.715</td>
<td>1.168</td>
<td>1.848</td>
<td>2.964</td>
<td>4.965</td>
<td>8.992</td>
<td>18.631</td>
<td>49.727</td>
<td>246.52</td>
</tr>
<tr>
<td>0.8</td>
<td>0.872</td>
<td>1.367</td>
<td>2.082</td>
<td>3.215</td>
<td>5.18</td>
<td>9.004</td>
<td>17.843</td>
<td>45.303</td>
<td>211.54</td>
</tr>
<tr>
<td>0.9</td>
<td>1.119</td>
<td>1.692</td>
<td>2.489</td>
<td>3.712</td>
<td>5.764</td>
<td>9.626</td>
<td>18.238</td>
<td>43.938</td>
<td>192.01</td>
</tr>
</tbody>
</table>

Table 4
F/f Values that Return SNMC_s_r = SNMC_i

<table>
<thead>
<tr>
<th>α</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.698</td>
<td>3.417</td>
<td>11.604</td>
<td>33.55</td>
<td>90.97</td>
<td>247.94</td>
<td>733.25</td>
<td>2,685</td>
<td>17,740</td>
</tr>
<tr>
<td>0.2</td>
<td>0.506</td>
<td>1.714</td>
<td>4.739</td>
<td>12.006</td>
<td>29.658</td>
<td>75.378</td>
<td>210.97</td>
<td>737.89</td>
<td>4,680</td>
</tr>
<tr>
<td>0.3</td>
<td>0.499</td>
<td>1.356</td>
<td>3.25</td>
<td>7.429</td>
<td>16.976</td>
<td>40.561</td>
<td>107.88</td>
<td>361.09</td>
<td>2,199</td>
</tr>
<tr>
<td>0.4</td>
<td>0.536</td>
<td>1.258</td>
<td>2.708</td>
<td>5.7</td>
<td>12.187</td>
<td>27.551</td>
<td>69.868</td>
<td>224.14</td>
<td>1,311</td>
</tr>
<tr>
<td>0.5</td>
<td>0.599</td>
<td>1.26</td>
<td>2.496</td>
<td>4.905</td>
<td>9.899</td>
<td>21.28</td>
<td>51.601</td>
<td>158.83</td>
<td>891.96</td>
</tr>
<tr>
<td>0.6</td>
<td>0.684</td>
<td>1.324</td>
<td>2.45</td>
<td>4.543</td>
<td>8.709</td>
<td>17.874</td>
<td>41.528</td>
<td>122.71</td>
<td>661.16</td>
</tr>
<tr>
<td>0.7</td>
<td>0.797</td>
<td>1.444</td>
<td>2.522</td>
<td>4.445</td>
<td>8.133</td>
<td>15.983</td>
<td>35.625</td>
<td>101.06</td>
<td>521.58</td>
</tr>
<tr>
<td>0.8</td>
<td>0.954</td>
<td>1.635</td>
<td>2.718</td>
<td>4.573</td>
<td>8.011</td>
<td>15.098</td>
<td>32.289</td>
<td>87.818</td>
<td>432.88</td>
</tr>
<tr>
<td>0.9</td>
<td>1.205</td>
<td>1.962</td>
<td>3.116</td>
<td>5.017</td>
<td>8.419</td>
<td>15.199</td>
<td>31.113</td>
<td>50.789</td>
<td>377.79</td>
</tr>
</tbody>
</table>

Table 4, above, presents the same type of results for a referral structure that requires a scheduled transfer of patients from the primary clinics to the hospitals. Again,
the strategy of scheduled referrals will be desirable if, for a given pair of parameters $\alpha$ and $\beta$, the existing $F/f$ ratio is greater than the corresponding value on the table.

As the discussion above anticipated, the figures in Table 4 are larger than the ones shown in the previous one, meaning that the implementation of a system with scheduled referrals will require scenarios in which the investment costs in hospitals are relatively more expensive and/or the fixed costs of the primary care clinics are relatively less expensive than a strategy with direct referrals.

As can be seen from both tables, for any given $\alpha$, as $\beta$ increases the ratio of fixed costs required to make a referral structure desirable also increases (similarly to what was observed with medical costs, i.e. $\partial(F/f)/\partial\beta > 0$): as the proportion of patients actually requiring hospitalization increases, the investment costs needed to add primary care facilities must be increasingly small, relative to a hospital’s fixed cost, in order to make a referral strategy economically viable, for there are extra transportation costs — imposed on an increasingly greater number of patients — that must be compensated: (note that larger $F/f$ ratios imply, “ceteris paribus,” larger $n/N$ ratios).

It is interesting to notice that for “smaller” values of $\beta$ the $F/f$ ratios presented in the last two tables initially decreases with $\alpha$, but as the parameter increases the ratio begins to rise. It must be noted that as $\beta$ increases, the value of $\alpha$ necessary to make the declining trend of the $F/f$ ratio shift into a growing pattern also increases. Consider, for instance Table 4: when $\beta = 0.2$, the minimum value of the $F/f$ ratio that satisfies expression (31) declines from $\alpha = 0.1$ to $\alpha = 0.4$ and increases afterwards; when $\beta = 0.3$ the $F/f$ ratio declines from $\alpha = 0.1$ to $\alpha = 0.6$. The constraining factor here is the
relatively “small” proportion of patients requiring hospital care: if $\beta$ is too small, there are relatively few individuals actually in need of hospital care, which means that there are no great demand pressures on these facilities. Moreover, low values of $\beta$ and large $\alpha$’s imply a relatively large proportion of the patient population (unnecessarily) seeking care in hospitals (i.e. $\alpha(1 - \beta)J$ is significant), in this sense, a referral strategy would divert many cases to the clinics, i.e. the clinics will not only have to screen all consumers, but actually treat a number of patients that is substantially greater than what would have been observed under a non-referral context, therefore the need for relatively more clinics (larger $n/N$ ratios) which, as seen, are imposed by the higher $F/f$ ratios shown in Table 4.

Figure 6 displays, for values of $\beta$ varying from 0.1 to 0.5, the data of Table 3. The clearly defined “u” shape format of the curve when $\beta = 0.5$ shows distinctly the change in the tendency of the $F/f$ ratio discussed above:
The relatively small values that the $F/f$ ratio can assume, must be noted, reaching figures of even less than one. If $\beta$ is small, very few individuals will be submitted to the greater transportation costs that mark the referral structure. In this sense, this setting becomes easily desirable. If, once again, the case of $\beta = 0.2$ is used as reference, the last two tables show that the values of the $F/f$ required to make the referral system desirable are indeed not substantial, even when the proportion of patients that would overestimate their health condition is small.

In conclusion, the results displayed in tables 2, 3 and 4 indicate that, for likely values of $\beta$, the referral system does seem to be an appropriate form of intervention when patient behavior can be described by one-level misinformation, thus enhancing the welfare of the population. The extra transportation costs and the additional burden placed on the clinics would be more than compensated by the reduction in the misuse of hospital care and lower investment in these facilities. Furthermore, as will be demonstrated below, the referral system is an appropriate choice whenever a second-best scenario is present.

II.2.4 Two-Level Misinformation

The objective of this subsection is to introduce into the analysis the possibility of misjudgments occurring in patients that need care at the two levels of the pyramid, i.e. patients can not only overestimate their health problem, but also underestimate the seriousness of their conditions. This scenario is termed two-level misinformation. With all the information available, a general comparison of all possible scenarios becomes possible.

The two-level misinformation framework removes the assumption that patients requiring hospitalization would be able to correctly identify the kind of medical care they
need (Assumption 12’). Accordingly, some individuals may underestimate their impairments and seek outpatient care when they need treatment in an inpatient basis. Thus, Assumption 12’, presented in subsection II.2.2 is replaced by:

Assumption 12’’: some of the patients in need of outpatient care mistakenly seek care in hospitals, and some of those patients that require inpatient care unknowingly seek help in the primary care facilities.

If that is the case, after being examined in the primary services these patients would have to be referred to the secondary level of care, since the clinics would not have the necessary inputs to treat them.\[32\] If \( \theta \beta J \ (0<\theta<1) \) defines the number of patients that, in need of hospital care, underestimate their health problem and seek the services of the primary care clinics, the scenario of two-level misinformation presents the following characteristics:

a) \( \beta J \) patients need hospital care;

b) \( (1-\beta)J \) patients only need primary care;

c) \( \alpha(1-\beta)J \) is the number of patients that only need primary care but, if allowed, would demand care in hospitals since they overestimate their health problem. This is what was called first-level misinformation;

d) \( (1-\alpha)/(1-\beta)J \) patients correctly seek outpatient care;

e) \( \theta \beta J \) patients will seek care in the clinics, where they will be examined (thus imposing the medical cost \( m \)). But since they actually need inpatient care, these individuals will have to be referred for hospitalization (which have marginal variable costs equal to \( M \)). This is what can be called second-level misinformation;

\[32\] As before, this referral can be made directly from the clinic to the hospital or on an appointment basis in which the
f) \((1-\theta)\beta J\) patients correctly perceive that they need inpatient treatment.

It is important to notice that second-level misinformation would only have any meaning in a system without a formal referral structure, since with referral all patients would have to show up first at the clinics. In this sense, the referral structure is not only independent of \(\alpha\), as shown before, but also of \(\theta\), which means that its costs and outcomes are the same whether one or two-level misinformation is the appropriate description of the real world.

As noted above, some transferring of patients will have to exist in world with a two-level misinformation: those patients that underestimate their health problem and seek care in outpatient facilities will have to be referred to a hospital, for the clinics cannot treat them adequately. As a result, the transportation schedule when two-level misinformation exists and patients are transferred directly from the clinic to the closest hospital is described by the following elements:

a) \((1-\alpha)(1-\beta) J\) patients correctly seek outpatient care and travel an average of \(1/2n\) to get to the clinic and return home;

b) there are \(\alpha(1-\beta) J\) patients that only need outpatient care but went to hospitals, thus for them the round trip averages \(1/2N\);

c) the \((1-\theta)\beta J\) individuals that correctly identify their need for hospitalization will also travel, on average, \(1/2N\) to get to the hospital and return home;

---

patient has to return home.
d) the average travel distance for the $\theta \beta J$ patients that present second-level misinformation amount to: $1/4n$ to get to the clinic, plus $1/4N$ (the average transfer distance), plus $1/4N$ (the average return trip from the hospital).  

Therefore the total social cost of this system is shown by the following expression:  

$$TSC_{d2} = (1 - \alpha)(1 - \beta) \frac{Jc}{2n} + \alpha(1 - \beta) \frac{Jc}{2N} + (1 - \theta \beta) \frac{Jc}{2N}$$  

$$+ \theta \beta Jc \left( \frac{1}{4n} + \frac{1}{2N} \right) + Jm[(1 - \alpha)(1 - \beta) + \theta \beta]$$  

$$+ Jm\{\alpha(1 - \beta) + \beta]\} + m + NF$$  

Note that all $\beta J$ patients will necessarily end up being treated at the hospitals; furthermore, the $\theta \beta J$ individuals that misjudged their health condition are only examined at the clinics and therefore do not derive any benefit from the encounter. The $\alpha(1 - \beta)J$ patients that mistakenly sought care in the hospitals, on the other hand, can be treated at these facilities but the benefits derived from the treatment are only equal to $b$. Expression (33) displays the corresponding welfare function:  

$$W_{d2} = \beta JB + (1 - \beta)Jb - (1 - \alpha)(1 - \beta) \frac{Jc}{2n} - \theta \beta \frac{Jc}{4n}$$  

$$- \left[\alpha(1 - \beta) + \beta\right] \frac{Jc}{2N} - Jm[(1 - \alpha)(1 - \beta) + \theta \beta]$$  

$$- Jm\{\alpha(1 - \beta) + \beta\} - m + NF$$  

The welfare maximizing solutions for the number of clinics and hospitals are:

---

33 As in the previous subsection, a general notation requires that: $(1 - \beta)J/n$, $\beta J/n$ and $\beta J/N$ be even integers. Which also implies that $n/N$ will be an even integer.

34 The subscript in (32) refer to the direct transfer of patients from the clinics to the hospitals within a scenario that presents two-level misinformation.
The same type of reasoning applies when patients are not transferred directly to the hospital from the clinic but have to return home prior to hospital admission. The only difference with respect to the case in which patients are transferred directly from the clinic to the hospital is the average distance traveled by those $\theta \beta J$ patients that underestimate their health condition. These individuals will now have to travel an average of $1/2n$ to get to the clinic and return home, plus a round trip to the hospital, which is on average equal to $1/2N$. Accordingly, the structure of social costs and the welfare function are defined by the following expressions:

\[ TSC_{s_2} = (1 - \alpha)(1 - \beta) \frac{J_c}{2n} + \alpha(1 - \beta) \frac{J_c}{2N} + (1 - \theta \beta) \frac{J_c}{2N} \]
\[ + \theta \beta J_c \left( \frac{1}{2n} + \frac{1}{2N} \right) + J_m \left[(1 - \alpha)(1 - \beta) + \theta \beta \right] \]
\[ + J_M \left[ \alpha(1 - \beta)+ \beta \right] + rf + NF \]

\[ W_{s_2} = \beta JB + (1 - \beta) Jb - \left[(1 - \alpha)(1 - \beta) + \theta \beta \right] \frac{J_c}{2n} \]
\[ - \left[ \alpha(1 - \beta)+ \beta \right] \frac{J_c}{2N} - J_m \left[(1 - \alpha)(1 - \beta) + \theta \beta \right] \]
\[ - J_M \left[ \alpha(1 - \beta) + \beta \right] - rf - NF \]

And the optimal number of clinics and hospitals is given by:

\[ n_{a_2}^* = \sqrt{\left[2(1 - \alpha)(1 - \beta) + \theta \beta \right] \frac{J_c}{4f}} \]

\[ N_{a_2}^* = \sqrt{\left[\alpha(1 - \beta) + \beta \right] \frac{J_c}{2F}} \]

---

35 The subscript indicates a scheduled hospitalization appointment for those patients that underestimated their health condition, within a two-level misinformation world.
If the results of the two-level misinformation are compared it can be easily seen that:

i. $N_{d2}^* = N_{s2}^*$; and

ii. $n_{d2}^* < n_{s2}^*$.

That is, welfare maximization in the presence of two-level misinformation requires more clinics when the system cannot transfer patients directly from the clinics to hospitals, as a way of compensating for the greater transportation costs that these individuals are forced to incur. In fact, it is easy to show that even with more clinics, a system that does not transfer patients directly from the clinic to the hospital will present greater total transportation costs:

$$\frac{1}{2} \left( \frac{1}{2n_{d2}^*} + \frac{1}{N_{d2}^*} \right) < \frac{1}{2} \left( \frac{1}{n_{s2}^*} + \frac{1}{N_{s2}^*} \right) \Rightarrow \frac{n_{s2}^*}{n_{d2}^*} < 2, \text{ for } N_{d2}^* = N_{s2}^*$$

Substituting (34) and (38) and squaring both sides of the last expression:

$$\frac{2 \left[ (1-\alpha)(1-\beta) + \theta \beta \right]}{2(1-\alpha)(1-\beta) + \theta \beta} < 4 \quad \text{qed.}$$

This result implies that, when patients are not transferred directly from the clinic to the hospital, the greater number of clinics that maximizes welfare do not offset completely the
greater distance traveled by patients, and since this is the only difference between the two alternatives it follows that: $TSC_{s2} > TSC_{d2}$.

**II.2.5 Summary of Results and Conclusions**

Table 5, below, summarizes the main results obtained so far, thus complementing Table 1 with the outcomes of the two-level misinformation.

![Table 5]

**Note**: Expression numbers are shown in parenthesis.

36 A formal proof for $TSC_{s2} > TSC_{d2}$ is given in the appendix.
By direct inspection of the table it is clear that the referral structure provides the number of hospitals that is equal to the first-best case, as it is expected from it, and it occurs independently on how the referral process is set. A second-best world without referral, on the other hand, will require relatively more hospitals in order to serve those that demand hospital care unnecessarily. Therefore:

a) \( N_{s2}^* = N_{d2}^* = N_i^* > N_{sr}^* = N_r^* = N^* \).

It can also be easily seen that:

b) \( n_{s2}^* > n_{d2}^* > n_i^* \).

That is, a scenario with two-level misinformation will need more clinics than a world in which only first-level misinformation exists, because these facilities will also have to examine patients that have underestimated their health problem.

If it is possible to assume that: a) \( \alpha > \theta \), i.e., the proportion of patients that overstate the severity of their disease is greater than the proportion of individuals that underestimate their health problem; and b) \( \beta < 0.5 \), i.e., less than half of the total patient population actually require inpatient care, then it can be shown that the number of clinics in a two-level misinformation world will always be less than the optimal number for the first best scenario (i.e. \( n_{s2}^* < n^* \)), and an unambiguous ranking of the number of clinics can be established. For \( n_{s2}^* < n^* \) it is (“ceteris paribus”) necessary that:

\[
\left[ (1-\alpha)(1-\beta) + \theta\beta \right] \frac{Jc}{2f} < (1-\beta) \frac{Jc}{2f}
\]

Which simplifies to:

\[
\beta(\alpha + \theta) < \alpha \Rightarrow 1 + \frac{\theta}{\alpha} < \frac{1}{\beta}
\]
The last inequality is always satisfied if conditions \((i)\) and \((ii)\), described above, are met. If this is case, and it must be noted that these are likely conditions, then the ranking for the optimal number of clinics in the several scenarios examined becomes:

\[
c) \quad n_{sr}^* > n_r^* > n^* > n_{s2}^* > n_{d2}^* > n_i^*.
\]

With the referral system, the optimal number of clinics is greatest, reflecting the extra workload on primary services. Without it, the number of clinics will always be less than that prevailing in a first-best scenario. The possibility that some patients may underestimate their conditions, thus increasing the number of visits to outpatient clinics, means that more of these facilities will be necessary \((n_{s2}^* > n_{d2}^* > n_i^*\), see item \((b)\) above), but since these patients cannot be actually treated there, the number of hospitals will not be reduced \((N_{s2}^* = N_{d2}^* = N_i^*)\). Note, however, that whenever imperfect information is present, hospitals will have to treat patients that could have been helped in primary care units. This will lead to a greater number of these more complex facilities than would be observed either in a first-best setting, or with the implementation of formal referral structure (see item \((a)\) above). In other words, without referral, hospitals can be a substitute for outpatient clinics, but these cannot substitute for the inpatient care units.

As seen, with two-level misinformation some patients, those that mistakenly seek care in the clinics, will have to be transferred to the upper levels of the pyramid. The establishment of the referral system is an attempt to generalize and formalize these procedures to the entire patient population. In this sense, a referral strategy could be seen as an active health policy put by a government facing a second-best scenario. Conversely, the solutions described under subscript \(i\) are the result of a health policy approach in which the government realizes the lack of perfect knowledge by the patients and passively react...
by providing the number of medical care facilities that would be maximizing welfare for such conditions.

It has been shown in the previous subsection that, under reasonable assumptions, the costs imposed by the referral structure are smaller than those that society would face with the inappropriate utilization of its resources, particularly those with high shadow prices such as the ones found in hospitals (i.e. we have shown that \( TSC_i > TSC_{sr} > TSC_r \)). Similarly, it can also be shown that the existence of two-level misinformation involves greater social costs than a one-level misinformation scenario without referral. Intuitively, this occurs because those patients that could not be helped at the clinics not only unnecessarily consumed the resources of the clinics, but also incurred in greater transportation costs. Therefore, \( TSC_{i2} > TSC_{d2} > TSC_r \). Since it has already been seen that \( TSC_{i2} > TSC_{d2} \), it only remains to be proven that \( TSC_{d2} > TSC_r \), which will occur if:

\[
(1 - \alpha)(1 - \beta)\frac{Jc}{2n_{d2}^*} + \theta \beta \frac{Jc}{4n_{d2}^*} + \left[\alpha(1 - \beta) + \beta\right] \frac{Jc}{2N_{d2}^*} + J_n [(1 - \alpha)(1 - \beta) + \theta \beta] + J_M [\alpha(1 - \beta) + \beta] + f_{n_{d2}^*} + F_{N_{d2}^*} > 0
\]

\[
(1 - \alpha)(1 - \beta)\frac{Jc}{2n_{i}^*} + \left[\beta + \alpha(1 - \beta)\right] \frac{Jc}{2N_{i}^*} + (1 - \alpha)(1 - \beta)J_M + \left[\beta + \alpha(1 - \beta)\right] J_M + f_{n_{i}^*} + F_{N_{i}^*}
\]

Since \( N_i^* = N_{d2}^* \) the above inequality reduces to:

\[
(1 - \alpha)(1 - \beta)\frac{Jc}{2n_{d2}^*} + \theta \beta \frac{Jc}{4n_{d2}^*} + \theta \beta J_M + f_{n_{d2}^*} > 0
\]

\[
(1 - \alpha)(1 - \beta)\frac{Jc}{2n_{i}^*} + f_{n_{i}^*}
\]

Rearranging and substituting (13) and (34), the expression can be written as:

\[
2 f_{n_{d2}^*} + \theta \beta J_M > 2 f_{n_{i}^*}
\]
Since \( n_i^* < n_{d2}^* \) the last inequality is satisfied and therefore \( TSC_{d2} > TSC_i \).

Thus it can be stated that under reasonable conditions:

\[
(40) \quad TSC_{s2} > TSC_{d2} > TSC_i > TSC_{sr} > TSC_r
\]

In summary, the results obtained so far show that under likely conditions the strategy of screening all patients at the low-level services and referring to the upper levels of the pyramid only those that need these more specialized types of care is an economically sound choice whenever imperfect information is present, independently of its type. In this sense, if need is the basis upon which the health system is structured, policy makers do not have to be concerned whether patients overestimate and/or underestimate their health conditions, health planners have only to know that reality is characterized by a second-best world in which consumers do not have perfect knowledge of their health status, to move into an interventionist approach and implement a referral strategy.
III. CHOICE CONSIDERED: PLANNING BASED ON DEMAND

The previous section showed how a health care system would be organized if each individual in need of medical assistance is to have access to a service. In such a context, the number and types of services are defined according to a potential level of utilization: the planning process does not take into consideration if the facilities are actually used. In this sense, full access is defined as the assurance that all consumers have the opportunity of seeking care.

As discussed elsewhere (Iunes, 1996), the quantity of services that individuals want to consume is not likely to be same as the one regarded by health planners as needed. If that is the case, it is argued that the system is liable to present excess capacity in some areas, or levels of care, and excess demand in others. The fact that, particularly in developing countries, hospitals tend to be overcrowded and emergency services have a tendency to display long lines, while ambulatory services remain practically empty, are seen as evidence to validate the argument.

Economists reason that individuals allocate their scarce resources (which include time) in order to maximize their level of well-being, given their structure of preferences. If prices are set according to their production costs, they are the appropriate signals to be used by consumers when defining the distribution of their assets. It is in this sense that economists favor the use of prices as the relevant tool for planning, and criticize the normative approach described in the previous section as conceiving an “ideal” world in which health need is the concern of society.
This section presents a demand-based model that, even though very simple, captures the main characteristics of the location framework analyzed in this paper, and is able to derive (i.e. it does not have to assume) results that are consistent with the second-best literature.

### III.1 The Basic Demand Model

In accordance with the spatial framework used throughout this paper, it will be considered as costs to consumers not only the monetary price paid for a service but also the transportation costs incurred to reach the facility. In this sense, a person’s decision to demand health care will be determined by the level of total expenditure $e$ that he or she incurs. Recalling the definition presented in this paper’s first expression:

\[
e = p + 2c x
\]

With $p$ representing the monetary price charged, $c$ the unit cost of transportation, and $x$ the actual distance that the consumer has to travel to reach the facility, $e$ reflects round-trip costs.

In this sense, there are two monetary variables that determine the demand for medical care: the price or fee charged, and the travel expenses incurred to get to the facility. Since the price is exogenous to the consumer, the total willingness to pay for a service is expressed by the maximum distance that the patient is prepared to travel: $x_m$.

In other words, each consumer associates an “ex-ante” value or benefit with a treatment, and it is this perception that determines, for each level of price, the maximum distance $x_m$ and the maximum expenditure the person is willing to incur $e_m$:

\[
e_m = p + 2c x_m = b
\]
It is important to note that $2cx_m$ can also be seen as a measure of the (maximum) consumer’s surplus. Consider, for instance, a person living just over a facility: this consumer is, like all consumers, willing to spend $e_m$ to obtain medical attention, however, he or she will only be paying the user fee $p$. Note, though, that the observed utilization data will provide information only about the realized, or “ex-post”, cost of medical care: $e = p + 2cx$ (in the case just presented $e = p \neq e_m$). It is in this sense, that: a) demand and utilization are different concepts; and b) data analysis based solely on utilization figures tend to be biased.

It follows directly from expression (41) that:

\begin{equation}
(42) \quad x_m = \frac{b - p}{2c}
\end{equation}

It must be noted that if the consumer perceives that the benefit that he or she will derive from medical care is so small that it is not worth the costs it imposes, this person will not demand assistance even if a “real,” but perceived as minor, health problem is present. If the problem is indeed minor, health professionals are likely to regard it as a case in which no need for care existed and in this case need and demand would agree. However, it is also possible that consumers may perceive his or her case as so serious that it is helpless, or even that the system is believed to be so ineffective that it would not be able to provide any substantial benefit. In these cases the opinion of the health professional may not agree with that of the individual and demand and need will differ.\textsuperscript{37,38}

\textsuperscript{37} See also Iunes (op. cit.).

\textsuperscript{38} It is no surprise, therefore, to observe in household surveys expressive proportions of individuals stating that a health problem existed without a corresponding demand for care.
It has been shown in the previous section that the maximum distance a person is from a health clinic is $1/2n$. Thus, if $x_m \geq 1/2n$ anyone may get assistance and the model developed in Section II.1.1 apply: utilization would be equal to market demand, which would be equal to $J$.\(^{39}\) If, however, $x_m < 1/2n$ some people, those that live farther, will not be served (see Figure 7).

![Figure 7](image)

It must be noted that even when $p$ and $x_m$ are fixed, i.e. prices and unit costs of transportation remain constant, the proportion of consumers accessing the services increases as the number of health care facilities increases, since the distance between any two providers is reduced. Accordingly, the market demand for medical services is a function of three variables: the monetary price charged, the number of services and the costs of transportation. Formally:

$$q = f(p, n, c)$$

$$\frac{\partial q}{\partial p} < 0; \frac{\partial q}{\partial n} > 0; \frac{\partial q}{\partial c} < 0$$

\(^{39}\) The market demand will be equal to $J$ only if the patients that need medical assistance are the ones that seek care.
Health policy makers can, however, affect the demand for medical care only through the two variables controlled by the government: the fee charged and the number of facilities offered to the population. In this sense, $c$ is an exogenous variable assumed to remain constant. Thus:

$$q = f(p, n, c) = f(p, n)$$

$$\frac{\partial q}{\partial p} < 0 \text{ and } \frac{\partial q}{\partial n} > 0$$

The set of premises that characterize this basic model are

**Assumption 1**: the country is defined by a circle of unit circumference;

**Assumption 2**: the health system provides only one type of service, called primary care;

**Assumption 3**: there are $n$ identical health care centers or clinics evenly distributed around the circle (therefore the distance between clinics is equal to $1/n$);

**Assumption 4**: the health care facilities provide only outpatient care;

**Assumption 5**: there are $Z$ consumers, uniformly distributed around the country, demanding medical care;

**Assumption 6**: the services are government-owned.

Note that, with the exception of Assumption 5, all other hypothesis from the model described in Section II.1.1 still apply. Assumption 5 must be adjusted for the demand model because the number of patients that demand medical care is not likely to be the same as the one that would be defined by the medical professionals as actually needing attention: some patients that need medical attention may not demand care, while some

See more on this point further below.
individuals may demand assistance without an underlying medical need, which implies that $Z < J$.

The distance $x_m$ is, in this spatial model, the length of an arc on the unit circle, therefore, the market area or demand of the $k^{th}$ health clinic is given by:

$$q_k = 2Zx_m$$

Because any given health center receives patients from both sides. The potential market demand is, accordingly, defined by:

$$q = 2nZx_m$$

Replacing $x_m$ into the last expression defines the demand function for primary care:

$$q(n, p) = Zn\left(b - \frac{p}{c}\right)$$

Which satisfies the conditions expressed in (43) since:

$$\frac{\partial q}{\partial p} = -\frac{Zn}{c} < 0 \quad \text{and} \quad \frac{\partial q}{\partial n} = Z\left(b - \frac{p}{c}\right) > 0$$

The elasticities of demand with respect to the monetary price and the number of facilities are respectively:

$$\eta_p = -\frac{p}{b - p} \quad \therefore \quad \eta_p < 0 \quad \text{if} \quad p > 0; \quad \text{and} \quad \eta_n = 1$$

It can be seen that the demand for primary care will be inelastic to prices if the value of the fee is less than half the monetary equivalent of the benefit generated by the service. This result reflects the fact that out-of-pocket expenditures have become a relatively small
proportion of the total cost incurred by the individual when consuming health care and is in accordance with Acton’s (1975) findings.\footnote{It must be remembered that the consumer’s total expenditure is given by the monetary price and transportation costs. If $p < b/2$, then the fee will be less than the person’s willingness to spend with transportation (see expression (41)).} Formally, the behavior of the price-elasticity of demand is described by the following expressions:

$$\begin{align*}
\text{if } p < b/2 & \Rightarrow |\eta_p| < 1 \\
\text{if } p = b/2 & \Rightarrow |\eta_p| = 1 \\
\text{if } p > b/2 & \Rightarrow |\eta_p| > 1
\end{align*}$$

Financing agencies have been encouraging developing countries to introduce user fees in their health care systems (see for instance de Ferranti, 1985) not only for their revenue generating properties, but also as a mean of inducing a behavior by the part of consumers that is more responsive to economic factors (i.e. a more “rational” conduct). The results in (46) do seem to provide some support for the proponents of such measures in the sense that they indicate that there is a range of prices in which a fee would have a relatively small impact on access. In these cases overall demand could actually be increased if the revenue obtained from the higher prices is used to built new facilities reducing transportation costs.\footnote{With $p < b/2$, $|\eta_p| < 1$ and $|\eta_n| = 1$.} In fact, the results of Gertler, Locay and Sanderson (1987) for Peru seem to corroborate this hypothesis. Their simulations indicate that the imposition of user fees would generate significant revenues for the government and a small reduction in the total demand for health care. Furthermore, their results show that if the extra revenue generated through the fees are reinvested in order to reduce transportation
costs, a welfare loss is transformed into a welfare gain. Two very important questions remain open, however:

i. whether the revenue generated through the fees would be sufficient to promote investments that are large enough to reduce transportation costs;

ii. the analysis developed so far does not take into consideration the distributive impact of introducing user fees. The result is not clear and the discussion will be left to the next sub-section, but it could even be argued that if the poor are the most affected by large transportation costs, the equity of the system could actually be improved with the reinvestments (depending on the answer to the first question above). Gertler et al. (op. cit., p. 85), however, indicate that welfare would be distributed from the poor to the rich: “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.”

Figure 8, below, displays the schedule of the elasticity of demand with respect to prices.

Figure 8
Since the maximum distance that the representative consumer is willing to travel to get to a clinic is given by $x_m$, the average travel distance is $x_m/2$, and therefore the average transportation expense necessary to get to a clinic and return home is equal to $cx_m$. The system’s total social cost is, therefore, equal to:

$$TSC = qx_m + nf + qm$$

Since total benefits are given by $qb$, the welfare function to be maximized is:

$$(47) \quad W = qb - qx_m - nf - qm - qp + qx_m - nf - qm$$

The last result is obtained by replacing the definition of $b$ given in (41).

In developing countries, the limitations of health sectors usually come from a government health budget that is insufficient to cover expenditures. In fact, as discussed above, the introduction of positive prices for health care services in these countries is frequently justified by governments and financing agencies as a source of funding necessary to maintain the system (see for instance de Ferranti, op. cit. and Lewis, 1993). With fiscal resources that amount to $H$ the sector’s budget constraint would then be equal to:

$$(48) \quad H + pq \geq nf + qm \Rightarrow H + pq - nf - qm \geq 0$$

In its first arrangement the budget constraint is displayed with revenue sources on the left-hand side and expenditures on the right-hand side.

The problem of the government is therefore to maximize the welfare function, expressed in (47), with respect to the variables it controls, the number of facilities and the fee charged (i.e. $n$ and $p$), constrained by the health care budget presented above. In which case the appropriate Lagrangian function is:
The first-order Kuhn-Tucker conditions necessary for welfare maximization are given the following set of expressions:

\[
\frac{\partial L}{\partial p} = \frac{\partial q}{\partial p}
( p - m)(1 + \lambda) + \alpha_m \leq 0 \quad \text{if} \quad \frac{\partial L}{\partial p} < 0 \Rightarrow p = 0
\]

\[
\frac{\partial L}{\partial n} = \frac{\partial q}{\partial n}
( p - m)(1 + \lambda) + \alpha_m \leq 0 \quad \text{if} \quad \frac{\partial L}{\partial n} < 0 \Rightarrow n = 0
\]

\[
\frac{\partial L}{\partial \lambda} = q(p - m) - r + H \geq 0 \quad \text{if} \quad \frac{\partial L}{\partial \lambda} > 0 \Rightarrow \lambda = 0
\]

This maximization process allows for the determination of the optimal number of primary care clinics, the pricing rule and the marginal utility of government income:

\[
r^* = \frac{1}{f}\left[ \frac{1 + 2\lambda}{1 + \lambda} \alpha_m \right] = \frac{1}{f}\left( 1 + \frac{\lambda}{1 + \lambda} \right)\alpha_m
\]

\[
p^* = m + \frac{\lambda}{1 + \lambda} 2\alpha_m
\]

\[
\lambda = \frac{1}{2}\left( \frac{r - H}{H} \right)
\]
The optimal number of clinics is directly proportional to the total transportation costs imposed on the population and inversely related to the clinic’s investment costs. It is important to be noted that if the budget is not binding, i.e. $\lambda = 0$, and full access exists, i.e. $x_m = 1/2n$ and therefore $q = Z$, the formula for the optimal number of clinics just obtained reduces to the expression derived in the previous section (expression (6)).\(^{42}\) In other words, the needs-based model described in Section II is a special case of the demand model presented in this section. Even though no definite statement can be made, the optimal number of clinics obtained from the demand model would not be smaller than the one from the needs-based approach only if the $1+2\lambda/1+\lambda$ ratio is large enough to compensate the fact that $q < J$ and $x_m$ is likely to be less than $1/2n$. Note, however, that if the $1+2\lambda/1+\lambda$ ratio is large, the marginal utility of government income ($\lambda$) is also large, which implies a small health budget ($H$) and high prices (in order to balance the overall budget constraint). If prices are set high, the willingness to travel will be reduced, which, in (53), will tend to offset the large $1+2\lambda/1+\lambda$ ratio.

Expression (54) shows that if the fiscal budget is not sufficient to cover all the sector’s expenditures, i.e. $\lambda > 0$, fees will have to be set to levels that are greater than marginal costs. This occurs because the price or fee serves more than one purpose (see for instance Harris, 1977): not only it is used to cover (marginal) operating costs, but also part of the sector’s investment costs. This can be clearly seen by looking at expression (55), which defines the marginal utility of government income: it shows that if the budget is binding, the fiscal funds available to the government to finance the sector are not enough

\(^{42}\) For the actual optimal number of facilities to be identical in both cases it would also be necessary that $Z=J$, i.e. all
to cover the country’s needs for investment resources, or $nf > H$. As put by Baumol and Bradford (1970, p. 265):

*Prices which deviate in a systematic manner from marginal costs will be required for an optimal allocation of resources, even in the absence of externalities*. . . .

. . . [since] one is dealing with a problem in the area of the second best. We are now faced with a problem involving [social welfare] maximization in the presence of an added constraint. Resource allocation is to be optimal under the constraint that governmental revenues suffice to make up for the deficits (surpluses) of the individual firms that constitute the economy. (italics by the authors)

Equation (54) is in fact showing that prices are providing an indication of the system’s social costs, and therefore could be seen as (optimal) shadow prices.

The term $\lambda/(1 + \lambda)$ present in (53) and (54) is known as the Ramsey number (see for instance Nelson, 1982). In the pricing rule described by the last expression, the Ramsey number indicates the proportion of the consumer surplus that must be taken away from patients in order to finance the health care system. Thus, the greater the lack of resources the larger will be the proportion of the consumer surplus used to meet the sector’s need of funds. 43 In the limit, as the marginal utility of government income increases, i.e. as $\lambda \to \infty$, $\lambda/(1 + \lambda) \to 1$, and the entire consumer surplus will have to be taken away.

It can also be shown that the pricing mechanism derived by the model does satisfy the rule that the percentage deviation of price from marginal cost is inversely proportional to the elasticity of demand, and therefore the model does provide a Ramsey

---

43 Note that the Ramsey number is less than one.
pricing rule (see, for instance, Baumol and Bradford, op. cit.; Nelson, op. cit.; and Barnum and Kutzin, 1993). The price-elasticity of demand is given in (46) as:

\[ \eta_p = - \frac{p}{b - p} \]

Using the definition of \( b \):

\[ \eta_p = - \frac{p}{2\alpha_m} \]

And from (54):

\[ p - m = \frac{\lambda}{1 + \lambda} 2\alpha_m \Rightarrow \frac{p - m}{p} = \frac{\lambda}{1 + \lambda} \frac{2\alpha_m}{p} = \frac{\lambda}{1 + \lambda} \eta_p \]

The last result is exactly the condition for Ramsey pricing (e.g. Nelson, op. cit.). Thus the percentage in which the fee will be allowed to deviate from marginal cost will increase as the elasticity of demand diminishes.

Expression (53), on the other hand, tells that, in order to compensate for transportation costs, the optimal number of clinics should be proportionately more than the relation between society’s average transportation expenditures and the cost of investing in another service.

The results obtained from the model show that the government can play with the variables it controls, prices and the number of facilities, to increase social welfare. It can increase prices — in an opposite direction to the elasticity of demand — to finance more facilities and therefore reduce transportation costs.
III.1.1 Some Equity Considerations

The basic model developed above does not take into consideration equity implications related to differences in income that may exist within a given society. The purpose of this sub-section is to provide a brief analysis of some results that are derived when income differentials are introduced. This is particularly important because, as shown by the results from Gertler et al. (op. cit.) presented above, policy decisions are likely to have completely different impacts on rich and poor.

The fact that the opportunity cost of being seek differs between the rich the poor, implies that the monetary equivalent of the benefit accruing from a treatment is a function of the level of income. Thus \( b = b(y) \). At any given price level, these differences are revealed through the (maximum) distance that a person is willing to travel. Formally, expressions (41) and (42) now become:

\[
\begin{align*}
(56) & \quad p + 2\alpha_m(y) = b(y) \\
(57) & \quad x_m(y) = \frac{b(y) - p}{2c}
\end{align*}
\]

From which follows the demand equation:

\[
(58) \quad q(p, n, y) = 2Zr x_m(y) = 2Zr \left( \frac{b(y) - p}{2c} \right) = Zr \left( \frac{b(y) - p}{c} \right)
\]

With \( \partial q/\partial p < 0 \) and \( \partial q/\partial n > 0 \), as before; and \( \partial^2 q/\partial y^2 < 0 \).

Assuming that there are two income groups in society, rich and poor, the three preceding expressions are rewritten as:

\[
(56') \quad p + 2\alpha^i_m = b^i
\]
\[(57') \quad x^i_m = \frac{b^i - p}{2c}\]

and,

\[(58') \quad q^i = 2Zx^i_m = Zn\left(\frac{b^i - p}{c}\right)\]

Where \(i = r, p\) for rich and poor, respectively.\(^{44}\) It must be noted that:

\[x^r_m > x^p_m \text{ and } q^r > q^p\]

The elasticities of demand are:

\[
\eta^r_p = \frac{-p}{b^r - p}, \quad \eta^p_p = \frac{-p}{b^p - p} \\
\eta^r_n = \eta^p_n = 1
\]

Since \(b^r > b^p\), it follows that:

\[|\eta^r_p| < |\eta^p_p|\]

Thus, as expected, the poor are more “sensitive” to price changes. The fact that Gertler and his colleagues (op. cit.) found that higher prices and reinvestments produce a substantial decline in the demand of the poor and a small increase in the demand of the rich,\(^{45}\) suggests that the prices charged in Peru reached a point in which the demand of the poor became elastic with respect to prices — more than offsetting the elasticity with respect to the number of facilities (which is equal to one) — while the demand of the rich remained inelastic.

\(^{44}\) Please note that these identifiers are used as superscripts. They should not be confused with the indicator of the referral strategy used in the previous section as subscript, and the price symbol.

\(^{45}\) See the quote in p. 62.
It has been shown above that the monetary equivalent of a health benefit received by an individual is small, the demand will become elastic fairly easy. It is, therefore, important to notice that:

i. if income is small, \( b(y) \) will also be (relatively) small; and

ii. the low quality of the services provided, as is frequently the case in developing countries, may render the benefit itself (actual or perceived), small.

Since these two conditions tend to affect the poor the most (the first one by definition and the second because their options are very limited or nonexistent), their demand can be severely curtailed by an increase in the fees charged by the government, specially if the increase is not preceded by an improvement in the quality of the services offered to the population.

These results put a qualifier into the outcomes obtained in the previous subsection, they show that governments, if they have any concern with the distributive impact of their policies, have to be very careful when setting their price strategies. In fact, they do suggest that some price discrimination would be desirable, which, as shown, is in fact the essence of the Ramsey pricing strategy.
IV. A FINAL DISCUSSION AND POLICY IMPLICATIONS

Several health systems, both, in developing and industrialized countries, have been erected with the idea that the necessary resources should be allocated to respond to the health needs of the population. In this sense differentials in variables that are not related to such needs, such as income and place of residency, for instance, should not affect the access to the system. In this sense, it must be planned as to assure full access: every person in need of medical attention can, potentially, use a service. In fact, some systems go even further, assuring horizontal equity: i.e. not only everyone must have access to care, but persons with the same health needs (irrespective of other variables) must have the same level of access.

In summary, this paper recognizes the fact that, despite the opposition from economists, the need approach has been, in reality, used as the conceptual basis for many health care systems, and therefore should be subject to an economic analysis. The models constructed — based on the spatial economics literature — evaluate alternative forms of organizing health care systems based on the concept of need. These alternative scenarios result from the perception of how consumers behave due to the differences in knowledge and information that exist regarding their own health condition. The models show that the absolute and relative number and types of facilities vary substantially depending on how individuals tend to assess their health needs.

Since in these systems there are no prices to induce a desired behavior, health planners must develop alternative mechanisms, such as a referral strategy, that will assure the appropriate utilization of resources without restricting access when needed. Because
these mechanisms do generate additional social costs, the paper has examined and compared the results obtained from each of the alternative scenarios to determine the conditions that will make a given option preferred to another in an economic sense, i.e. make it less costly to society.

It has been shown that if individuals tend to underestimate and/or overestimate their health condition, an interventionist approach through a referral strategy is likely to be less costly to society than a “laissez-faire-type” policy, i.e. the model shows that health planners should intervene and organize the system in such a way that patients will use appropriately the more complex (and costly) services.

The referral strategy is no panacea, however. The international evidence suggests that most attempts made to implement referral strategies have failed, particularly in developing countries. One of the reasons for that is the extremely limited ability shown by the clinics and posts in solving most health problems, if that is indeed the case, it becomes very hard to justify the presence of gate keepers, for almost every case will have to be referred anyway. Furthermore, these facilities usually do not have the capacity to handle the extra workload imposed by the referral structure. Consequently, the implementation of a referral strategy must be proceeded by investments in primary care services that would make them more responsive to the health needs of the population. There two major implications arising from this discussion:

i. health systems will have to move away from the traditional model that perceives the primary care facilities as extremely simple units with very little resources available;

ii. the referral strategy is not the appropriate solution for very poor economies. If the health sector is organized, in terms of number and types of facilities, for a referral
system, but the strategy does not perform properly, the needs of the population and the structure of services available will conflict.

The fact that the referral structure is no solution for many countries means that a more passive approach must be followed. However, the model shows that those are actually the cases in which the understanding of the type and level of imperfect information becomes critical, for otherwise some services will end-up being overused and overburdened while others will remain underutilized, and as shown, the impact such misspecification can be substantial.

While the model of Section II evolves around the concept, so much favored by health professionals, that the health need of the population is the only variable that should determine the way in which the system is to be organized by the government; the discussion of Section III incorporates the traditional economic conception that individuals and governments have different options when allocating their resources, and health care is only one of them. Which means that its (private and social) costs and benefits are going to be weighted against those of other activities, and failing to realize that will lead to resources being improperly allocated. In this context prices serve are the appropriate signals to be used by consumers and suppliers when taking their decisions.

The results obtained from the demand model of Section III show that the needs-based approach can be seen as a special case of the demand model, where access and budgetary constraints do not exist. The Ramsey pricing rule derived shows that prices serve two objectives: to recuperate the costs of production and to cover for the deficits in the government (investment) budget.
It is important to note that demand is an “ex-ante” concept and should be contrasted with the “ex-post” notion of utilization. In this sense, a consumer might be willing to travel a distance equal to \( x_m \) to get medical care, however, if the actual distance to clinic, \( x \), is greater than that, this person, even though wanting, will not use the service.

According to this basic model, the demand would be inelastic to price changes if the fee charged represents less than half the expected benefit provided by the service, thus implying that higher prices may have a small impact on the demand for health care. In fact, it is shown that it is theoretically possible to conceive a price range that would produce an increase in overall demand, if the extra revenue generated through higher fees is used to expand the number of facilities and, therefore, reduce transportation costs.

Proponents of the needs approach argue, on the other hand, that the demand framework is inequitable and would have a disproportionately negative impact on the poor, a point that seems to have been substantiated by empirical studies on the impact of user fees in developing countries. Sub-section III.1.III.1.1 examines the validity of this reasoning. The analysis shows that the poor do present a demand that is much more sensitive to price changes than the rich, which means that prices are likely to, indeed, have an inequitable impact on health care consumption.

Note that the demand model assumes an uniform spatial distribution of the rich and the poor. If, however, the latter tend to live farther away from the services, the actual utilization of these facilities by the poor will be significantly smaller than the (potential) demand described by \( q^p \), i.e. in practice the access of the low income populations to health care will be even more restricted. Furthermore, if the rich tend to live closer, their consumer surplus is likely to be generally higher than the poor. It must be, therefore, clear
that an uncritical use of the demand approach may have important distributive consequences.

These considerations do suggest that some form of price discrimination is desirable. In fact, if the poor and the rich have different price elasticities, the Ramsey rule, shown to apply to the model, assures that price discrimination will also be optimal.

Thus, while health planning based on the concept of need may be limited by the fact that it conceives an “ideal” world, as if health were the only concern of individuals and governments (thus society); health planning based on the traditional concept of demand also falls in a similar trap of an “ideal” world, here symbolized, for instance, by the representative consumer. The development of the field of health economics has shown that many economic concepts have to be adjusted or “reconstructed” in order to be able to capture the special characteristics of the health sector, in this sense, in the need versus demand debate does not necessarily require rejecting one concept for the other, but rather to incorporate the idea of need (and equity) — issues that are particularly important for the reality of developing countries — into economic modeling. The recent work developed by Thomas Rice (1992), pointing out the limitations of the traditional welfare loss analysis, goes in this direction. The discussion that emerged from Rice’s paper (see Rice, 1993a, 1993b; Feldman and Dowd, 1993; and Peele, 1993), can be viewed (in a negative sense) as a reaction from the paradigm, or more optimistically as the revival of a fruitful debate.
APPENDIX

This appendix will present two proofs:

a) first, it will be shown that, in a world characterized by two-level misinformation, social costs will be greater when the transfer of patients from the clinic to the nearest hospital has to be scheduled to another date instead of being done directly from one facility to the other, i.e. $TSC_{s_2} > TSC_{d_2}$; and

b) it will also be proven that the average distance traveled by a patient from a clinic to the nearest hospital (referral distance) is equal to $1/4N$, when $n/N$ is an even integer, and equal to $(1/4)[1/N - N/n^2]$, when $n/N$ is an odd integer.

A.

For $TSC_{d_2} < TSC_{s_2}$ it must be that:

$$(1 - \alpha)(1 - \beta)\frac{Jc}{2n_{d_2}^*} + \alpha(1 - \beta)\frac{Jc}{2N_{d_2}^*} + (1 - \theta)\beta\frac{Jc}{2N_{d_2}^*}$$

$$+ \theta\beta Jc\left(\frac{1}{4n_{d_2}^*} + \frac{1}{2N_{d_2}^*}\right) + Jn\{(1 - \alpha)(1 - \beta) + \theta\beta\} + JM\{(1 - \alpha) + \beta\}$$

$$+ f_{n_{d_2}}^* + FN_{n_{d_2}}^* < (1 - \alpha)(1 - \beta)\frac{Jc}{2n_{s_2}^*} + \alpha(1 - \beta)\frac{Jc}{2N_{s_2}^*}$$

$$+ (1 - \theta)\beta\frac{Jc}{2N_{s_2}^*} + \theta\beta Jc\left(\frac{1}{2n_{s_2}^*} + \frac{1}{2N_{s_2}^*}\right)$$

$$+ Jn\{(1 - \alpha)(1 - \beta) + \theta\beta\} + JM\{(1 - \alpha) + \beta\} + f_{n_{s_2}}^* + FN_{n_{s_2}}^*$$

Since $N_{d_2}^* = N_{s_2}^*$ the inequality simplifies to:

$$\left[(1 - \alpha)(1 - \beta) + \frac{\theta\beta}{2}\right]\frac{Jc}{2n_{d_2}^*} + f_{n_{d_2}}^* < \left[(1 - \alpha)(1 - \beta) + \theta\beta\right]\frac{Jc}{2n_{s_2}^*}$$

Substituting (34) and (38) into the expression, the following result is obtained:
\[ \frac{\sqrt{2(1-\alpha)(1-\beta)+\theta\beta}}{2} < \frac{\sqrt{2(1-\alpha)(1-\beta)+\theta\beta}}{2} \quad \text{q.e.d.} \]

**B.**

**B.1 \( n/N = \text{even integer} \):**

Consider, as an illustration, Figure A1.1, below. It displays the case in which \( n = 16, N = 2 \) and therefore \( n/N = 8 \).

![Figure A1.1](image)

In locations 1 and 9 there are hospitals and clinics, therefore the referral distance in these sites is equal to zero. In all other locations there are only health care centers. Table A1.1, below, displays the referral distance from each site.

The total distance traveled during referrals is obviously the sum of all the distances shown at the table. Two results appear immediately: i. in \( N \) locations the referral distance is zero (the minimum distance); and ii. there are also \( N \) mid-points in which the referral distance is equal to \( 1/2N \) (the maximum referral distance). In between these minimum and maximum referral distances, there will be patients traveling at least \( 1/n \), and
at most $3/n$ from both sides of each hospital. The numerator ($3$) of this last result ($3/n$), is expressed in general notation form by the formula: $(1/2)(n/N) - 1$.

Table A1.1
REFERRAL DISTANCES FOR $n=16$ and $N=2$

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>REFERRAL DISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$1/16 = 1/n$</td>
</tr>
<tr>
<td>3</td>
<td>$2/16 = 2/n$</td>
</tr>
<tr>
<td>4</td>
<td>$3/16 = 3/n$</td>
</tr>
<tr>
<td>5</td>
<td>$4/16 = 4/n = 1/4 = 1/2N$</td>
</tr>
<tr>
<td>6</td>
<td>$3/16 = 3/n$</td>
</tr>
<tr>
<td>7</td>
<td>$2/16 = 2/n$</td>
</tr>
<tr>
<td>8</td>
<td>$1/16 = 1/n$</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>$1/16 = 1/n$</td>
</tr>
<tr>
<td>11</td>
<td>$2/16 = 2/n$</td>
</tr>
<tr>
<td>12</td>
<td>$3/16 = 3/n$</td>
</tr>
<tr>
<td>13</td>
<td>$4/16 = 4/n = 1/4 = 1/2N$</td>
</tr>
<tr>
<td>14</td>
<td>$3/16 = 3/n$</td>
</tr>
<tr>
<td>15</td>
<td>$2/16 = 2/n$</td>
</tr>
<tr>
<td>16</td>
<td>$1/16 = 1/n$</td>
</tr>
</tbody>
</table>

Accordingly, the total referral distance amounts to:

\[
(A1.1) \quad 0N + N \frac{1}{2N} + 2N \left[ \sum_{k=1}^{(1N-1)/2N} \frac{k}{n} \right] = \frac{1}{2} + 2N \left[ \sum_{k=1}^{(1N-1)/2N} \frac{k}{n} \right]
\]

The average referral distance is simply expression (A1.1) divided by $n$:

\[
(A1.2) \quad \frac{1}{2n} + \frac{2N}{n} \left[ \sum_{k=1}^{(1N-1)/2N} \frac{k}{n} \right]
\]

Since a summation from $1$ to $z$ is equal to $[(1+z)/2]z$, it is possible to write:
\[
\left( \sum_{k=1}^{\left\lfloor \frac{n}{2N} \right\rfloor} \frac{k}{n} \right) = \frac{1 + \left( \frac{n}{2N} - 1 \right)}{2} \left( \frac{n}{2N} - 1 \right) = \frac{n - 2N}{8N^2}.
\]

The result of substituting (A1.3) into (A1.2) is:

\[
\frac{2N}{n} \left( \frac{n - 2N}{8N^2} \right) + \frac{1}{2n} = \frac{n - 2N}{4Nn} + \frac{1}{2n} = \frac{n - 2N + 2N}{4Nn} = \frac{1}{4N}.
\]

**B.2. n/N = odd integer:**

Again it is convenient to start with an example. Table A1.2 displays the same type of information as the previous one for the case in which \( n = 14, N = 2 \), and therefore \( n/N = 7 \).

**Table A1.2**

<table>
<thead>
<tr>
<th>LOCATION</th>
<th>REFERRAL DISTANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1/14 = 1/n</td>
</tr>
<tr>
<td>3</td>
<td>2/14 = 2/n</td>
</tr>
<tr>
<td>4</td>
<td>3/14 = 3/n</td>
</tr>
<tr>
<td>5</td>
<td>3/16 = 3/n</td>
</tr>
<tr>
<td>6</td>
<td>2/16 = 2/n</td>
</tr>
<tr>
<td>7</td>
<td>1/16 = 1/n</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1/16 = 1/n</td>
</tr>
<tr>
<td>10</td>
<td>2/16 = 2/n</td>
</tr>
<tr>
<td>11</td>
<td>3/16 = 3/n</td>
</tr>
<tr>
<td>12</td>
<td>3/16 = 3/n</td>
</tr>
<tr>
<td>13</td>
<td>2/16 = 2/n</td>
</tr>
<tr>
<td>14</td>
<td>1/16 = 1/n</td>
</tr>
</tbody>
</table>
It can be seen from the table that the main difference between the two cases is that, if \( n/N \) is an odd integer, there are no health care centers located at mid-point between any two hospitals.

As a result, expression (A1.1), that described the total referral distance, now becomes:

\[
\begin{align*}
(A1.4) & \quad 0N + 2N \left( \sum_{k=1}^{(n-1)} \frac{k}{n} \right) = 2N \left( \sum_{k=1}^{\frac{n}{2}} \frac{k}{n} \right) \\
& = 2N \left[ \sum_{k=1}^{n} \frac{k}{n} \right] \\
& = \sum_{k=1}^{n} \frac{k}{n} \\
& = \frac{2N}{n} \left[ \sum_{k=1}^{n} \frac{k}{n} \right] \quad \text{(A1.5)} \\
& = 2N \left( \frac{n^2 - N^2}{8N^2} \right) = \frac{n^2 - N^2}{4Nn} = \frac{1}{4} \left( \frac{1}{N} \right) \left( \frac{N - n^2}{n^2} \right) \quad \text{q.e.d.}
\end{align*}
\]
REFERENCES


Iunes, R.F. 1996. “Need, Demand and Equity in Health Care,” paper prepared for the Takemi Program in International Health, Harvard University, Boston.

