The financing of public hospitals: 
a study of possible alternatives

Silvia Montoya
Manuel Willington

Córdoba, Enero de 1998
The financing of public hospitals: a study of possible alternatives

Silvia Montoya and Manuel Willington

Over recent years, the provinces’ need to rationalize their public expenditure has heightened the problem of how to finance public hospitals. In addition to their existing problems of administrative efficiency and own resource allocation, they are now suffering pressure stemming from the crisis in the health schemes, whose poorer members have had to resort to the public hospitals, thus increasing demands on their budgets.

These financing problems are being met jointly by the provinces and the federal government through cost recovery laws and a decree for hospital self-management. This legislation, although aimed in the right general direction (it enables hospitals to use their resources with a certain degree of autonomy and to charge for services when they are provided to those able to pay for them) is insufficient, even without taking into account the problems in its implementation.

After reviewing the Argentine Health system and the main features of its development over the last few years, the paper carries out a theoretical analysis of the different resource transfer mechanisms between the government and hospitals, and the incentives generated by these mechanisms. No doubt, under a fixed budget together with a fixed wage payment program for staff, in which available resource do not depend on the hospital’s financial results, the incentives generated to harder work and greater efficiency are minimal or absent. The main conclusions are drawn from an agency model in which the government (as principal) must design both the fund transfer mechanism to the hospital (as agent) and the payment program for its staff.

The current restrictions in public sector hiring programs and the information requirements necessary would preclude implementation at present of prospective payment programs associated with diagnosis (as theory demands, and as already put into practise by other countries). However, there are intermediate solutions, such as budgets linked to management commitments (allowing hospital resources to keep in step with financial results, thus achieving a considerable improvement in global resource allocation), and these have better chances of proper implementation.
The financing of public hospitals: a study of possible alternatives

Silvia Montoya and Manuel Willington

1. Introduction

Among the proposed modifications of the Argentine health system, the change in public hospital financing - introduced in 1993 with a national decree for hospital self-financing - is one of the issues generating most discussion.

General public sector organization, and that of hospitals in particular, has generated few incentives toward efficiency, and the proposed change toward self-financing implies major cultural changes for hospital administrators, physicians and patients, which should be backed with regulatory action, introducing appropriate incentives for these agents.

According to international evidence, it is possible to obtain similar type of results in different health systems with different expenditure levels and different institutional organization. This has generated growing interest in the structure, incentives, viability and results of different health systems, which has been reflected in a considerable amount of theoretical and empirical literature attempting an explanation of the differences in the results and the necessary trade-offs according to the organization of the system.

One of the aspects deserving major attention is health service financing; not only because of who pays, but also regarding how much is paid and how payment is made. All these dimensions of the chosen payment system affect health system operations on two levels that should be used to guide the analysis of all sectoral reform: that of efficiency

1 This paper has been published in the Revista de Análisis Económico, vol.11, N°2, November 1996.
(referred to global costs, quality and incentives generated) and that of equity (referred to the possibility of effective access for different population sectors).

The purpose of this study is to carry out an analysis of the financing of public hospitals that, despite being only partial due to the number of problems faced by the Argentine health sector, enables conclusions to be drawn from theoretical and empirical evidence, and suggests changes in the financing mode, improving the efficiency and equity of the public sub-sector. For the purpose, first of all the organization of this sector in the country is briefly described, with particular insistence on the organization of the public sector and public hospitals. This is followed by a theoretical review of the main payment mechanisms with their advantages and disadvantages and the optimal mechanism for hospitals is drawn from a theoretical model. Finally, the last section gives policy recommendations for Argentina.

2. The health sector and public hospitals in Argentina

Health expenditure in Argentina is around 7.2% of GDP\(^2\). The public sector contributes 22% of the total expenditure, social security 36% and the private sector 42%. Within the public sector, 75% of direct expenditure is made by the provincial governments, 16% by municipalities and only 9% by the federal government. Beyond the amount of the expenditure involved, the efficiency of health expenditure in the different sub-sectors is cast in doubt.

From the viewpoint of service production, three kinds of suppliers may be identified: the public sector (including the national government, the provinces and the municipalities, with varying participation), private suppliers and a third group - of lesser relative importance - corresponding to some obras sociales (health insurance institutions of the social security system) supplying services on their own account.

As shown in Diagram 1, there are three major financing entities that also have their own regulations (mainly in the case of the provincial public sectors). These are the public

\(^2\)IRDB (1996)
sector (national, provincial and municipal), the Social Security and the private sector. Financing originates from three sources: national, provincial and municipal taxes (in the case of the public sector), social security contributions and the direct purchase of insurance plus the co-payments in the case of the private sector.

Given the number of intervening agents (and despite the relatively clear definition of the users of the different sub-systems), in practice there are problems of overlapping functions, as well as the domestic geographical differences in coverage and quality of services. These differences occur between the different sub-systems and within each one of them.

Diagram 1
Key Participants in the Health Care System in Argentina

2.1. The Social Security sub-sector

The magnitude of the expenditure by the obras sociales system (2% of GDP) and the size of the beneficiary population (around 60%)\(^3\) show the importance of its operation for population welfare.

There is much controversy within the obras sociales system about the quality of its services and its cost to society. Actually, its lack of effectiveness has led the low-income beneficiary population to resort to public hospitals (thereby generating pressures on the sector), while the higher income beneficiaries enjoy a reasonably good health coverage\(^4\) from the private system. By 1992, almost 30% of the population covered by obras sociales had some other kind of additional coverage\(^5\).

The roughly 300 national obras sociales concentrate more than 8 million beneficiaries. To these must be added 24 provincial obras sociales with some 4 million beneficiaries who are not covered by the national laws\(^6\), and the Integral Medical Assistance Program (PAMI) with a little over 4 million beneficiaries\(^7\).

Membership is obligatory according to activity sector with a 5% (employer) and 3% (employee) payroll contribution. Within each activity sector the obra social is usually split in two (one for the management staff, and the other for blue collar workers).

Sixty percent of national obras sociales’ beneficiaries belong to the unionist type, thirty percent belong to mixed administration schemes\(^8\) and eight percent to those of

---

\(^3\) Flood, C. Harriage, M. Lerner L. and L. Montiel (1994)

\(^4\) In 1992, some 27% of consultations financed by the public sector were made by individuals with obra social coverage.


\(^6\) Except for cases in which they explicitly adhere. In the other situations the calculation method for contributions differs from that of the private sector.

\(^7\) PAMI is more or less similar to the MEDICARE system in the USA.

\(^8\) The obras sociales are those in which administration is shared by beneficiaries, management and the State. In most of the cases they add to the funds obtained from wage contributions other from other sources, usually calculated as a percentage of the output “sold” by the sector. Thus, for example, in the case of the banking sector other funds are obtained as a percentage (2%) of interests charged.
management personnel. This quick classification enables one to appreciate that despite the relatively low contribution per member of the unionist scheme, these concentrates a significant amount of economic power, bearing in mind that the 1995’s income of the national obras sociales was some 3 billion pesos.

Due to the fact that each obra social groups workers from one particular industry or sector of activity, it shows very little inner income variance. The low wage dispersion per branch is reinforced by the segmentation that occurs within each branch between the workers and the management staff. Because of this grouping and of the way in which the obras sociales’ income is determined, important differences exist in the per capita income and, then, in the medical services quantity and quality.

In this situation, there are obras sociales that are able to finance a good and comprehensive range of services for their members, and others that are not even able to offer minimal coverage. The way in which membership of the system is determined makes almost half of the 200 syndicate obras sociales group less than 1,000 beneficiaries. This indicates the inefficiency deriving from conditions for membership, which are totally unrelated to the minimal threshold enabling an adequate supplying of services.

As a rule, health services are not supplied directly. Instead, obras sociales usually hire private suppliers and clinics under the system of fee for service, according to a service catalog agreed between the obras sociales and the suppliers. Co-payments are usually the rule (both in outpatient and hospitalization services), including the payments for medication. In the last two years, both the unionist obras sociales and PAMI began to sign per-capita-type contracts, not only with medical suppliers, but also with public hospitals.

This sector has experienced a major number of regulatory changes, that was completed with the entry in force of the law providing for the free choice of the health insurance agent. This freedom of choice will spawn greater competition between obras sociales for new members who, on their part, will insist on a greater emphasis on service quality when contracts are negotiated with medical suppliers, including public hospitals.
2.2. The private sub-sector

Argentina’s private sector consists of the so-called “pre-paid medicine” firms (with high-income membership) - a very small sector of health insurance companies - plus mutual benefit societies and clinical hospitals grouped together in closed, HMO\(^9\)-type systems. Direct private sector expenditure must be added to this. The sector has two basic regulatory frameworks: the insurance companies are regulated by the corresponding Insurance Law, while the “pre-paid medicine” firms are regulated by the laws for commercial contracts.

Some 200 “pre-paid medicine” firms exist, covering over 2 million people, at a cost of 1.5 billion dollars a year. The average expenditure per beneficiary is US$ 54 per month, of which approximately 30% corresponds to administrative expenditure\(^10\). Over recent years, the system has developed heterogeneously with growing competition.

The private sector basically serves the sector of the population with the highest purchasing power, but has also begun to cater for other sectors, among other things due to the deterioration of service quality in the obras sociales system. Because of the system’s captive membership feature, double membership\(^11\) situations are very common.

2.3. The public sub-sector

The public sub-system - through the national government, the provinces and the municipalities - is in charge of the functions of health promotion and preventive action, as well as providing services through hospitals and health care units. In recent times, the sub-system has tended toward a progressive decentralization (in key with general health policy), and this tendency has accelerated over the last ten years. Due to this, although the

\(^9\) Health Maintenance Organization.
\(^10\) The issue of administrative expenses is a problem shared by all the insurance companies.
\(^11\) According to IRDB estimates (1996), Argentina is one of the Latin American countries in which the purchase of private health services is least correlated with family group income level.
laws governing health policies are federal, more and more policies are being put into effect by the provinces. In practice, primary medical attention has historically been provided both by the specific establishments for the purpose (dispensaries and low complexity hospitals) and by high complexity hospitals, which doubtless represents a high degree of inefficiency from the viewpoint of resource allocation.

The role of the public sector at the provincial level is not merely restricted to preventive programs - the main programs of this kind were transferred from the national government to the provinces in 1992\(^{12}\) - and to hospital availability, but also, thanks to the existence of provincial \textit{obras sociales} for State employees (such as the provincial social security sub-system), the provinces care for an important percentage of their own population either directly (through the labor relationship of wage-earners), or indirectly\(^{13}\)\(^{14}\).

Since every province is able to exercise self-determination of its health policies, there are important regional differences. Most of them divide their territory into health zones, but their management and their share in the running of the health system is totally different. In this sense, an aspect clearly illustrating the policy differences is the degree of decentralization.

The 24 provincial public sectors are currently faced with service reductions due to strikes and a lack of medicine and other inputs, to which must be added the semi-official tariff system applied through cooperatives. Provincial health policies are static, limited to maintaining the \textit{status quo}, and have very little leeway for maneuvering. Thus, the measures adopted by the provinces to face their financing crisis over the last two years

\(^{12}\) Each province is responsible for the primary treatment of its inhabitants, although policies are debated and coordinated in the Federal Health Council.

\(^{13}\) In some provinces more than half of all employment is public, so over 60% of the total populations depends either on the State.

\(^{14}\) Even though from the legal viewpoint the entire population are beneficiaries, in practice, public sub-sector users are the lowest income sectors, and are the groups covered by \textit{obras sociales} without effective access to all of their services. In 1992, almost 37% of persons treated in the public sector were from the lowest quintile of income distribution; while only 5% of these consultations were made by persons from the top quintile (Flood, C., Harriage, M., Lerner, L., and L. Montiel, 1994)
concentrated on trimming down excess staff, controlling working hours, eliminating incompatibilities, etc.

Table 1
Public Health System Performance Indicators

<table>
<thead>
<tr>
<th></th>
<th>I Occupancy Coefficient</th>
<th>II Average Stay</th>
<th>III Bed Turnover Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal District</td>
<td>0.83 0.83</td>
<td>14.53 14.53</td>
<td>20.77 20.77</td>
</tr>
<tr>
<td>Buenos Aires</td>
<td>0.63 0.42 0.68 0.64</td>
<td>10.45 40.36 11.60 8.41</td>
<td>22.16 3.78 21.26 27.65</td>
</tr>
<tr>
<td>Catamarca</td>
<td>0.29 0.29</td>
<td>4.21 4.21</td>
<td>25.30 25.30</td>
</tr>
<tr>
<td>Córdoba</td>
<td>0.67 0.69 0.67 0.48</td>
<td>16.70 6.26 19.00 4.06</td>
<td>14.57 39.92 12.87 42.65</td>
</tr>
<tr>
<td>Corrientes</td>
<td>0.57 0.57</td>
<td>6.50 6.50</td>
<td>32.01 32.01</td>
</tr>
<tr>
<td>Chaco</td>
<td>0.48 0.48</td>
<td>4.51 4.51</td>
<td>38.69 38.69</td>
</tr>
<tr>
<td>Chubut</td>
<td>0.50 0.50</td>
<td>6.95 6.95</td>
<td>26.52 26.52</td>
</tr>
<tr>
<td>Entre Ríos</td>
<td>0.50 0.50</td>
<td>9.46 9.46</td>
<td>19.13 19.13</td>
</tr>
<tr>
<td>Formosa</td>
<td>0.44 0.44</td>
<td>4.87 4.87</td>
<td>33.07 33.07</td>
</tr>
<tr>
<td>Jujuy</td>
<td>0.55 0.55</td>
<td>7.89 7.89</td>
<td>25.40 25.40</td>
</tr>
<tr>
<td>La Pampa</td>
<td>0.38 0.38</td>
<td>5.91 5.91</td>
<td>23.67 23.67</td>
</tr>
<tr>
<td>La Rioja</td>
<td>0.31 0.31</td>
<td>4.83 4.83</td>
<td>23.12 23.12</td>
</tr>
<tr>
<td>Mendoza</td>
<td>0.66 0.67</td>
<td>7.31 7.31</td>
<td>33.17 33.22</td>
</tr>
<tr>
<td>Misiones</td>
<td>0.64 0.64</td>
<td>5.05 5.05</td>
<td>46.17 46.17</td>
</tr>
<tr>
<td>Neuquén</td>
<td>0.47 0.47</td>
<td>4.52 4.52</td>
<td>38.23 38.23</td>
</tr>
<tr>
<td>Rio Negro</td>
<td>0.49 0.49</td>
<td>5.13 5.13</td>
<td>34.63 34.63</td>
</tr>
<tr>
<td>Salta</td>
<td>0.58 0.58</td>
<td>6.96 6.96</td>
<td>30.31 30.31</td>
</tr>
<tr>
<td>San Juan</td>
<td>0.54 0.54</td>
<td>6.31 6.31</td>
<td>31.48 31.48</td>
</tr>
<tr>
<td>San Luis</td>
<td>0.45 0.45</td>
<td>5.18 5.18</td>
<td>31.37 31.37</td>
</tr>
<tr>
<td>Santa Cruz</td>
<td>0.40 0.42</td>
<td>6.65 6.71</td>
<td>21.32 22.85</td>
</tr>
<tr>
<td>Santa Fe</td>
<td>0.63 0.50 0.58 0.93</td>
<td>8.50 8.85 8.50 8.49</td>
<td>26.86 20.61 24.95 39.97</td>
</tr>
<tr>
<td>Stgo. del Estero</td>
<td>0.39 0.39</td>
<td>5.33 5.33</td>
<td>26.50 26.50</td>
</tr>
<tr>
<td>Tucumán</td>
<td>0.56 0.56</td>
<td>7.51 7.51</td>
<td>27.48 27.48</td>
</tr>
<tr>
<td>Tierra del Fuego</td>
<td>1.06 1.06</td>
<td>5.51 5.51</td>
<td>0.29 70.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>I Occupancy Coefficient</th>
<th>II Average Stay</th>
<th>III Bed Turnover Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.54 0.53 0.53 0.72</td>
<td>7.12 18.49 6.95 8.87</td>
<td>30.09 21.43 30.37 32.81</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.16 0.11 0.15 0.17</td>
<td>3.02 15.50 3.10 3.73</td>
<td>10.85 14.76 11.03 8.99</td>
</tr>
<tr>
<td>Variation Coefficient</td>
<td>0.30 0.21 0.29 0.24</td>
<td>0.42 0.84 0.45 0.42</td>
<td>0.36 0.69 0.36 0.27</td>
</tr>
</tbody>
</table>

Indicators calculated as: I = Patients-days / (nº of available beds * 365); II = Patients-days / Discharges; III = Discharges / nº of available beds.


2.3.1. Public hospitals

The provinces - and Buenos Aires city municipality (MCBA) - concentrate almost 80% of the establishments with hospitalization facilities and 65% of those without (the
establishments mostly providing health care services are normally run by the municipalities).

There are some 1,200 hospitals with hospitalization facilities (around 78,000 beds) in the country. This implies that the public sector has 2.3 beds available for each 1,000 inhabitants (this represents approximately 60% of the total supply of hospital beds in the country)\(^\text{15}\). The private sector possesses some 2,000 establishments, mostly in the major cities, with around 50,000 beds\(^\text{16}\).

The *provincial* public establishments (including MCBA) have, on average, a yearly occupancy of only a little over half their capacity\(^\text{17}\). As with all averages, these figures hide major differences between the jurisdictions.

The organizational model rests on major structures that cover the range from primary care up to high-complexity services. There is no explicit system for deriving patients, hospital doctors deciding instead which patients should be interned and for how long, and what laboratory analyses should be made.

The allocation of provincial health expenditure among hospitals is done, in most provinces, according to historical patterns. Actually, the central government does not transfer a whole budget for the hospital to administrate, but rather pays wages and buys inputs itself\(^\text{18}\). This leaves the hospital administrators a rather small leeway to administrate their resources and compete in the market. Diagram 2 shows the scheme of public hospital relationships.

In practice, conditions are set for competition among hospitals, since patients are free to choose the hospital they attend. However, on the public hospital side there have

\(^{15}\) Data for the developed world (Chernichovsky, 1995) show that the number of physicians and beds per 1,000 people is 2.5 and 8.5, while the average for low-income countries (excluding China and India) is 0.09 and 2.3.

\(^{16}\) For the global system, the relationship is 3.8 beds per 1,000 inhabitants and 335 inhabitants per physician.

\(^{17}\) Each bed is used by 27 patients, with each hospitalization lasting around 8 days according to data from MSAS (1995). The low average compared with international levels suggests that the data are not very trustworthy. Average days of hospital stay, however, do vary significantly between countries. For 1988 the average stay was 13 days in Spain, 9 in the U.S., 15 in the U.K., 19 in Sweden and 17 in Germany.

\(^{18}\) This is an important aspect, since expenditure on staff represents some 80% of total expenditure.
been no incentives to obtain new patients and compete, since their State funding is unrelated to the number of patients treated or, for that matter, to any other measurement of output\textsuperscript{19}.

The system of payments to physicians and administrative staff tends to aggravate the problem to a certain extent, since staff work for a fixed wage. Thus, their income does not depend on the number of patients treated either, nor on staff efforts to identify those patients able to pay (either on their own, or through a third party).

Under these conditions, no incentives exist to increase income or control costs either on the part of the directors of the establishment, nor for physicians or administrative staff.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{public_hospital_intersectoral_relationships.png}
\caption{Public Hospital’s Intersectoral Relationships}
\end{figure}

\textsuperscript{19} On the other hand, if patients cross jurisdictions, no compensation mechanisms are available. The most serious problem of this kind occurs between the Federal District and Greater Buenos Aires.
2.3.2. Changes in the regulatory framework

Hospital self-financing

The National Executive Decree of 1993 setting up Self-Financing Public Hospitals includes, among other changes, the invoicing of services for which cost recovery is possible. Within this context, and to the extent to which staff are able to obtain marginal utilities, one might expect competition to increase among service suppliers, even within each level of government, thereby transforming hospitals into entities with administrative autonomy.\(^\text{20}\)

Before the aforesaid decree, some provinces had begun to convert their public hospitals into autonomous decentralized entities, but they did not produce the changes

---

\(^{20}\) To complement the reform, a Self-Financing Hospital Services Catalog was published with tentative prices for different services. In practice, these prices may be modified upwards or downwards.
necessary to enable the hospital administrators to manage their resources according to needs. Among other things, no clear criteria were established for allocating budgets nor for distributing funds obtained through the invoicing system\textsuperscript{21}.

Even though the changes introduced are correctly aimed, they are still insufficient. The inherent inefficiencies of public offices persist, there is little or no staff incentivation, there is meager information on service costs, hospitals of differing complexity do not provide complementary services, etc.

To these “historical” problems another of no lesser importance should be added: due to the signing of contracts between the hospital and the obras sociales (or other insurers) there is a loophole for discrimination against persons without payment capacity or insurance coverage (end beneficiaries of the public hospitals), and for discrimination in favor of those whose treatment represents a certain utility for the hospital (be this through some kind of co-payment or through the subsequent collection of service payments from the insurer). This kind of problem, associated with the existence of multiple paying agents, occurs because there is no mechanism guaranteeing treatment for the population without coverage. This lack of regulations may be due, among other things, to the assumption that there is an excess of supply in the public sector subsystem.

The consequences from equity viewpoint is generated by a fixed-payment budget mechanism, added to the meager incentives created for efficiency, show the need to restate the relationship between the State (as regulator and insurer of persons without coverage), the public hospitals and their professional and administrative staff.

3. Mechanisms for hospital financing

A study of the most efficient payment mechanism for payments from the financial backer to the supplier in the health sector has some special features distinguishing it from

\textsuperscript{21} The experience of provincial governments is quite varied. An attempt was made in some provinces to recover certain costs, but when the funds were absorbed by the provincial budget, there was no incentive for either physicians or administrative staff, since these costs did not alter wages nor the funds allocated to hospitals.
other “industries”: it is extremely difficult to measure output; the consumer’s role is
different to that in other markets; there are important information asymmetries between the
doctor and his patient; there is no possibility of storing the product, and the stochastic
nature of the demand for hospital beds leads inescapably to an excess capacity of available beds.22

The different regulations on hospital financing affect the efficiency, the degree of
equity in access to services and the consequences of the functioning of any health system.
Due to this, the reviewing of the topics of health treatment expenditure and hospital
efficiency leads to a restatement of the optimal financing mechanisms to link payers and
hospitals and trigger efficient consumption patterns. Following this line of thought, there
are studies analyzing the incentives and effects generated by the different forms of linkage
among third-party payers and the suppliers of health services (covering vertical
integration, service-package contracts, prospective payments associated with diagnoses,
refunds for services rendered, etc.) and the use of mechanisms for cost-containment (co-
payments and deductibles, second opinions, hospitalization authorizations, ex-post
reviews, etc.).

It is, in fact, an agency problem between a payer (the State or the insurers) and the
suppliers (physicians or hospitals), all with different target-functions and among whom - as
has already been pointed out - there are important information asymmetries. The classical
agency problem, in this case, is compounded by the presence of consumers, whose
relationship (both with the insurers and with the health service suppliers) also features the
asymmetries mentioned above.

Meanwhile, consumer behavior is affected by the type of payment linking payers
and suppliers. This is not only due to the different cost-sharing mechanisms - co-payments
and deductibles - , but also to the non-monetary costs related to treatment (waiting and
traveling time, etc., associated with the possible lack of incentives faced by the staff) and
to the incentives that physicians may have for inducing demand.

So as to be able to analyze the incentives generated by the different payment mechanisms, it is necessary to differentiate between the target functions involved. Theoretical and empirical studies with different purposes involve utility functions of physicians, hospital managers, financial backers (*trustees*) or of the hospital as a whole. The utility functions involved for physicians have as arguments their income, service quality provided (for customer satisfaction or reasons of prestige) and some parameter related to the effort made.

### 3.1. Payment mechanisms and incentives

First, it is necessary to clearly explain the effects on incentives of different payment mechanisms and their potential advantages and disadvantages from a theoretical point of view. The analysis considers the “pure” payment systems (refunding of costs and prospective payments) and the different variations of the same that are usually found.

The need to generate the right incentives for public hospitals by reformulating the regulatory framework and the funds transfer system excludes from the analysis the risk division mechanisms on the demand side (co-payments, deductibles, periods without coverage, etc.). This omission may be justified for two reasons: on one hand, the public sub-system beneficiaries are, in general, people of meager resources, that limits the chances of implementing mechanisms to transfer the financial risk to beneficiaries and, on the other hand, as hospitals are public, and all their staff is employed by the State, this is an

---

23 Different papers propose alternative utility functions. In McGuire and Pauly (1991), the physicians’ utility function depends on their income, their leisure time and (negatively) on the degree of demand induction; in Breyer and Schneider (1992), the utility depends on the quality of services and, also positively, on an inefficiency measurement explained by the effort made (they assume that the physician’s monetary income is a fixed wage); in Hodgkin and McGuire (1994) the arguments are the utilities and the intensiveness of treatment; in Wedig (1993) the monetary utility and the patients’ welfare; and in Ligon (1994) patient recovery, income and the doctor’s reputation (as related to the technology used). In regard to the utility functions of hospitals, they are usually considered to be the result of the interaction of hospital-affecting agents: physicians, managers and financing entities in the case of non-profit hospitals (Newhouse (1970)), or instead they are simply considered to be utility maximizers (Dor and Farley (1996)).

24 Throughout the analysis, it is implicitly assumed that the hospital’s utility function depends positively on staff income and quality of services and negatively, on the effort staff must make. Additionally it is assumed that hospital staff can appropriate, at least in part, the utilities that may be generated. Otherwise, the incentives to reduce costs generated by some of these systems disappear.
impediment to the switch, at least in the relatively short run, of a demand subsidy system in which the State beneficiaries, as in other countries, may freely choose a service supplier, increasing the competition between private and public hospitals.

In any case, it is necessary to bear in mind the importance of complementing the supply and demand mechanisms. If the policy objective is to generate efficient consumption patterns while at the same time reducing the financial risk of the beneficiaries (limiting *moral hazard*, adverse selection and demand inducement), the risk division mechanisms on the demand side will only allow second-best solutions to be obtained, since by imposing co-payments or deductibles, the financial risk of the insured is increased. The complementing of those mechanisms with incentives on the supply side, however, will allow the obtaining of better results in terms of efficiency\(^\text{25}\).

### 3.1.1. Cost reimbursement

Cost reimbursement presents several problems: no incentives exist for the efficient production of intermediate services (minimization of costs in the services used) nor for the providing of the socially optimal quantity and intensiveness of services. Since benefits are zero and payments to physicians are not affected by their own behavior, the utility maximization problem of the hospital is solved according to the rest of the variables. Thus, if it is assumed that the other variables are patient welfare or the intensiveness of treatment, the consequence will be an oversupplying of services; and if variables related to effort are introduced into the maximization (as is possible in many cases of public employment), the efficiency losses associated with the reimbursement of costs may be significant.

As advantages of this mechanism may be mentioned that it does not generate incentives for risk selection by the provider, it recognizes that some cost differences among hospitals may not be due to differences of efficiency or intensiveness, it does not

\(^{25}\) Ellis and McGuire (1993).
create incentives for the undersupplying of services and the physician is able to act with greater leeway.

3.1.2. Prospective payments

Even though prospective payment systems share the feature of generating incentives for cost minimization in the production of the services used, in other aspects there are relevant differences among them. The most common variants are: fee for services, payments associated with diagnosis, the fixed global budget and the global budget associated with some kind of output variable.

Fee for services

Following this mode, the payer and the service supplier agree a price for each service, and the payer transfers funds according to the quantity of services provided. Although the mechanism generates incentives for unitary service cost minimization, these incentives also operate toward reducing quality, since the supplier will tend to save as much as he can. In the optimum case, the marginal cost of the service supplied should equal the marginal utility for the patient deriving from the service supply (assuming that there are no externalities of any kind). However, in this system the supplier has incentives to provide services above and beyond this optimum.

It is logical that with this system there will be an explosion of expenditure, due to which it is normal to combine it with cost containment mechanisms on the demand side (incomplete insurance) or the existence of co-payment and/or control mechanisms on the supply side (procedural reviews, hospitalization permits, etc.). As an advantage it may be said that the system does not generate incentives for risk selection, since the hospital is compensated for the extra services it renders to the more seriously ill patients.

26 If the hospital is interested in both the staff’s income and the quality of services, the latter would be “lower” in step with a preponderance of the income effect. This is more likely when costs are more sensitive to quality, and staff incomes are lower. The assumption of a dominant income effect also underlies the analyses of payments associated with diagnosis and of a budget with self-financing commitment.

27 Some authors such as Pauly (1996) suggest that this payment mechanism can be the optimal one under certain assumptions that are difficult to comply with in real life.
Payments associated with diagnosis

According to this system prices are agreed for different diagnosis that are expected to equal the average cost for patients with those diagnosis. The main advantage of this method is that by transferring the risk from the treatment costs to the physician, he will have incentives to limit his expenditure and the inputs and intermediate services he uses. This payment mechanism generates incentives to undersupplying of services, since the supplier appropriates not only what is saved in the production of services, but also the savings from all those services he decides not to provide, for which reason he has incentives to undersupply quality in medical attention. The undersupplying will be limited by the features of the medical practice, professional ethics, patients’ grievance ability (malpractice suits, etc.), the need of the physician to maintain his reputation and the competition among suppliers. If the payments associated with diagnosis are well designed (without differing too much from average case costs) and the aforementioned factors play a significant role, the undersupplying of services and incentives to diminish quality should not be a major problem.

Payments associated with diagnosis generate risk selection incentives for suppliers, since they will be benefited by discriminating against those patients whose treatment implies an above-average cost.28

Fixed budget

This method shares the disadvantages that characterize fixed budgets in any activity, namely: as years go by, the amount of funding bears no relationship to the relevant production level or cost variables, the budget tends to be subjected to negotiations where other interests play a part, etc. If one applies this mechanism, there is no possibility that the hospital staff will appropriate any surplus. Controls on the use of funding are scarce and incentives would lead to very few services being provided. As long as the number of patients being attended in the hospital were not a relevant variable, the interest in the quality of the treatment and the welfare of the patients in the hospital would
create an excess of services for a small group of patients, who will receive services that are above optimum in both amount and quality (given the budgetary restrictions). If there is any interest in treating a greater number of patients, the quantity-quality *trade-off* will be solved according to the shape of the utility function. Despite all its disadvantages, due to its simplicity this is one of the methods most used by the national government and the provinces for transferring funding to public hospitals. Its advantage is not generating incentives for risk selection.

*Budget associated with an output commitment*

Following this mode, the supplier negotiates one or more variables with the third-party payer, making a commitment that they will attain a predetermined value over the next period, and the payer transfers the appropriate funding according to this commitment. The incentives generated by the system will depend on two basic factors: the reward and penalty system that the payer can apply, and the definition of the output variable being negotiated.

In regard to the first point, the institutional framework should allow the effective application of rewards and penalties since, if not, the incentive system designed would lack real power. In regard to the second point, it is fundamental that the variable being negotiated be perfectly observable and measure, as thoroughly as possible, the supplier’s output level. This, together with the reward and penalty program, will determine the incentive scheme imposed on the supplier, and through the supplier’s utility function and its maximization, hospital behavior as well. From the theoretical viewpoint, hospital output is the change in patient welfare, but this is not observable. On the other hand, the initial diagnosis and final status of patients upon entering and leaving the hospital can be observed, due to which output should be measured in relation to these variables.

---

28 This, in fact, happens in the case of “dumping”, which, in this context, occurs when persons with costs over the average transfer themselves to State-financed hospitals.

29 A basic principle of every regulation is that the variable for defining the contract or regulation be perfectly verifiable.
Again, so as not to generate incentives for oversupplying of services and to avoid procedural controls, etc. a possible solution is to set payments according to diagnosis and to the expected costs of treatment. In order to solve, if only in part, the problem of undersupplying of services, the set of services corresponding to different diagnosis could be standardized, albeit leaving the physician a certain freedom of choice.

Table 2 summarizes the discussion about payment mechanisms and incentives.

<table>
<thead>
<tr>
<th>Incentives of alternative methods of payment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production of services</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Cost reimbursement</td>
</tr>
<tr>
<td>Fee per service</td>
</tr>
<tr>
<td>Diagnosis associated prices</td>
</tr>
<tr>
<td>Fixed budget**</td>
</tr>
<tr>
<td>Budget with self financing commitment</td>
</tr>
<tr>
<td>D.R.G. Services</td>
</tr>
</tbody>
</table>

* Competition limits these incentives. ** The hospital is assumed not to appropriate the surplus. *** Interest in quantity is assumed not to exist.

Source: IERAL of Fundación Mediterránea.

3.1.3. Payment form and staff incentives

An additional issue which should be considered is the designing of the payment structure for administrative and medical staff of the hospitals intervening in the production process. From the theoretical viewpoint, the different forms of payment to staff are analogous with (and generate similar incentives to) some of the already analyzed hospital funding transfer modes.

A fixed wage unrelated to worker productivity over time does not generate adequate incentives; this is comparable to the fixed-budget case already analyzed.

If the payment depends in some way on productivity, the incentives generated will clearly depend on the definition of the product. If “product” is taken to mean the quantity of services rendered, incentives will be generated for the induction of demand and the reduction of costs per service (perhaps affecting quality). This physician payment form
may occur in the cases where they are not hospital employees, and rent facilities for patient treatment. In the United States this factor plus the third-party fee for service payment modality (both with hospitals and physicians) explain to some extent the explosion of health service expenditure between the sixties and the early eighties (doubtless, another fundamental factor was the technological progress that took place over the last few decades).

Another way of conditioning staff income to productivity is by linking it to the hospital’s global income. This should be rounded off with reward and penalty schemes for individuals or hospital sectors or services, so as to more closely link individual productivity to income. In any case, management techniques seeking total staff commitment in the attainment of global objectives would be highly important.

3.2. Theoretical Model

Considerations about incentives and different payment mechanisms lead to the presentation of a model attempting to solve this crucial point of hospital inefficiency—not only in regard to the choice of treatment intensiveness, but also in regard to some measurement of effort—, and the need to combine adequate hospital and staff payment mechanisms so as to reduce this problem. Concentrating on these aspects would imply simplifying the model in others, which would impose certain limitations on the conclusions.

In order to analyze the incentive problem generated by the different payment mechanisms, a simple analysis is suggested for a general-type model (bearing several of the features found in the literature) and, although it does not lead to a final result, it nevertheless allows a rapid visualization of some of the effects on incentives generated by different payment mechanisms.

The model is a classical one of the principal-agent type, in which the government (as principal) must design the funding transfer system to the public hospital (the agent) and

---

30 Actually, instead of the hospital paying the doctor, he charges private individuals or their insurers.
to its staff. In the model, the State can not determine the amount of services for the hospital, simply having to set the mechanism(s) of payment and through these induce the desired behavior from the hospital, which (once it has determined the value of the variables it controls) must satisfy all the demand that arises.

The proposed utility function for the hospital is given as the result of the interaction between physicians and hospital managers or directors: \( U = U(I,K,E,) \), where \( I \) is staff income, \( K \) is a measurement of the intensiveness of the treatment provided (for example: devices used, average number of days of hospital stay, etc.) and \( E \) is a measurement of staff effort, such that \( U_3 \) is negative\(^{31} \ 32\). For the next analysis it makes sense to include staff income, if this is allowed to vary according to some measurement of the results. In the first approximation, however, \( I \) will be assumed constant, this being consistent with the historical focus for Argentina, according to which public hospital staff has been paid a fixed wage independent of any result. The intensiveness variable is included because it also affects physicians’ and directors’ utilities (whether due to the prestige arising from the implementation of advanced technology, a legitimate concern for patient welfare, or both). Finally, the measurement of effort is included because it affects hospital staff’s utility; and the incentive schemes proposed will also seek to affect this variable in some way.

The government’s objective in this model consists of the minimizing of the global amount of funding transferred to the hospital, restricting this minimization to keep the

\[^{31}\text{U}_i \text{ or } X_i \text{ will be the first derivative in respect of the } i\text{th argument, } U_{ii} \text{ or } X_{ii} \text{ will be the second derivative in respect of the same argument, } C' \text{ and } C'' \text{ will be the first and second derivatives in respect of the single argument.}\]

\[^{32}\text{Rogerson (1994) proposes a principal-agent model in which the State must determine the optimal non-profit hospital payment mechanism. In the model, both the hospital and the State have the same utility function (depending on the quantity and quality of the services), but the hospital maximizes without worrying about costs. The principal result indicates that the State, in order to achieve an efficient outcome, must set the markups on the marginal cost of each product in inverse proportion to the product’s demand treatment intensiveness elasticity. In the efficient solution, some products (those of low elasticity) will have a positive markup, and others (with high elasticity) will have a negative one, making the global utility zero.}\]
resulting output at the same level or higher than the minimum level considered socially desirable by the government (in general, the restriction will be satisfied with equality)\textsuperscript{33}.

The accurate solution of the problem is that of a sequential game, in which the government moves first, proposing a staff payment function and the mechanism for transferring funds to the hospital, and then the hospital responds by setting the value of the variables under its control ($K$ and $E$) that maximize its utility function, taking the government’s parameters as exogenous. The game is solved recursively, in such a way that the principal will consider the response or reaction function of the agent at the moment of designing a system that will maximize its target function\textsuperscript{34}. The fact of not working with explicit functional forms for utility and cost functions, etc. prevents one from finding the final solution of the model, which would indicate the optimal combination, not only of fund transfer mechanisms to the hospital (price per unit of output, global budget independent of any output and cost reimbursement variable), but also of staff payment forms (fixed component plus a variable amount).

Due to the impossibility of accurately solving the model a different solution strategy is adopted, consisting of the consideration of three pure hospital payment mechanisms (fixed budget, price per unit output and cost reimbursement) combined with different physician payment systems (fixed income, fixed income plus a variable component as a function of the surplus on variable costs, and fixed income plus a variable component as a function of the output level) and of the solving of the hospital maximization problem subject to the restriction imposed by the payment mechanism. Once a solution is obtained to this problem, and in order to analyze the efficiency of each system, minimum expenditure levels required for output to be at least equal to the floor levels proposed by the government are compared.

\textsuperscript{33} The fact that the government minimizes the expenditure subject to a merely quantitative restriction (without considering intensiveness) implies that the intensiveness variable ($K$) in the margin will not be adding anything to the “social valuing” of the output (at least in the government’s eyes), due to which “$K$” might be considered as a variable related to the number of days of hospital stay or an excessive use of existing apparatus or to some other variable that does not add value to the output.

\textsuperscript{34} This would be the incentive-compatibility restriction. The participation restriction is assumed to be satisfied with the fixed staff salary level, which will not be modified throughout the analysis.
The main equations for the model are as follows:

\[ U = U(I, K, E), U_1 > 0, U_2 < 0, U_{ii} = 0, U_{ij} = 0, U_3 \rightarrow 0 \text{ si } E \rightarrow 0 \text{ y } U_2 \rightarrow \infty \text{ si } K \rightarrow 0. \]

\[ CT = F + C(K) \cdot X, C' > 0, C'' = 0 \]

\[ X = X(K, E), X_i > 0, X_{ii} > 0 \]

in which \( U \) is the hospital’s utility function with the usual assumptions on first and second derivatives, \( CT \) are the total costs, \( F \) is the fixed cost (including the payment of the fixed component of staff wages), \( C(K) \) is the average variable cost which is assumed to be a convex function of the intensiveness level and does not depend on staff effort, and \( X(\cdot) \) is the demand function faced by the hospital, which depends positively on \( K \) and \( E \). It is assumed that consumers do not face monetary costs in exchange for services, due to which demand depends positively on the intensiveness or quality of the treatment and positively on the effort made by staff (a greater effort by the staff is reflected in lower non-monetary costs for the patient, such as the reduction of queues or waiting time, for example, which gives the positive mathematical sign to \( X_2 \)). \( E \) is assumed to be limited, so that \( E \geq 0 \).

The variable or variables on which both the payment to the hospital and the incentive scheme are based must be “socially desirable”, due to which the immediate possibility of fixing them in accordance with the \( X \) output variable emerges. Although this alternative is commonly found in many agency problems, the fact that the services concerned are health services gives the model certain special features (some of them not considered in the model) that must be explained. These features refer to the problem of output definition and the possibility of generating different incentives accordingly. If output is taken in terms of hospital discharges (associated with diagnosis), on one hand we have problems related to the great diversity of the cases that can be treated and the different degree of seriousness of these which cannot be contemplated in the diagnosis and, on the other hand, an incentive is created to increase the patient turnover rate and the rate of re-admissions to hospitals (patients are discharged and return for treatment). However, if output is defined in terms of services, the problem of demand inducement will
appear as relevant (enhanced by the fact that beneficiaries face a nil monetary cost). Both are examples of moral hazard on the hospital’s side, and should be borne in mind.

The model avoids the problems of multiplicity of diagnosis (services) and of readmissions and works with one, homogeneous output, which is defined as “discharge associated with a single diagnosis”\textsuperscript{36}. This is implicitly understood as it is assumed that the State is a good agent for its beneficiaries and is concerned about their health, and not about the services rendered in themselves. A similar rationalization strategy would consist of considering one homogeneous service only (for example a medical examination), with no possibility of demand inducement, with this service producing an identical improvement in the health of all patients. Thus, the government’s aim is to minimize expenditure subject to an output level of at least $X^*$. 

Solutions are now proposed for the different cases: first, for the case in which staff income is fixed, then for the case where it depends on the surplus on variable costs and, finally, for the case where it depends on the output level (mechanisms of fixed budget, cost reimbursement and price per unit of output are analyzed for each alternative)\textsuperscript{37}.

3.2.1. Fixed payment to staff and zero fixed cost ($I=S$, $F=0$)

To illustrate the kind of solution sought, a simplified first alternative is proposed, in which it is assumed that the fixed costs ($F$) are nil and that the staff payment is fixed ($I=S$). In this first solution, the fixed budget and unit price mechanisms are proved to be equally efficient and better than that of cost reimbursement; later we will analyze the case of fixed costs, an assumption that will be maintained for the remainder of the cases.

\textsuperscript{36} Fisher, Wennberg, Sukel and Sharp (1994).

\textsuperscript{37} No analysis has been made of the extent to which results are projectable to multi-product contexts with more complex cost functions, nor of what strategic behavior could emerge in the case of multiple payers.

The form of notation to be used will be: * for cases of fixed budget, ° for cases of price per unit and ° for cases of cost reimbursement; for the case of fixed payment to staff ($I=S$) one symbol only will be used, for that of incentives on the economic surplus ($I=S+f(.)$) two, and for that of payment according to output ($I=S+g(x)$) three.
Fixed budget

In this case, the hospital’s problem consists of maximizing $U(I,K,E)$ - with regard to $K$ and $E$ - subject to the restriction that total cost $(C(K).X(K,E))$ be less than or equal to the global budget that the State transfers to it $(Y)$. Thus, the problem and the conditions of the first order are expressed:

Max.: $U(S, K, E) - \lambda \cdot (C(K) \cdot X(K,E) - Y) + \mu \cdot E$

FOC:

$U_2 \cdot \lambda \cdot (C' \cdot X + C \cdot X_t) = 0$  \hspace{1cm} (1)

$U_3 \cdot C \cdot X_2 + \mu = 0$  \hspace{1cm} (2)

$\lambda \geq 0$  \hspace{1cm} (3)

$\mu \geq 0$  \hspace{1cm} (4)

$\lambda \cdot (C(K) \cdot X(K,E) - Y) = 0$  \hspace{1cm} (5)

$\mu \cdot E = 0$  \hspace{1cm} (6)

where $\lambda$ and $\mu$ are Lagrange multipliers and $S$ is a constant. Since $U_2$ and the second term parenthesized factor of equation (1) are positive, it may be deduced that $\lambda$ will be strictly positive, which implies through (5) that the budgetary restriction will be active ($C(K).X(K,E)=Y$) and through (2) that $\mu$ will be strictly greater than zero; this in turn means that, through (6), $E=0$.

In the hospital’s optimum, effort is nil, since it does not represent any additional income ($I=S$ is fixed), it causes a loss of utility ($U_3<0$), and also, a positive effort level would imply a greater output level ($K$ given) which, in order to satisfy financial restriction would imply a lower intensiveness level for the solution. In this way, with $E=0$, the hospital chooses the highest level of $K$ which will satisfy the budgetary restriction.

The solution to this problem therefore indicates that the effort will be nil ($E=0$), and the intensiveness level ($K$) as high as possible. In other words, $K=K^*/C(K^*).X(K^*,0)=Y$. Obviously, $Y$ is equal to the level of expenditure ($G$), and this bears a positive relationship to the chosen level of intensiveness. In this way, it is possible
to choose \( Y = G^*/X(K^*(G^*),0) = X^* \), where \( X^* \) is the output level set as a lower limit by the government.

### Unit output price

The problem in this case is very similar to the preceding one, but instead of receiving a budget, the hospital receives a fixed amount \( p \) per unit of output:

\[
\text{Max.: } U(S, K, E) - \lambda \cdot ((C(K) - p) \cdot X(K, E)) + m \cdot E
\]

**FOC:**

\[
\begin{align*}
U_2 & - \lambda \cdot (C' \cdot X + (C \cdot p) \cdot X_t) = 0 \\
U_3 & - \lambda \cdot (C - p) \cdot X_t + \mu = 0 \\
\lambda & \geq 0 \\
\mu & \geq 0 \\
\lambda \cdot ((C(K) - p) \cdot X(K, E)) & = 0 \\
\mu \cdot E & = 0
\end{align*}
\]

Again, equation (1) implies that \( \lambda \) will be different from zero (and positive by restriction(3)) and through (5) that the budgetary restriction is satisfied with equality \( (C = p) \). Equations (2) and (6) imply that effort is nil in this case also, since if it were positive, (6) would imply \( \mu = 0 \) and (2) would no longer hold (since \( C = p \)). The solution is analogous to that of the previous case, with the hospital now choosing \( K = K'/C(K') = p \).

This defines a function \( K(p) \) such that the government may choose \( p = p'/X(K(p'),0) = X^* \).

Logically, the level of intensiveness necessary for output to be \( X^* \) is the same in both cases \( (K^* = K') \) since the effort is also identical in both. In this case, the total expenditure is \( G' = p'.X(K'(p),0) = p'.X^* = C(K'), X^* = C(K^*).X^* = G^* \), due to which both mechanisms would be equally efficient.

### Cost reimbursement

The solution in the case of cost reimbursements is trivial, the hospital maximizes its utility without any type of restrictions; the hospital will then choose the maximum value for \( K \) and the minimum value for \( E \) (an upper limit \( K_{\text{max}} \) is assumed for \( K \)). Consequently the government cannot induce the desired level of output, and the resulting
expenditure level will be $G^o = C(K_{max}) \cdot X(K_{max}, 0)$. Obviously, $G^o > G^* = G'$ since it is assumed that $K_{max}$ is greater than $K^* = K'$.

Graph 1 illustrates these first results: when budgetary restrictions are taken to the plane $K, E$ they slope negatively (fixed budget) or to infinity (price per output) while the indifference curves are positively sloped, due to which the maximizations in both cases are shown as corner solutions. In the cases of fixed budget and price per unit it will be possible to induce the choice of an intensiveness level that, given the nil effort, will lead to an output level equal to the minimum proposed as a restriction by the government, but in the case of cost reimbursement the solution will always be at nil effort and the maximum level of intensiveness, which would induce an $X(X_R)$ too high.

3.2.2. Fixed payment to staff and fixed costs greater than zero ($F > 0$, $I = S$)

Fixed budget and cost reimbursement

The fact that $F$ is greater than zero does not alter the main conclusions of the cases of fixed budget and cost reimbursement. In both, the resulting level of expenditure would be identical to the previous one plus fixed cost, since the only change in the problem is the addition of a constant ($F$) to the budgetary or financing restriction. However, with the introduction of a fixed cost, the analysis for the case of unit output price does vary.
The problem is conceptually similar, but the need to finance fixed cost through the payment of a price per unit changes the nature of the budgetary restriction and induces the hospital to choose a positive $E$. In the previous cases there was no sense in doing so, since in no case did the level of the possible (in view of the restriction) $K$ increase thereby.

Graph 2 clearly illustrates the different situations: in the case in which there was no fixed cost, the budgetary restriction in the plane $E, K$ was a straight line parallel to the y-axis; due to which maximization was a corner solution with $E=0$ and $K$ as high as the restriction would allow. In the case of fixed costs, the budgetary restriction transferred to plane $E, K$ is a convex function, due to which, since the indifference curves are concave and positively sloped, utility maximization will be obtained at an inner point (B) with a positive effort level. Intuitively, the fact of achieving effort induction makes one think that there may have been a gain of efficiency (drop in total expenditure) through the use of payment per unit of product.

Formally, the problem may be stated:
Max.: \( U(S, K, E) - \lambda ((C(K) - p) \cdot X(K, E) + F) + \mu \cdot E \)

**FOC:**

\[
U_2 \cdot (C' \cdot X + (C - p) \cdot X_2) = 0 \tag{1}
\]

\[
U_3 \cdot (C - p) \cdot X_3 + \mu = 0 \tag{2}
\]

\( \lambda \geq 0 \) \tag{3}

\( \mu \geq 0 \) \tag{4}

\[
\lambda \cdot ((C(K) - p) \cdot X(K, E) + F) = 0 \tag{5}
\]

\( \mu \cdot E = 0 \) \tag{6}

Equations (1) and (3) again indicate that \( \lambda > 0 \), due to which the financing restriction is satisfied with equality. Since \( C < p \) is a necessary condition to satisfy the budgetary restriction, equation (2) indicates that \( U_3 \) should be strictly negative, due to which \( E \) will be greater than zero (it was assumed that \( U_3 \) tends to zero when \( E \) tends to zero) and \( \mu \) will equal zero (to comply with (6)).

To see the expenditure level associated with this payment mechanism, it is worthwhile analyzing graphs 3.a and 3.b. In graph 3.a, further increases of \( p \) (\( p^0 > p' \)) gradually delineate the \( E(K) \) function showing the pairs of \( E \) and \( K \) that may be induced through different prices. In graph 3.b one may observe the demand levels while \( K \) varies, keeping \( E = 0 \) (full line) and varying \( E \) according to the function \( E(K) \) (broken line). Since \( E(K) > 0 \), the same levels of \( X \) will be associated with lower levels of \( K \) when \( E \) varies according to function \( E(K) \) (asymptotically, when \( E(K) \) tends to zero, the \( X[K, E(K)] \) function will tend to \( X(K, 0) \)). Thus, the level of \( K \) necessary to reach \( X^* \) will be lower with this payment mechanism, as will be the total necessary expenditure level:

\[
G' = p' \cdot X(K', E(K')) = p' \cdot X^* = C(K') \cdot X^* + F < Y = C(K^*) \cdot X^* + F = G^* \tag{38, 39}
\]

---

38: \( d^2 E/dK^2 > 0 \), since \( p \) is necessarily greater than \( C \).

39: The equality of the third and fourth member is verified, since as \( \lambda \) is greater than zero, equation (5) implies that the budgetary restriction is satisfied with equality \( ((C - p) \cdot X + F) = 0 \).
In summary, if fixed costs are present, the system of price per unit payments will be more effective (it will be associated with a lower expenditure level for the same level of output) than that of cost reimbursements and global budgeting. This is simply due to the fact that, because of to the nature of the budgetary restriction, in the case of unit price payment a greater effort will allow the increase of intensiveness while still satisfying the restriction, due to which the hospital, maximizing its utility, will be willing to make a positive effort in exchange for an increase in the level of $K$. The principal will exploit this behavior on the part of the agent by setting a final price inducing a lesser level of intensiveness and a greater effort ($K', E'$) compared with the fixed budget case ones.

### 3.2.3. Payment to staff as a function of the surplus on costs

$$I = S + f(TI - VC)$$

The incorporation of any staff income incentive system will aim to induce a determined pattern of behavior. Therefore, it is logical to define the incentive in terms of some kind of efficiency-related result, such as the generation of a surplus on costs (as an incentive to save funds) or higher production (as a means of incentivating effort):
Incentives are now fixed according to the surplus on variable costs, this being essentially similar to a system that sets incentives for total utilities, since it will be assumed that \( f(F) = 0 \), which would be the same as assuming that \( f(0) = 0 \) if the function’s argument were the level of utilities. The assumptions on the function are \( f'(.) > 0, f''(.) \leq 0, \) and \( f'(F) \leq 1 \).

Fixed budget

The formal statement of the problem is:

**Max.:** \( U(S+f(Y-C(K)\cdot X(K,E)), K, E) - l \times (C(K) \times X(K,E)-Y+F) + m \times E \)

**FOC:**

\[-U_1 f'(C'X+C\cdot X_1)+U_2 - \lambda \cdot (C'X+C\cdot X_1) = 0 \]  
\[-U_1 f'(C\cdot X_2)+U_2 - \lambda \cdot C\cdot X_2 + \mu = 0 \]  
\[\lambda \geq 0 \]  
\[\mu \geq 0 \]  
\[\lambda \cdot (C(K)\cdot X(K,E)-Y+F) = 0 \]  
\[\mu \cdot E = 0 \]

Thanks to the system that has been proposed according to the conditions of the first order, it is no longer necessary for \( \lambda \) to be strictly greater than zero, so the two possible cases will be analyzed: with \( \lambda \) greater than zero and the financing restriction active, and with \( \lambda \) equal to zero and the restriction inactive.

When \( \lambda \) equals zero, equation (1) may be stated as \( U_1 \cdot \delta I/\delta K + U_2 = 0 \), which indicates that in the optimum case, the marginal utility of the increase in \( K \) (\( U_2 \)) will be equal to the marginal disutility derived from the loss of income due to the increase of \( K \) (\( \delta I/\delta K < 0 \)).

To see the level of expenditure associated with this incentive mechanism one must analyze what happens with the optimum effort level. Since the budget is fixed, an increase of \( E \) would trigger an increase of \( X \), and therefore one in total variable costs also, which in turn would imply a lower remuneration level. In this way, increasing the level of effort
would imply a loss of utility for the hospital, due both to the greater effort \((U_3\) negative) as well as the drop in the value of \(f(.)\) that this would imply. In terms of the equations, this is seen in (2) where \(\delta I/\delta E=-f’(C.X_2)<0\), due to which the sum of the three first terms will be negative, this requiring that \(\mu\) be strictly positive to satisfy the equation. Through (6) this means that the optimum effort level will be zero.

Since \(E=0\), the level of \(K\) necessary for demand to equal \(X^*\) will be \(K^{**}=K^*\). If this level of intensiveness can be induced, the associated expenditure level will be:

\[
G^{**}=C(K^{**})X^*+f(Y(K^{**})-C(K^{**})X^*)+F=C(K^*)X^*+F=G^*^{40}
\]

In general, \(G^{**}>G^*\) since, starting from the assumption that \(\lambda=0\), only a mere coincidence would produce \(Y-C(K^*)X^*=F\) and therefore \(f(.)=0\) and \(G^{**}=G^*\).

In the case of the restriction being active in the solution \((\lambda>0,Y-C(K^*)X^*=F)\) the expression \(U_1\delta I/\delta K+U_2\) will be positive and, therefore, the increment in utility derived from increasing the intensity will be greater than the utility loss caused by this increase on the income side. This means that the hospital would wish to increase \(K\), but this would mean infringing the budgetary restriction. If this is the case, the relevant component of payments to staff will be zero \((f(F)=0)\) and, since the total of the first, second and third terms of (2) is negative, then \(\mu>0\) and the incentive scheme will not induce an \(E>0\). The solution to the problem will be identical to that of the case without a variable component with a required expenditure level equal to \(G^*\).

Since in the case of a non-active restriction the budget required by this incentive scheme is equal to or above that required with a fixed wage, and with an active restriction the budget required is the same as that without incentives, it must be concluded that there is no sense in using a scheme of this kind when hospital payments are made by budget. The effectiveness of the mechanism is now analyzed by setting a prospective price.

\[
^{40}\text{In } G^{**} \text{ it is being implicitly assumed that the government recovers all surpluses or economic utilities that are not destined to the payment of the variable component of staff income } f(.)\text{.}
\]
Unit output price

The problem and the conditions of the first order of maximization in this case are:

\[
\text{Max.: } U(S+f((p-C(K))\cdot X(K,E)), K,E) = \lambda \cdot ((C(K)-p)\cdot X(K,E)+F)+\mu \cdot E
\]

**FOC:**

\[
\begin{align*}
-U_1 f'((C'\cdot X+(C-p)\cdot X_1)+U_2 \lambda (C'\cdot X+(C-p)\cdot X_1) &= 0 \quad (1) \\
-U_1 f'((C\cdot X,2)+U_3 \lambda (C-p)\cdot X_2+\mu &= 0 \quad (2) \\
\lambda &\geq 0 \quad (3) \\
\mu &\geq 0 \quad (4) \\
\lambda ((C(K)-p)\cdot X(K,E))+F &= 0 \quad (5) \\
\mu \cdot E &= 0 \quad (6)
\end{align*}
\]

Initially it should be stressed that when the payment mechanism is that of prospective prices, in contrast with the budget case, this incentive scheme implies that the derivative of staff income in respect of effort will be positive \((\delta I/\delta E = f'.((p-C).X_2)>0)\), which leads one to think of the possible efficiency profits derived from the incentive scheme. By finding \(\lambda\) in equation (1) and replacing it in (2) it is proved that the level of effort will be necessarily positive, and therefore \(\mu\) will be zero\(^{41}\).

In the case that both \(\mu\) and \(\lambda\) are zero, as the restrictions are not active, equations (1) and (2) will be: \(U_1 \delta I/\delta K+U_2=0\) and \(U_1 \delta I/\delta E+U_3=0\). In both, the marginal utility of increasing variables \(K\) or \(E\) by one unit \((U_2 \delta I/\delta E)\) will be equal to the marginal loss produced by such an increment \((U_1 \delta I/\delta K \neq U_3)\). To the extent that the financing restriction is active \((\lambda>0)\) it will be verified, on one hand, that the marginal utility of increasing the intensity level is higher than the utility loss on the income side and, on the other hand, that the marginal utility loss derived from increasing the effort \((U_3)\) will be greater in absolute value than the increase on the income side. The explanation of this is that the desire to increase the intensity level \((U_1 \delta I/\delta K+U_2>0)\) induces an additional effort that contributes

---

\(^{41}\) Equation (2) remains as: \((C-p)X_2\cdot(-U_2/(C'X+(C-p)X_1))+U_3+\mu=0\) (2'), where the first term denominator is necessarily positive by equation (1), making the first term positive and therefore making it necessary for \(U_3\) to be strictly less than zero. This makes it necessary for the level of effort to be positive.
to relaxing the financial restriction (when \(E\) is increased, \(X(K,E)\) increases and therefore \(p.X(K,E)\)), allowing a marginal increase of \(K\).

What about the expenditure level necessary to reach \(X^*\)? Compared with an identical incentive scheme but with a fixed budget, it would be logical to think that efficiency would increase, since the mechanism failed to induce staff effort (due to which expenditure was related to an excessively high intensiveness level). In this case effort is always positive, due to which a lower intensiveness level is expected in order to achieve \(X(K,E)=X^*\), which in turn could reflect in lower costs to reach the desired level of output.

If it is assumed that those surpluses on total costs plus the variable component of payment to staff \((p.X-F-C.X-f(.))\) are retained by the government itself, the total expenditure necessary will be:

\[
G''=C(K'')\cdot X^*+F+f((C(K'')-p'')\cdot X^*)
\]

The comparison between \(G''\) and \(G**\) can not be error free, since the functions relating \(Y\) and \(K\) (in the previous case) and \(p\) and \(K\) (in this one) are unknown, preventing us from discovering the value of function \(f(.)\) in each case. It is possible, however, to carry out a comparison between the expenditure level \((G'')\) and that associated with a prospective price under a fixed wage scheme \((G')\). Intuitively, one might think that, in view of the adding of an extra factor for effort induction to the case without incentives and a prospective price (the possible increase of income derived from greater effort), this case should be at least equal to the previous one.

However, in order to make some conclusion on the expenditure necessary to induce an \(X^*\) output level, the amounts of inducible effort for each intensiveness level \((K)\) must be analyzed: this is an \(\hat{E}(K)\) type function, as in Graph 3.a. By calculating the quotient of the first two conditions of the first order in the problem, equations (1) and (2), and considering \(\mu=0\), one obtains:

\[
U_2/U_3=(C'\cdot X+(C-p)\cdot X_1)/(C-p)\cdot X_2,
\]

\[42\] In the case of \(\lambda>0\), function \(f(.)=0\), and therefore, \(G''<G**\).
This expression may be verified irrespective of the fact that \( \lambda = 0 \), and is identical to that obtained for the case of a prospective price without incentives. Assuming that \( U_2 \) and \( U_3 \) do not depend on staff income level (this is one of the initial assumptions: \( U_i = 0 \) for every \( i \) different from \( j \)), it is concluded that the \( E, K \) pair set inducible through some kind of price is identical in both problems, and is given by the \( \hat{E}(K) \) function of Graph 3.a.

Thus, as the effort and intensiveness levels to solve this problem \( (E'', K'') \) are equal to those of the case without incentives \( (E', K') \), the expenditure level necessary to induce \( X^* \) will be equal to or greater than that of the case without incentives (it would be equal in the case that \( \lambda > 0 \) - and therefore \( f(.) = 0 \) - and greater when \( \lambda = 0 \) and \( f(.) > 0 \). In short, this incentive scheme will not be effective either with a prospective price hospital payment system.

According to the results obtained up to the moment, the proposed incentive scheme is not efficient under the analyzed payment mechanisms \( (G^{**} \geq G^* \text{ and } G'' \geq G') \). Up to now, the most efficient mechanism has turned out to be that of prospective prices without a staff income incentive scheme. In the next section a second incentive scheme will be analyzed where staff income depends on output level\(^{43}\).

3.2.4. Payment to staff as a function of the output level \((I=S+g(X))\)

The fact that a staff incentive scheme based on utilities (or on a surplus on the variable costs) is not convenient leads us to think of another incentive structure based on the output level. Although it would be interesting to compare the results to be obtained with those of the case without incentives, we will once again be limited by the fact of not working with explicit functions. The analysis presented will only consider a function \( g(.) \) monotonously growing, with a second derivative lower than or equal to zero\(^{44}\).

---

\(^{43}\) Analyzing this incentive scheme under a cost reimbursement mechanism makes no sense, since the difference between income and variable costs would always be \( F \) (therefore, \( f(F)=0 \)), giving an identical result to the case without incentives.

\(^{44}\) One might think of a function \( g(X) \), such that \( g'(X^*) = 0 \) and \( g''(X)<0 \), or simply of a payment scheme such that the variable component were positive (and large enough) only if \( X = X^* \). Although either of these schemes would probably give a more efficient result in the model than the one with \( g'>0 \) and \( g''<0 \), it is hard to believe that these could be implemented. There are, in fact, countless factors adding to the uncertainty of
Fixed budget

The problem and conditions of the first order in this case are expressed as follows:

Max.: $U(S+g(X), K, E) - \lambda(C(K)X(K,E)+F+g(X)-Y)+\mu E$

FOC:

$U_1g'X_1+U_2\lambda(C^+X+(C+g')X_1) = 0$ (1)

$U_1g'X_2+U_3\lambda(C+g')X_2+\mu = 0$ (2)

$\lambda \geq 0$ (3)

$\mu \geq 0$ (4)

$\lambda(C(K)X(K,E)+F+g(X)-Y) = 0$ (5)

$\mu E = 0$ (6)

Since the two first terms of equation (1) are positive, $\lambda$ must be greater than zero and the financing restriction will be active in the optimum case. With regard to the value of $E$, it cannot be stated that it is necessarily positive; an increase of $E$ has three basic effects: the first (negative) through the reduction of utility ($U_3$ negative), a second, positive, through the increase in the output and consequently in income ($g' > 0$); and a third negative effect also derived from the increase in the output which, through the budgetary restriction and with fixed total income imposes a lower intensiveness level. Since $U_3$ tends to zero when $E$ tends to zero, the first effect will not be important for low effort levels (to determine whether the value of $\mu$ is zero). Logically, as incentives become greater ($g'$) and the fixed component of staff income turn to be lower (greater $U_1$), the effort will more likely be positive.

Graph 4 illustrates the situation: the budgetary restriction has a negative slope and is identical to that of Graph 1, but the indifference curves map in the plane $E, K$ is no longer the same. Since: 1) increases of $E$ produce increases of staff income (through increased quantity), 2) $U_3$ tends to zero when $E$ is close to zero, and 3) $X_2$ is “large” when $E$ is low, it should be expected that these “indifference curves” in plane $E, K$ might have

the demand faced by the hospital, and it would complicate the possibility of setting a certain level of $X$ as a target; to this must be added the restrictions imposed by the hiring schemes enforced (especially in the public sector) that probably would make schemes of this type non-viable.
an initial stage with a negative slope, followed by one with a positive slope. Beyond a
certain point, the effects mentioned above lose relative importance and others,
determining the positive slope, begin to prevail (in the negative stage of the slope, $U_3$ is
greater in (absolute) value, $X_2, U_1$ and $g'$ are less important, and $X_1$ and $U_2$ increase).

The two possibilities mentioned above are shown in Graph 4: a) that the slope of
the budgetary restriction when $E=0$ is greater (in absolute value) than that of the
indifference curve, in which case $E$ will be positive; or b) that it is lesser than that of the
indifference curve, in which case the solution will be a corner one and identical (as regards
the $E, K$ solution pair) to that of the case without incentives ($E=0$ and $K=K^*$;
$X(K^*,0)=X^*$). Logically, if this last were the case, the incentive scheme wouldn’t make
any sense; but if, as expected, the relevant case is that with an inner solution and positive
effort levels, the introduction of the incentive mechanism could generate efficiency gains.

---

These indifference curves incorporate the marginal relationships of "K" and "E" corresponding to the utility
function, the changes in utility deriving from changes in income through $g(X(K,E))$, which are due, in the
final analysis, to variations in the values of "K" and "E".
In any case, in order to be able to compare expenditure levels associated with $X^*$ in each individual case, the different functions involved should be defined. It may, however, be compared with the case of a fixed price and the same incentive scheme.

**Unit output price**

We now analyze the case where the government pays a unit output price to the hospital and sets an incentive scheme on an output level identical to that analyzed. The problem is:

**Max.:** $U(S+g(X), K, E)-\lambda \cdot ((C(K)-p) \cdot X(K, E)+F+g(X))+\mu \cdot E$

**FOC:**  
\[
\begin{align*}
U_1 \cdot g' \cdot X_1 + U_2 \cdot \lambda \cdot (C' \cdot X + (C-p+g') \cdot X_1) &= 0 \quad (1) \\
U_1 \cdot g' \cdot X_2 + U_3 \cdot \lambda \cdot (C-p+g') \cdot X_2 + \mu &= 0 \quad (2) \\
\lambda &\geq 0 \quad (3) \\
\mu &\geq 0 \quad (4) \\
\lambda \cdot ((C(K)-p) \cdot X(K, E)+F+g(X)) &= 0 \quad (5) \\
\mu \cdot E &= 0 \quad (6)
\end{align*}
\]

Since the two first terms of (1) are positive, $\lambda$ will be positive and the financing restriction will be active in the optimum case. Therefore, as $C-p+g'$ is under zero, equation (2) implies that $U_3$ must be strictly negative, and therefore $E$ will be greater than zero and, through (6) $\mu$ will be zero.

To compare the expenditure level associated with this mechanism with that of the fixed budget, one need only observe that the function associating pairs $E, K (\dot{E} (K))$ which the hospital would choose in the case of different prices is above that corresponding to the different budgets (Graph 5). This is so because the budgetary restriction faced by the hospital when it is paid by price has a positive slope (in the relevant section) and therefore tangency will occur in the positive section of the indifference curves. Then, since for each $K$ a greater effort may be induced, the intensiveness needed for the output level to be equal
to the required minimum ($X^*$) will be lower ($K'''<K^{**}$) and the total costs will also be lower:

$$G'''=C(K''')X(K''',\hat{E}(K'''))+g(X(K''',\hat{E}(K''')))=$$
$$C(K''')X^*+g(X^*)<C(K^{**})X^*+g(X^*)=G^{***}$$

Graph 5

On the other hand, a comparison with the expenditure associated with the prospective price scheme without incentives ($G'$) cannot be directly made while the functions are not defined. As $g(X)$ is not specified, the total cost represented by the variable component of payments to staff cannot be established, nor can the total expenditure level be determined. It is possible, however, to compare function $\hat{E}(K)$ with that corresponding to the case without incentives $\hat{E}(K)$ (Graph 6): it may be shown that for identical values of $E$ and $K$ the budgetary restriction for the case with incentives has a

46 From the budgetary restriction one has $p-F/X=C(K)+g(X)/X$, since $g''(X)\leq 0$, then $g(X)/X\geq g'(X)$, due to which $p-F/X\geq C(K)+g'(X)$ and, consequently, $p>C(K)+g'(X)$.

39
lesser slope than for the case without incentives, while the indifference curve including the effect of change in income due to the incentive system will have, for the same pair of \( E, K \), a greater slope than the original curve.

For the case without incentives, the slope of the restriction was \( \frac{dE}{dK} = C' \cdot \frac{X}{(p^o - C) \cdot X_1 / X_2} \), while for the case with incentives it is \( \frac{dE}{dK} = C' \cdot \frac{X}{(p' - C - g') \cdot X_1 / X_2} \). If we evaluate these slopes in the same pair of \( E, K \) (and therefore the same \( X \)) it is shown that the slope in the case without incentives will be lesser to the extent that \( g' \) is lesser than \( p' - p^o = \Delta p \). Using the fact that in both cases the restrictions are satisfied with equality, it is deduced that \( \Delta p = g(X)/X \) and, since \( g' \leq 0 \), \( \Delta p > g'(X) \).

The slope of the indifference curve without incentives is simply \(-U_2/U_3\) valued at that point. The slope of the curve including the income effect will be \(- (U_2 + U_1 \cdot g' \cdot X_1) / (U_3 + U_1 \cdot g' \cdot X_1)\) which is greater (for the same pair of \( E, K \)) than \(-U_2/U_3\) since the denominator is lower in absolute values and the numerator is higher.
In this way, the tangencies between the budgetary restriction and the indifference curves (with the effect of the change in income) that outline the function \( \hat{E}(K) \) will necessarily be over those corresponding to \( \hat{E}(K) \). Since in the solution to both problems the output level will be \( X^* \), the solution to the problem with incentives will be in the section above and to the left of point \( A''' \) of the curve \( \hat{E}(K) \), with a lesser level of intensiveness and a greater level of effort\(^49\).

Although we cannot compare the final level of expenditure associated with both solutions, the fact of being able to induce a greater level of effort for the same levels of intensity leads us to think of the existence of possible efficiency gains.

\(^{49}\) The solution cannot be in the \( A''B \) section of the curve since both "E" and "K" would be greater than \( E' \) and \( K' \), due to which \( X \) would be greater than \( X^* \). Additionally, since the output level grows as we go down \( \hat{E}(K) \) (p goes up), the points lying below B cannot be the solution either.
Cost reimbursement

The solution to the problem under a cost reimbursement scheme is simple: just as in the case without incentives, the highest possible intensiveness level will be chosen, since as well as the increased utility derived from the increase of $K$ itself, there will now also be a positive relationship between $K$ and staff income. In this way, since $K$ will be $K_{\text{max}}$, the unit cost will be identical to that of the case without incentives, but now the total expenditure will be greater since $E$ will be positive to balance the trade-off between effort and the income increase (since an increase of $E$ will increase $X$ and, through $g(.)$, staff income).

In this case, the fact that effort is positive is not reflected in an increase of efficiency. This, contrasting with the other analyses where $E$ was positive, is simply due to the government’s being unable to take advantage of the hospital’s optimum decision to induce it to choose a certain combination of $E, K$ which will lead to a demand level for the output of $X^*$. Thus, the hospital will always choose $K=K_{\text{max}}$.

3.2.5. Results

Through the model presented, it is sought to reflect the importance to efficiency of the different payment mechanisms, through the incentives generated by them\textsuperscript{50}. As mentioned at the beginning, the two prominent features refer to efficiency in regard to the choice of treatment intensiveness and staff effort. The analysis combined two “instruments” to be used by the principal (the government); the transfer mechanism of funds to hospitals and the definition of the hospital staff payment system. Due to the impossibility of solving the problem recursively, a simpler strategy was chosen which,\textsuperscript{50}

\begin{footnote}
50 The model has several simplifications that have been mentioned here and there, but that should be stressed. They refer mainly to: the non-taking into account of: the existence of multiple payers, incentives for demand induction that can generate the introduction of incentive mechanisms on the level of production (this problem is worsened when there is no sharing of risks with supply), and the possible incentives for risk selection on the hospital’s side in the (more realistic) case in which treatment required for different patients (even in the case of identical pathology) is not homogeneous as assumed in the model.
\end{footnote}
albeit not enabling the determination of the “optimum mechanism”\textsuperscript{51}, did nevertheless allow a comparison between the different alternatives.

Basically, three systems of hospital payment were compared, testing each one with three different systems of staff payment. The basic results obtained were the following: The cost reimbursement mechanism is less efficient under any incentive scheme; the unit price payment scheme leads to more efficient results than that of a fixed budget in the cases without incentives (and an over-zero fixed cost) and in those with incentives for quantity; the incentive mechanism defined according to the surplus of total income over variable costs is not efficient in any of the payment systems tested; the introduction of an incentive system on quantity is effective in inducing effort. Although it was not possible to compare expenditure levels with those of the cases without incentives, it is possible that the greater amounts of effort induced may be reflected in lower expenditure levels. These results are summarized in Table 3.

<table>
<thead>
<tr>
<th>Main Results</th>
<th>$I=S+I(TI-VC)$</th>
<th>$I=S$</th>
<th>$I=S+g(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed Budget</strong></td>
<td>$G''$ ≥ $G^*$ &gt; $G'''$</td>
<td>$G^*$ &gt; $G'''$</td>
<td>$G'''$</td>
</tr>
<tr>
<td><strong>Unit Output Price</strong></td>
<td>$G''$ ≥ $G'$ &gt; $G'''$</td>
<td>$G'$ &gt; $G'''$</td>
<td>$G'''$</td>
</tr>
<tr>
<td><strong>Cost Reimbursement</strong></td>
<td>$-$ &gt; $G^*$ &lt; $G'''$</td>
<td>$-$ &gt; $G'''$</td>
<td>$G'''$</td>
</tr>
</tbody>
</table>

Note: The light inequalities are relationships non-demonstrated, but expectable according to the analysis.

3.3. Other factors to be considered

3.3.1. Multiple payers

In the context of Self-Financed Public Hospitals (in which one of the main concerns - especially in the initial stages - is the recovery of costs through the charging of

\textsuperscript{51} No intermediate combinations of payment mechanisms to hospital and staff were analyzed.
services rendered to persons with some type of coverage and the signing of agreements with health care insurers) problems associated with the existence of multiple payers take on added significance. Thus, two of the aspects meriting the attention of specialists and concentrating theoretical and empirical studies have been cost displacement among payers and patient discrimination according to the payer.

The possibility that changes in the prices fixed for a payer may alter the prices, quality or quantity of the services rendered to other payers, has been analyzed at both theoretical and empirical levels. This type of behavior, known as cost-shifting, refers to the possibility that hospitals have of charging patient treatment costs to another payer\textsuperscript{52}. Predictably, cost transfer operates from those “payers” of reduced payment capacity towards those of greater purchasing power. Empirical evidence on cost transfer is weak, as is the theoretical basis for this hypothesis. Additionally, for this to happen, the hospital must necessarily be “price-accepting” in respect of a payer (or finds its negotiating power strapped) and is able to set the price for the other payer (or at least wield certain negotiating power).

Empirical evidence shows that (theoretically at least) the cost-adjusting hypothesis - according to which treatment intensiveness will be related to the payer’s generosity - is more plausible\textsuperscript{53}. This hypothesis would imply that the service supplier reacts by adjusting the quality or intensiveness of the treatment to be provided to the expected payment (so as to bring marginal cost closer to marginal income), without affecting the prices of other payers nor the quality provided to beneficiaries. According to this hypothesis, if the State desires the treatment received by its beneficiaries to be similar to that received in public hospitals by the beneficiaries of a health insurer, it should equate its payment levels with those of the insurer.

Among the different studies carried out on this topic, Glazer and McGuire (1994) propose a model with two payers who must determine the optimum payment system to a

\textsuperscript{52} Most of the papers analyze this kind of behavior for the case of hospitals in the U.S., which are able to charge higher prices for services in the case of patients whose payer is not the State.

\textsuperscript{53} Dor and Farley (1996).
supplier. The result (symmetrical Nash equilibrium) shows the incentives that both payers have to transfer costs to the other (cost-shifting\textsuperscript{54}), and how the choice of supplier inputs is affected by the strategic behavior of payers leading, generally, to inefficient results.

Patient discrimination in the context of multiple payers is a second topic, but closely linked to the above. So that this may happen, in the first place there must be a restriction on the supply side (assuming that payments in all cases exceed marginal costs). If this restriction exists, one may expect a discrimination against those patients whose mark-up was lower (this analysis is valid if both payers use mechanisms such as those of fee for services or payment associated with diagnosis). In the event that one of the payers is linked to the hospital by a system implying that the hospital’s marginal income derived from patient treatment is zero, one might expect a greater tendency to discriminate beneficiaries\textsuperscript{55}.

3.3.2. Demand inducement

When analyzing how to generate and regulate adequate incentives on the supply side, both the payments and regulations affecting hospitals as well as those directly affecting physicians must always be considered. As these are key operators of the system, the contractual relationship between hospitals and physicians is a decisive component of the industrial organization of the health system. Literature has frequently dealt with the problem of demand inducement and the hypothesis of target income. It suggests, basically, that physicians have a target income level, and that in the face of changes in the prices of services they provide, they will respond with changes in the quantity of services provided (through demand inducement) so as to reach the desired income level. This can only happen to the extent that their income depends on their behavior. In this way, in cases

\textsuperscript{54} The hypothesis of cost-shifting in this paper is not the usual one since the hospital does not set prices (it only accepts or rejects contracts). In this model, payer incentives are taken to mean offering a contract that produces losses that are, nevertheless, compensated for by utilities obtained by the hospital from another contract (the hospital accepts the contracts if and when global utilities add up to zero or more than zero).

\textsuperscript{55} This would almost always happen with patients without coverage in contrast to those who have a third-party payer, since the hospital will consider the need of financing contracts in subsequent years.
where the physician receives a fixed income for his work, the problem of demand inducement disappears\textsuperscript{56}.

These models of demand inducement assume that the patient does not have enough information on the advantages and disadvantages of different medical services - or that the monetary cost to them is zero -, and that the professional physician sets a lower value on the loss of reputation deriving from oversupplying services than on the additional income deriving from this type of behavior. Gaynor (1994) analyzes in detail the theoretical and empirical evidence of demand inducement by physicians, concluding that the “games of persuasion”\textsuperscript{57} framework seems to be the one most suitable for analyzing the interaction between physicians and patients, although no incontrovertible empirical evidence exists on inducement by physicians.

From the viewpoint of demand inducement, as there are multiple payers, changes in price by one of them will generate cross-effects altering induction levels in services provided to the other payers. McGuire and Pauly (1991) propose a model with multiple payers (or multiple services) in which the physician’s utility depends on his income, his leisure time, and negatively, on demand inducement. Empirical evidence supports the existence of an income effect, but not the pure target income hypothesis.

Another interesting aspect is the consideration of multiple products. In a context of this type, price change in one of the services will affect the induction level in the others. Wedig (1993), for example, analyzes price regulation in a multi-product context, with physicians able to induce demand (the physicians’ utility function depends on their economic utility and the patients’ net utility) and with high enough fixed costs for services to prevent setting an equal price for the marginal cost (due to the financing restriction). Working from these assumptions, and without considering split prices or per-capita

\textsuperscript{56} Gruber and Owings (1990) test this hypothesis for obstetricians in the U.S. during the seventies, finding that greater reductions of fertility rates in different states were related to higher rates of Caesarean sections in births, which would be an evidence of a certain degree of demand induction.

\textsuperscript{57} Some authors, such as Dranove (1988), model the relationship between doctors and patients as one of stages, during which the doctor decides his strategy of recommendation and prescription and the patient (who shares a degree of knowledge on the doctor’s recommendations, reducing the market feature of
contracts, he deduces an optimal Ramsey-type price scheme in which the highest margins over the marginal cost correspond to those services in which there is less possibility of inducing demand.

### 3.3.3. Competition among hospitals

Just as we considered the problems deriving from the existence of multiple payers that would appear in a self-financing hospital scheme competing for patients, the advantages and disadvantages associated with competition among hospitals are worthwhile examined. Although most theoretical and empirical literature deals with cases in the U.S. with special focus on cost explosion, the real possibilities opened up in Argentina by the Self-Financing Decree are worthwhile going into.

Considerations on demand inducement by physicians have important implications for the functioning of hospitals and, more to the point, for competition among them. One of the first aspects to emerge is the type of relationship operating between the physicians and the hospital. Here we have two types of model.

In one (perhaps that most widespread internationally) the hospital only offers hospitalization services and, therefore, must compete for the patients derived from outside the hospital by physicians. Many papers point to the importance of physicians’ using their own installations for the diagnostic methods that they themselves have prescribed, thus avoiding the need to derive patients to hospitals. This implies a competition between the hospitals and the diagnostic services provided by physicians in their consulting rooms. This point is significant, not only due to the possibility of competition among those doctors who possess diagnostic facilities, but also due to the implications for the hospitals. Since the physicians are the only persons able to arrange for hospital admission, it is understood that they also control the demand for hospital-based health services. In other

---

58 This is not the Argentine model, in which the hospital is not only for hospitalization, but also for outpatient treatment.
words, it is by no means clear from the literature if hospitals are competing for patients or for the physicians themselves, in those cases where they act as the gatekeepers\textsuperscript{59}.

A second kind of organization is that in which the hospital is not only used for patient hospitalization, but also for outpatient treatment. In these cases, the hospitals usually receive patients in emergency cases or through agreements with insurance companies. In this type of organization, the medical staff are subordinated to the hospital directors.

This issue leads us directly to the question of competition among hospitals, a frequent topic in the literature, that has been pointed out as one of the causes of the increase in medical costs\textsuperscript{60}. Hospitals accordingly compete (Medical Arms Race - MAR) by investing in equipment and infrastructure to attract physicians.

Few empirical studies attempt to detect whether the predominant preferences are those of the physician or those of his patients\textsuperscript{61}. The subject can become less relevant in the case of the existence of a third-party negotiating payer.

A special instance of competition is that which exists among hospitals for Medicare patients, in which payments for each type of diagnosis are determined by the State in such a way that hospitals are only left with the chance of competing for quality. Dranove (1987) argues that under this payment system hospitals specialize in certain types of diagnosis, while some patients - whose expected costs are above average (and over the price) - are derived to hospitals supported by the State\textsuperscript{62}.

\textsuperscript{59} Dranove and White (1994), pgs. 178-9.

\textsuperscript{60} Dranove and White (1994) hold that in the competition to lower prices, the quality of medical attention will not necessarily be lower, since evidence would seem to indicate that there are excessive costs that do not improve the quality of medical attention. Thus there is a chance for competition to reduce prices without affecting quality.

\textsuperscript{61} Frech III and Wooley (1992) find from an analysis made for the U.S. that the patient preferences are those that eventually predominate.

\textsuperscript{62} This practice is known as “dumping” and leads to inefficiency in the overall hospital system in the case that the hospital of last resort is less efficient than that from which the patients are derived.
Studies on competition among hospitals are not absolutely conclusive due to certain empirical problems such as the identification of the market where the hospital competes, the degree of concentration, and how to measure hospital performance.

The first two aspects are closely linked, and are important in attempting to identify the restriction operating on patients. In the case of hospitals, as they are a “business” producing different services, the number of competitors can vary significantly according to the type of service. This implies differing degrees of competition among the different types of hospital service offered by the same hospital. This may occur within a single jurisdiction or even between different geographical regions, which makes it necessary to define the “geographical area of competition” for each of the “products” handled by the hospital.

The second aspect, defining the degree of concentration, is closely linked to the question of the nature of the market. Corresponding studies have tended to apply concentration measurements developed for the study of other types of markets, and find that prices tend to rise the higher the degree of concentration.

With regard to performance, this appears to be quite difficult to define and compare among the different kinds of hospitals.

3.4. The optimal system

From the theoretical viewpoint, without considering the potential problems in the application of different mechanisms, a review of the literature would indicate that the optimal payment mechanism would be a combination of prospective payment associated

---

63 In this vein, there are papers such as that of Manheim, Bazzoli and Sohn (1994), which attempt to define the geographical market for a hospital and conclude that competition takes place through quality and services more than through prices.

64 Dranove and White (1994) highlight the personal preferences of patients in hospital choice, these preferences becoming more marked as diagnostic complexity increases, a situation that can lead to a strong preference for a remotely located hospital, thus widening the market.
with a diagnosis and cost reimbursement (Ellis and McGuire (1986,1993), Hadley (1996), Pope (1990)). The fact of incorporating a share of his costs within the “price” paid to the provider is basically justified by factors that have been mentioned such as the advantages of the pure cost reimbursement system, and the disadvantages of the pure prospective system associated with a diagnosis.

Effectively, the partial reimbursement of costs limits the incentives toward a risk selection on the supplier’s side (since he does not have to face the higher costs associated with patients who require more treatment), limits the incentives toward undersupplying of services that could exist in a pure prospective payment scheme associated with diagnosis (through a similar argument to the previous one) and, finally, recognizes (and compensates) the part of the cost differences that could be explained for reasons other than efficiency or equity.

Beyond the problem of defining the hospital payment mechanism, it seems relevant to define a hospital staff payment scheme jointly with the government’s objectives. Thus, if the real situation of public hospitals is studied, considering the meager or non-existent incentives toward efficiency (share by many government agencies), the low staff wages, and the costs of waiting time or queues faced by the beneficiaries (these could be reduced with a more efficient administration), it should be concluded that the potential efficiency and welfare gains of patients deriving from the implementation of incentive mechanisms affecting staff income are considerable (non-monetary incentive mechanisms have not been analyzed). Although in the model only two mechanisms have been considered - variable payment according to surplus on variable costs and according to output level -, the latter would seem the more appropriate due to the results it produces and its relatively

---

65 Although this alternative was not analyzed in the model, its simplicity precludes the existence of any “force” that could lead to a result of this kind (such as the existence of multiple agents with cost differences (Pope, 1990) or that of cost differences among patients generating incentives toward risk selection on the side of the service supplier (Ellis and McGuire, 1993)).

66 Low income levels imply a high $U_1$ in the model, and therefore a greater response to incentives on income, independently of the value of $g()$. A similar result is found by Hodgkin and McGuire (1994) when analyzing the potential effects of changes in the marginal and average prospective payments; and they conclude that changes in the average payment also affect hospital decisions. Specifically, a higher payment will trigger an increase in the level of intensiveness chosen, which will tend to equal the marginal benefits of intensiveness and income.
simple administration compared with others, which could involve the control by the authorities of the hospitals’ financial variables (costs, utilities, etc.).

4. Conclusions

From a theoretical point of view, bearing in mind the considerations on incentives arising from the different definitions of output, the optimal system of funds transfer should be signaled as that of prices associated with diagnosis, with a cost reimbursement component\(^{67}\) (that seeks to correct differences in severity, in the price of inputs, etc.). If the State were able to draw up health service contracts with both public and private suppliers, leaving the choice of services up to patients, this system would be similar in conception to that of a demand subsidy, and in this context one might expect that competition itself would induce public hospitals to reach greater internal organizational efficiency (with economic incentive schemes for staff, or other types of control mechanisms).

A demand subsidy system in which the State becomes the payer of its insurees would nowadays be inapplicable due to numerous factors such as the lack of information necessary to apply a system of payment per diagnosis, the poor information on specific hospital costs, the existing labor regulations for public employees, the lack of technology and computing systems necessary for minimal control, and the inconvenience of opening up competition between public and private hospitals\(^{68}\).

Beyond the possible price corrections in a prospective payment system (cost-based component, mark-up function of the possibility of inducing medical services demand, etc.), it is important to emphasize the results that indicate the superiority of the unit output price system compared with the fixed-budget or cost-reimbursement system (with any of

\(^{67}\) The Medicare prospective payment system incorporates three specific aspects of the hospital (Hadley, 1996). The prices to be received are adjusted upward according to the following points: a) a greater number of hospitalizations and hospital residents in relation to the number of beds; b) a greater number of patients qualifying for (Medicaid) assistance programs for the needy and c) a greater expenditure by the hospital schools.

\(^{68}\) At least without a restructuring period for the public hospitals and a recognition of the additional costs of the hospital schools, etc.
the staff incentive schemes analyzed). An alternative system that has not been analyzed in the model (possibly vital for the transition toward a demand subsidy system), but which in a certain way is connected with that of unit output price, is that of budgets with management commitments. According to this mechanism, the hospital and the government (or, better said, the supplier and the payer) agree to certain output levels whose supply is committed by the agent, and the government consequently transfers budgeted funds to cover this commitment. Obviously, under this scheme there must be an award and penalty system according to whether or not the hospital has attained the set objectives, to be applied in the subsequent periods (this case should be modeled as a repeated game) and to include its staff as well, so as to obtain staff commitment in line with the hospital’s global income. A change in this direction would constitute a first step toward linking funds received by hospitals with their end results.

In regard to the problem of service quality, it should be expected that competition among hospitals will guarantee a quality floor. This floor level will not necessarily materialize in the market, for several reasons. On one hand, the fact that obras sociales do not compete among themselves makes them more concerned about the price than about the quality of services they negotiate with suppliers. The freedom of choice introduced from 1997 in the social security sub-system will help to make competition for patients also operate through quality enhancing mechanisms. On the other hand, up to 1994 public hospitals were not too interested in attempting patients due to the fact that funds received from central government were not related to the number of the patients treated or the quality of the services provided. A reform such as that proposed, which would lead hospitals to compete for patients whose third-party payer is the State (and in the absence of any co-payment) should induce a greater competition in service quality.

Other aspects deserving special attention under this type of scheme would be the possible incentives toward demand inducement faced by staff (to the extent that

---

69 Mechanisms of this kind have been used in other countries in the first stages of public health system reform. See Miranda (1996), for example, in which the Chilean case is analyzed.

70 This would happen in the case of budgets associated with self-financing commitments to the extent that the objectives agreed with the hospitals are ambitious enough that they have to make an effort to achieve them.
committed goals exceeded the natural demand faced by the hospital), toward risk selection on the physicians’ side (showing a preference for those patients of lower expected cost), and the possible incentives to discriminate against public system beneficiary patients once the agreed goal has been reached. This effect would be attenuated if a certain flexibility were allowed regarding the initial commitment. If the State transfers additional funds when the hospital exceeds its targets and these marginal funds are lower than the average initial payment, the incentives to discriminate will be lesser and the incentive to seek an agreement on a very low target figure in the following periods will not be generated. An additional factor that should also be considered refers to the incentives toward undersupplying services if output is defined by diagnosis; in this sense, a certain amount of procedural standardization according to diagnosis could be an important factor.

If the above factors allow, it should be possible to advance first toward a budgetary system with financing commitment (that should be progressively refined according to the definition of output, diagnosis and State control mechanisms), and later toward a prospective price payment scheme (associated with diagnosis) that allows some corrections for severeness, input price differences, etc., as is being successfully put into practice already by other countries.

References


Chernichovsky, D. (1995), “What can developing economies learn from Health System Reforms of developed economies?”, en Peter Berman (Ed.), Health Sector Reform in Developing Countries, Harvard University Press.


IRDB, (1996), Report Nº P-6851-AR.

Kreps, D., (1990), Curso de Teoría Microeconómica, Mc Graw Hill.


