

## CHAPTER 8

# HOSPITALS AND RELATED INSTITUTIONS: IF NOT-FOR-PROFIT THEN FOR WHAT?

### THE SCALE AND GROWTH OF NFP INSTITUTIONS IN CANADA

The previous chapter traced out the way in which economists' analysis of professional practices, as special types of firms, has evolved over time. Earlier analyses attempted to treat practices as essentially similar to ordinary private, for-profit firms, differing only in their ability to limit (collectively) the rate of entry of new professionals/firms and to exercise some co-ordination of economic behaviour (particularly pricing). These profession-as-monopoly models explain some aspects of professional behaviour quite satisfactorily, but are clearly inadequate to capture its full diversity. More extended and sophisticated models of professional practices treat them as managed to seek a wider range of objectives, with broader sets of both constraints and strategies. The things that matter to professional practices, or rather to their owner/managers, include profits in the economic sense, but they also include other components of the professional's income, the style, nature and amount of work, and a professional interest in the well-being of the patient.

The process of balancing these multiple objectives, in the context of the economic freedom conferred by the legal status of the profession (particularly freedom to suppress potential competitors) generates a much wider range of questions about the relative efficiency and effectiveness of the outcomes. The greater realism and interest of such extended models, however, and the expanded range of issues they address, are gained at the expense of greater ambiguity in a priori predictive power, which must be supplemented by direct observation.

Throughout this evolution, however, models of NOFP firms have retained the conceptual framework of an owner/decisionmaker/residual claimant who directs the practice-firm's behaviour with some set of objectives in view. And those objectives include practice net revenues, whether expressed as profit or more realistically as practitioner net income. The owner (or owners) of the practice has the legal status of residual claimant to such net revenues: they belong to him/her; and this legal status is logically independent of the regulatory question of whether such residual claimant is a natural or a legal (corporate) person, or possesses a professional qualification.

While such firms are of central importance to the overall functioning of the Canadian health care system, their share in its economic activity (employment or sales) is substantially less than that of not-for-profit (NFP) firms. The leading example of these is the voluntary hospital, governed by a board of trustees on behalf of a voluntary society, municipality, or religious order who are the legal owners. But the category also includes government-run public health programs, co-operative-owned medical clinics, university programs providing health services or research, not-for-profit insurance plans, and in general any organization in which no individual can be identified as having a direct or indirect claim to net revenues.

In Canada, hospitals of all types (general and allied special, mental, and federal) made up 41.4 percent of total health care spending in 1982, compared with 21.8 percent for all professional services. Another 13.7 percent was spent in "homes for special care," split between the NFP and the for-profit sector. For-profit hospitals are virtually non-existent. Moreover, of the 12.2 percent expenditure on public health, research, prepayment, capital formation, and miscellaneous items, about three-quarters passes through the public sector and is thus controlled by NFP organizations.

The largest, most dynamic, and most extensively studied component of the NFP sector is, of course, hospitals. Table 8-1 displays the growth over time of the Canadian general and allied special (G. and A.S.) hospital sector ("allied special" includes chronic and convalescent), in terms of total spending, share of GNP, and share of total health spending. It includes, when available, other institutional spending and various components of all other health care net of institutions, professional services, and drugs and appliances (the "bureaucratized" sector). The latter, not wholly but largely NFP, has grown relatively slowly, while the hospital sector not only expanded dramatically in the general expansion of the 1950s and 1960s, but has been able to maintain some relative growth in the much more restrained climate of the 1970s. The G. & A.S. sector has had to absorb some of the workload of the provincial mental, and federal, hospitals, but it has also been able to shed some of the long-term maintenance of the elderly to the rapidly-growing "homes for special care" sector.

TABLE 8-1

**The Growth of the NFP Sector of the Canadian Health Care System,  
in \$Mn and as Percentage of GNP (Bracketed Figures), 1946-1982**

	G. & A.S. Hospitals	Other Hospitals	"Homes for Special Care"	Prepayment & Administration	Public Health	Capital Expenditures	Health Research
1946	150.7 (1.27)		--	--	--	--	--
1951	326.4 (1.51)		--	--	--	--	--
1956	380.8 (1.25)	149.0 (0.48)	--	--	--	--	--
1961	722.1 (1.82)	227.0 (0.57)	--	52 <sup>1</sup> (0.10)	81 <sup>1</sup> (0.20)	176 <sup>1</sup> (0.50)	9 <sup>1</sup> (0.02)
1966	1319.0 (2.13)	349.7 (0.56)	--	83 <sup>1</sup> (0.10)	110 <sup>1</sup> (0.20)	221 <sup>1</sup> (0.40)	31 <sup>1</sup> (0.10)
1971	2585.5 (2.74)	567.2 (0.60)	516.0 (0.55)	123.4 (0.13)	214.7 (0.23)	420.2 (0.44)	78.3 (0.08)
1976	5778.4 (3.01)	793.1 (0.41)	1437.2 (0.75)	216.3 (0.11)	484.6 (0.25)	646.2 (0.34)	139.2 (0.07)
1977	6355.1 (3.02)	573.3 (0.28)	1743.3 (0.83)	256.5 (0.12)	544.5 (0.26)	668.3 (0.32)	168.3 (0.08)
1978	6929.3 (2.98)	554.2 (0.24)	2005.1 (0.86)	251.5 (0.11)	594.7 (0.26)	765.9 (0.33)	191.8 (0.08)
1979	7696.9 (2.91)	542.5 (0.21)	2293.4 (0.87)	273.8 (0.10)	697.1 (0.26)	822.7 (0.31)	206.2 (0.08)
1980	8920.4 (3.01)	564.4 (0.19)	2710.4 (0.91)	311.0 (0.10)	762.6 (0.26)	1219.8 (0.41)	243.4 (0.08)
1981	10364.8 (3.06)	359.6 (0.11)	3530.7 (1.04)	403.3 (0.12)	872.1 (0.26)	1332.4 (0.39)	290.2 (0.09)
1982	12045.4 (3.30)	424.6 (0.11)	4117.7 (1.15)	441.4 (0.12)	952.8 (0.27)	1585.4 (0.44)	327.1 (0.09)

SOURCES: See Data Sources Appendix.

<sup>1</sup> These data are for 1960 and 1965, not 1961 and 1966.

This continual growth in expenditure share has only to a limited extent been related to utilization patterns, at least as measured by admission and patient-day loads. As Table 8-2 indicates, the fast-growth period of general hospital bed capacity, relative to the population, was in the 1950s. Rates of hospital utilization continued to increase along with bed capacity during the 1960s, but they have been virtually stabilized by the restraints of the 1970s. The major sources of expenditure growth throughout the period are not population or utilization increases, but expenditure per admission or per patient-day.

**TABLE 8-2**  
**Hospital bed Capacity and Utilization, Canada, 1946-1982/83**

	Public General Hospitals				Public Allied Special Hospitals		All Hospitals	
	Beds per 1000 Pop'n	Patient-Days per 1000 Pop'n	Expenditure per patient-day		Beds per 1000 Pop'n	Patient-Days per 1000 Pop'n	Beds per 1000 Pop'n	Patient-Days per 1000 Pop'n
			\$current	\$1971				
1946	3.87	1101.4	5.27	11.72	0.5	104.8	10.13	--
1951	4.19	1212.1	9.05	13.72	0.71	201.2	10.42	--
1956	4.67	1350.2	15.95	23.28	0.66	195.4	10.84	3202.3
1961	4.86	1434.4	24.34	32.47	0.65	205.1	10.58	3368.4
1966	5.27	1527.7	38.56	46.18	0.84	266.5	10.35	3261.4
1971	5.49	1595.9	66.61	66.61	0.92	300.8	9.78	3002.0
1975	5.66	1554.0	121.82	87.93	1.02	333.7	9.18	2675.7
1976	5.15	1435.0	140.13	94.13	1.55	500.9	7.19	2062.4
1977/78	5.25	1478.1	151.87	94.45	1.57	506.8	7.14	2075.1
1978/79	5.19	1472.9	166.26	94.90	1.57	515.9	7.10	2081.9
1979/80	5.11	1438.6	182.18	95.27	1.57	523.0	6.97	2079.2
1980/81	5.11	1474.8	208.89	99.20	--	513.5 <sup>1</sup>	9.98	2039.4 <sup>1</sup>
1981/82	--	1471.4	236.75	99.94	--	501.8 <sup>1</sup>	--	2030.2 <sup>1</sup>
1982/83	5.04	1464.2	276.67	105.40	1.51	493.4	6.82	2028.8
<b>Annualized Rates of Growth:</b>								
1946-56	1.53	1.78	10.74	7.03	2.48	4.58	0.29	--
1956-66	1.23	1.07	10.59	7.28	3.54	3.90	-0.78	-1.14
1966-76	0.77	-0.66	16.29	7.19	2.61	2.63	-1.57	-2.84
1976-82	-0.36	0.34	12.01	1.90	-0.43	-0.25 <sup>1</sup>	-0.88	-0.27 <sup>1</sup>

<sup>1</sup> These data are drawn from preliminary annual reports, which in earlier years are inconsistent with the subsequently published annual reports.

NOTE: Data from 1946 to 1975 are from Fraser (1983). Data from 1976 on are from Canada, Statistics Canada (1982, 1984). See also Data Sources Appendix. In the 1946-75 data, paediatric hospitals are included with "Public General", from 1976 on they are part of "Allied Special" as are maternity hospitals throughout. The "Allied Special" category is, however, predominately long-term care. "All Hospitals" in 1976 and after excludes mental and tuberculosis hospitals. There appear to be additional inconsistencies between the two sources; pre- and post-1975 data are obviously not comparable. Expenditure per patient-day data are adjusted by the Consumer Price Index.

Table 8-2 also shows that overall hospital capacity and use was relatively stable down to 1966, and has trended down since then, with dramatic reductions in mental and T.B. hospital use balancing the growth of the general hospital sector. Allied special hospitals, primarily long-term care, have shown the most consistent growth since 1961. But the most prominent growth pattern in the hospital sector has been the increase in costs per patient-day, adjusted for general inflation. Average annual increases in *per diems* of over seven percent, for thirty years, make up the lion's

share of hospital, and indeed of health, cost increases over the post-war period. The second half of the 1970s represents a very significant break in this trend: virtually no increase from 1976 to 1979/80. Since then, costs per patient-day have averaged 3.43 percent per year growth, after adjustment for inflation, or about half the rate of the 1946 to 1976 period.

These increases reflect a combination of more intensive servicing per patient-day or per episode, and increasing costs of hospital "inputs." The principal input to hospital care is, of course, human time and skills; thus the 1950s and early 1960s saw a substantial increase in hours worked per patient-day, from 9.2 in 1953, to 13.0 in 1965. Since the mid-1960s, paid hours per patient-day in G. & A.S. hospitals have fluctuated, from 13.3 in 1966, to 14.0 in 1968, back to 13.3 in 1971, up to 14.4 in 1975, and back to 13.9 in 1980/81.

At the same time, the incomes of hospital workers were rising substantially faster than those of the general population, a trend which continued through the 1970s. From 1960 to 1971, an index of hospital workers' wages (Barer and Evans 1983) rose at an average annual rate of 7.53 percent per year, compared with 5.58 percent for the industrial weekly wage, for a 2 percent per year relative gain. From 1971 to 1980, the same index rose 11.65 percent, and the average industrial wage rose 9.73 percent. In the late 1970s the relative gains of hospital workers were greatly slowed, to 9.08 percent per year from 1976 to 1980, compared with 8.62 percent for the average worker. Prices of non-labour inputs to hospital care, particularly food, fuel, and imported equipment (though not drugs), also rose faster than the general price level during the 1970s, but the main source of escalation is labour costs.

The patterns of increase in hospital costs in Canada find an interesting counterpoint in United States experience. The United States showed the same rapid increases in the 1950s and 1960s; between 1950 and 1970 hospital care moved from 1.35 percent of United States GNP to 2.79 percent. This compares with the Canadian percentages of 1.51 percent in 1951 and 2.74 percent in 1971. But by 1979 United States hospitals were up to 3.56 percent of GNP, compared with Canada's 2.87 percent, and the 1982 United States figure is 4.43 percent (Gibson *et al.* 1983). And, as in Canada, utilization changes have contributed little to these increases. In the 1970s, days of care per capita in short-stay hospitals were virtually static -- from 1970 to 1978 community hospital beds per thousand people rose from 4.3 to 4.6, but occupancy fell from 77.3 percent to 73.2 percent (United States Department of Health and Human Services, 1980), implying an increase in patient-days per capita of only 1.3 percent over the period.

The chief difference seems to be the "intensity of servicing" factor. In the United States, it is estimated that after allowing for all increases in prices of hospital inputs, wages and other, increases in service intensity per day raised costs per day at an average annual rate of 4.5 percent from 1970 to 1978 (Freeland *et al.* 1979). A similar calculation for Canada, 1971-80, yields 1.7 percent (Barer and Evans 1983). On the other hand, the same United States source indicates substantially smaller relative wage gains for hospital workers. The indication, therefore, is that the cost impact of increasingly sophisticated technology in hospitals, ever more widely used, with more and more expensive people and machines, has been much less severe in Canada, and that this is the principal source of our cost advantage. On the other hand, the relative income status of hospital workers in Canada, compared to the rest of the population, has advanced more rapidly than in the United States.

Whereas the universal public health insurance program contains medical care costs through its impact on fee schedules -- prices -- and physician incomes, rather than outputs, it controls costs in hospitals by limiting the growth of service intensity, of quantities of services used. If this very broad generalization is correct, it immediately raises the issue of the potential cost in health benefits from such foregone servicing -- the extent to which it is "needed" in the sense of chapter 1. If increased servicing levels per patient-day or stay would lead to improved outcomes, then the

Canadian system may be underfunded. But if they do not, as is implicit, and sometimes explicit, in the widespread concern in the United States for controlling and containing hospital costs, then why do hospitals choose to provide them?

## **MODELLING HOSPITAL BEHAVIOUR: "ORGANIC" NFP FIRMS IN A HYPOTHETICAL MARKET ENVIRONMENT**

Put another way, we can "explain" hospital cost trends, in an accounting sense, by identifying the various components of hospital expenditures -- wages, prices, utilization, service intensity -- and observing which appear to be the primary sources of increase. But this approach addresses only the "how" aspect of cost behaviour, not the "why." The more fundamental, but more difficult, explanation in terms of "why" requires us to explore the behaviour of individuals and organizations which underlies such trends. Whose decisions and acts, in response to what stimuli or constraints, are reflected in the aggregate data? If these questions can be answered, then one can begin to consider ways of influencing aggregate trends.

The most natural, indeed almost instinctive, approach to such questions by economists is to apply the theory of the firm developed in the private sector, treating the hospital (or the government bureau) as a single decision-making and transacting entity, creating what Jacobs (1974) calls "organic" models of the hospital. Hospitals *look* like firms. They have a well-defined physical plant and organizational/administrative structure, purchase productive inputs and convert them into services which are valued by customer/patients, and (depending on the reimbursement system) the "sales revenue" from such services makes up their budgets.

The role of physicians is a bit of a puzzle when this model is applied to North American hospital systems. Physicians admit patients, direct their treatment, and, through the medical staff, exercise important managerial powers, yet they are not paid by or (with some qualifications) responsible to the hospital administration. Are they customers or management? Organic models have tended to sweep this embarrassment under the rug, often with unfortunate results.

In these models, the hospital-as-firm is conceptualized as seeking some objective, or set of objectives, subject to constraints imposed by the external environment. The simplest version expresses the constraints in the usual demand curve form,<sup>1</sup> thus presupposing a reimbursement environment in which patients, or their insurers, pay the hospital on an arm's-length fee-for-service basis.<sup>2</sup> Patients are assumed to choose the type, amount, and source of hospital care they will utilize, responding *inter alia* to the price of such care in terms of out-of-pocket payments. Total care utilization then depends on the average level of hospital prices faced by users, while each individual hospital's output level also depends on its prices relative to the general average. Levels and patterns of utilization by patients with full insurance coverage are rather difficult to explain in this context; analyses applying such models usually include an implicit assumption that hospital insurance is of the private, for-profit variety, with significant patient cost-sharing. Again, the specific United States institutional environment underlies the "general" theoretical framework.

The only modification to the theory of the private firm in the simplest organic model is that hospitals are obviously not profit-maximizers. Some other objective must then be identified, to be maximized under the constraint of the demand curve. Total output is a natural alternative, on the grounds either that hospitals exist to serve the community, or that physicians (entering the model by the back door) value hospital services as "free" inputs to their own practices and also dominate hospital management. One might then postulate that hospitals try to produce as large a

quantity of output as possible, subject to the constraints of the demand curve, the production function, and input supply functions.

Such a model has strong implications. Prices will be set as low as possible, to attract patients and increase output. If the hospital must break even financially it will price at average cost, but if it has access to external subsidies, it will price below cost and expand utilization to exhaust the subsidies. Cross-subsidization may occur -- profits on some services covering losses on others -- depending on the weights with which different types of services enter the overall hospital objective of "quantity" and on the price-responsiveness of demand for different services. In general, price-insensitive services will be over-priced to subsidize the price-sensitive. But technological efficiency will be assured, since any excess costs due to inefficiency will have to be passed forward in higher prices and hence will lead to lower output.

Break-even operation assumes a static environment. More realistically, hospitals may try to earn surpluses in any one period in order to finance future growth. Since a surplus implies that prices could be lower, and the demand curve constraint implies that lower prices would draw more patients, any surplus represents failure to maximize output in that period. But the surplus can be used to expand or modernize facilities. These lead either to lower production costs for present output, or support anticipated future expansion, and may also induce an outward shift in the demand curve if capacity or "quality" are perceived by potential patients or their physicians as enhanced. Thus present output is traded for future, just as private firms invest by trading present profits for future.

The thrust of such analyses, however, is to divert attention away from the internal workings of hospitals. Whatever is done is assumed to be done efficiently, and the choice of what is to be done is determined externally by patients.<sup>3</sup> Thus, the focus of analysis comes to rest squarely on patients and, in particular, on the role of "excessive" insurance in encouraging "overuse" as discussed in chapter 2 (recall Figure 2-2). If input markets are competitive, hospital care production is technically efficient, and levels and patterns of use are patient determined, then (by prior assumption) nothing else *can* be wrong! Users of some variant of this model accordingly interpret the historical record of hospital cost escalation in terms of growth of hospital insurance stimulating patients to demand more care as out-of-pocket costs fell, thus moving the "demand curve" restraint outward, and enabling hospitals to expand supply.<sup>4</sup>

## **BRIDGING THE GAP FROM MODEL TO EXPERIENCE: THROW IN MORE ASSUMPTIONS**

The apparently contrary observation that utilization increase has *not* been the primary source of expenditure growth in the last thirty years, anywhere in North America, can be dealt with in two ways. First, one can postulate that despite the apparent existence of excess capacity in the hospital industry (occupancy rates in the United States typically average in the 70-75 percent range, in Canada about 80-85 percent), for various reasons the industry is in fact at full capacity well below 100 percent occupancy.<sup>5</sup> Thus increases in utilization stimulated by, for example, the introduction of a public hospital insurance program, either Canada's universal system in the late fifties or the United States limited form in the mid-sixties, cannot be immediately accommodated. In a United States-style system, prices would then rise to ration demand; in Canada, "direct rationing" -- waiting lists, queuing, turning patients away -- would according to this model be a necessary response.

The problem with this approach is that costs per patient-day rose rapidly after the introduction of public insurance in *both* countries, and in neither was the utilization response prominent. And the predicted associated phenomena did not appear, neither overt evidence of direct rationing in Canada in the sixties, nor a delayed surge of utilization in the United States in the late 1960s or 1970s. According to the model, the growth of insurance in a situation of capacity restraint should have led to an initial jump in United States hospital prices which would then fall back as capacity and utilization increased. The second stage has never occurred.

A second line of response, therefore, redefines output as it enters the hospital/firm's objectives. Patient-days, or admissions, per se are an implausible representation of the things which matter to hospital managements; "quality of care" is also important. The firm's objective function must reflect an interest in the dual aspects of care: quantity and quality. But quality of care requires resources, too. So an increase in costs per patient-day or per admission might be desired by hospital managements insofar as it corresponds to an increase in intensity of servicing, which *is defined as* increased quality (in this analysis), which is "good."

Furthermore, one may hypothesize that patients also value quality, so that they are willing to pay more for "higher quality" episodes or days of care. The demand curve constraint, defined over admissions or patient-days, thus moves outward to the right as service intensity increases, and there will be some optimal quality-quantity combination which responds to consumer willingness-to-pay and meets hospital management objectives. Increases in insurance coverage thus set off a two-stage increase in "quantity demanded." The demand curve for hospital care -- utilization weighted by service intensity/"quality" -- moves outward first as insurance coverage increases, lowering patients' out-of-pocket costs. In the short run, prices of care increase, and the additional revenues are used to pay for more intensive servicing per day or care episode, which further expands consumer demand, which drives prices up further, and so on (Newhouse 1970; Feldstein 1971). Thus, the standard theoretical model of the consumer, determining utilization in response to price signals, can be reconciled with the observation that hospital cost increases are not associated with utilization increases as conventionally measured, either admissions or patient-days, and appear to be driven almost entirely by supply-side forces.

The concept of quality, in this framework, must play several different roles. In the hospital's objective function, "quality" is equated with service intensity. In terms of Donabedian's (1966) triad of structure, process, and outcome concepts of quality, this is a purely structuralist definition. More (input) is better. But as everyone who has thought about the quality of health care has emphasized, what matters to patients, and to society generally, is the quality of *outcomes*. Technological extravaganzas which leave the patient no better, or even worse, off are *not* quality in any meaningful sense, however satisfying they may be to the hospital where they occur or the personnel who carry them off. There is in fact no necessary relation between intensity of servicing and quality of care in the sense of improving health status, however plausible such intensity may be as a description of hospital objectives. Analyses which equate intensity of service -- (price adjusted) input costs -- with quality, and which then assume that such "quality" is of value to informed consumer/patients for its own sake, require one either to disregard the extensive clinical evidence that more is *not* in general better -- more intervention does not always imply health improvement -- or else to assume that patients derive direct utility from health care interventions independently of their health-enhancing effect -- Münchhausen's Syndrome again.<sup>6</sup>

The extended definition of output then brings the issue of cost inflation back to insurance coverage, sidestepping the apparently contradictory utilization data by redefining utilization. Patients are assumed to choose -- demand -- higher levels of intensity of service, in full knowledge of the effects. In this context, overuse and cost escalation are only problems insofar

as knowledgeable consumers are demanding care which they would not want if they had to pay its full price, and the solution is to reduce insurance coverage. Any concept of "need" has, like the agency role of the physician, dropped out of the model.

A Canadian, or any resident of a country where hospitals are not financed through fee-for-service reimbursement by patients or their private insurers, might reasonably ask, if insurance leads to overuse, would it not be as effective to respond to the overuse by direct rationing rather than by reducing coverage? The answer is no, within this framework, for two reasons. First, consumer/patients are assumed to be the best judges of their own interests. Thus any level of overall expenditure and use achieved by direct rationing will be inferior to (or at least no better than) that achieved by price-rationing and consumer responses.<sup>7</sup> Moreover, and also by assumption, rationing techniques must themselves be costly. They must take the form of non-price barriers (*e.g.*, waiting times) which must rise to a point such that demand adapts itself to constrained supply. Such non-price barriers are a pure overhead cost to society, unlike price barriers which are also someone's income. Rising prices transfer wealth from buyer to seller; non-price barriers are deadweight losses. Since the agency relationship has been excluded by the perfect information assumption, there is no possibility in this model that rationing could take the form of professionals simply changing their criteria of what to advise, in which case the rationing *per se* could be largely costless. But, of course, in the real world that is precisely what happens.

Despite its manifold inadequacies, the analysis outlined above remains quite popular in the United States. This may be partly a result of its strong linkage with conventional economic theory; doctrinal familiarity is a value in and of itself independent of explanatory power. But it also serves important political objectives. The model presupposes a particular reimbursement environment, dominated by multiple, competitive private suppliers of insurance. And it recommends as a solution to cost escalation further reliance on the type of insurance, embodying group-specific deductible and coinsurance features, which private corporations have a comparative advantage in supplying. It has thus served to protect their market against the threat of universal public coverage which, whatever else it did, would greatly reduce the costs of insurance administration. These costs are the sales of private insurance companies (see chapter 2).

Moreover, the analysis also serves providers by diverting attention from the questions of the technical efficiency with which hospitals produce their services, or the appropriateness of the service mix. The former is assured because hospital managements are assumed to be trying to achieve the highest level of output, in terms of either quantity or quality of services, under the constraint of limited resource inputs. They will thus be under as strong incentives as any manager of a for-profit firm to use those resources efficiently in a technical sense, not to waste them in ways which do not contribute to output. The service mix which the hospital produces is also, by assumption, appropriate, since it is determined by whatever consumer/patients choose to use (the consumer sovereignty and informed consumer assumptions again).

It is literally unthinkable, in this analytic framework, that cost escalation or overuse could be the outcome of inappropriate behaviour by providers, either hospitals or physicians. Given the basic assumptions, it cannot happen, so need not be discussed. If there is an economic problem, or a misuse of resources, it is the fault of patients, wanting too much and shopping too carelessly, and behind them "the government" heedlessly providing or promoting excessive insurance coverage.

## MAKING EXPLICIT THE PHYSICIAN'S ROLE: "CO-OPERATIVE" MODELS

The two most prominent weaknesses of this "organic" model of the hospital as a single firm, striving to maximize some combination of quantity and quality subject to an exogenous demand curve, are its neglect of the role, or even the existence, of physicians, and its assumption of an arm's-length relationship between hospitals and reimbursing agencies. The implausibility of the latter is most apparent in systems such as Canada's, where a single public reimbursing agency, now typically the provincial Ministry of Health, determines each hospital's annual operating budget as well as approving, and to a considerable extent funding, the hospital's capital expansion plans. Supposedly general economic theories of hospital behaviour, originating in the United States, tend to be seriously culture-bound. But the role of physicians, in determining who gets into hospital and what happens to them there, is of central significance in all jurisdictions. In this process physicians act on both sides of the demand-supply relationship, individually influencing the patient's desired utilization pattern, and then collectively, through the medical staff, influencing the hospital's response.

Alternative formal models of the hospital have therefore been constructed which make the physician's role explicit, and indeed bring her to centre stage. The hospital can be represented as a physician's co-operative, run by and for its physician staff members (Pauly and Redisch 1973). Such models retain the organic structure -- the hospital is still managed with a specific end in view -- but that end is now the net income of the physician staff members who form the effective management.<sup>8</sup>

In this model, physicians supply a comprehensive package commodity or service -- treatment of or relief from, a state of illness. They are assumed to face the usual exogenous demand curve determining how much of this composite commodity can be sold at each price. In producing "care" they may use their own time and effort, the services of employees, or other inputs purchased directly by them. Or they may admit the patient to a hospital, where capital and labour services will be paid for through the hospital budget. But all hospital activity is directed by the physician.

The physician's net income is, in this model, the residual after hospital costs and practice expenses have been deducted from the amount patients are willing to pay for treatments. Accordingly it is in the physician's interest to ensure that hospital costs are minimized at every level of hospital activity. Not-for-profit hospital organization ensures that any surplus of payments by patients over input costs accrues to the physician. And the physician's interest in cost-minimization ensures that hospitals will neither waste resources in their production of hospital services (technical inefficiency) nor supply unnecessary hospital services, since unnecessary hospital use for any given patient condition reduces the net income left over for physicians from the treatment of that condition.

The total amount the patient is willing to pay for treatment of a given condition is thus assumed to be independent of the specific services received for it. This implicitly introduces once again the counterfactual assumption that patients are fully informed about how health care produces health, with all the usual consequences. In effect, the hospital has disappeared as an independent entity. The "physician's co-operative" model is a model, not of a hospital, but of a large physician-owned and managed clinic with overnight beds, and this clinic is assumed to be run to maximize profits, or utility as a function of income and leisure, for the physician-owners. The co-operative aspect introduces complications of the type found in the labour-managed firm literature (Evans and Williamson 1978, Chapter 6; Vanek 1977; Meade 1972, 1974). Physicians

supply labour as well as owning the firm, and managerial decisions are made by a group whose size is itself a managerial decision. Under some assumptions, this can lead to inefficient "firm" sizes. But the issues of effective *hospital* operations, of technical efficiency, and of the appropriateness of utilization, are dismissed by the assumption of cost-minimizing behaviour under a demand curve constraint.

Over-use of hospitals *could* arise if hospital care were more extensively covered by insurance than physician care, in the sense that more costly forms of treatment in the hospital might then yield higher returns to physicians. But such overuse would be defined in the way depressingly familiar in economic models -- as a response to price distortions created by insurance -- rather than being related to medical need or its absence. The idea that overuse of hospitals results from asymmetric insurance coverage of inpatient and outpatient care is a very old one, and not just among economists. It has now been largely exploded by empirical findings that expanded insurance coverage of ambulatory care is as likely to *increase* hospital use, perhaps because it encourages more initial patient contacts with the health care system.

More generally, however, the usefulness of analytic models which define away the most significant policy problems in the hospital sector, the appropriateness of hospital use and the technical efficiency of production of hospital services, is distinctly limited. The further assumption in such models, that inputs and particularly labour inputs are purchased in competitive factor markets by cost-minimizing hospital firms, sidesteps the whole question of wage-determination in hospitals. As a result, the dramatic increases in hospital workers' relative wages over the past twenty-five years are dismissed, implicitly, as the result of external market shifts in supply which are unrelated to hospital behaviour.<sup>9</sup>

## **TRANSACTION MODELS: THE HOSPITAL AS A FRAMEWORK FOR NEGOTIATION**

The very limited application of formal organic models to actual experience can be extended in an *ad hoc* manner by inserting additional variables in the managerial objective function. One might postulate that managers have a direct preference for "organizational slack" -- inefficient production -- because it economizes on managerial time and effort. Or they may derive prestige and gratification from high technology interventions ("conspicuous production"), or have a charitable interest in the incomes of their workers. In this way one can come to grips with some of the behavioural realities of the hospital sector, but the arbitrary nature of the process casts doubts on the usefulness of the formal modelling exercise itself.

Accordingly, more recent efforts to describe and understand the economic behaviour of hospitals have moved away from formal organic models of the hospital as a single entity, seeking some well-defined objectives) under constraint, and back toward an older tradition of the hospital as a setting where various groups with differing objectives interact with each other in a mixture of co-operation and competition. The characteristic "two lines of authority" in hospitals, administration and medical staff, and the problems created by conflicts between them (Smith 1958), can be represented as two separate firms-within-a-firm, in which the administration assembles inputs and produces services which are then supplied to physicians, who demand and direct such services on behalf of patients (Harris 1977). This transaction process will obviously be governed by the objectives and constraints of the two different internal "firms." But it is not a "market" in any conventional sense, nor can either set of transactors, administrators or physicians, be adequately characterized as profit or income maximizers, or as functioning at

arm's-length from each other. Their relationship is a very complex, non-zero-sum game, not a sequence of self-contained spot contracts at explicit or implicit prices. Realistic description and analysis of such complex processes can lead to very useful generalizations about how hospitals are likely to behave, and to respond to changes in their external environments, but they may never be expressed in a formal analytic framework which is either realistic or useful.<sup>10</sup> In any case, it has not happened yet.

If this assessment is accurate, one may reasonably ask why economists have spent so much effort on attempts at formal modelling. The intent, hope, was to try to go beyond informed generalizations about a particular setting, and to construct a framework which would predict behavioural responses more universally, on the basis of relatively limited information. If the results have been unsatisfactory, the problem nevertheless remains. It cannot be too often emphasized, at least not in an economics text, that *any* statement or prediction about how an individual or organization will respond to a change (or even absence of change) in its environment necessarily implies some sort of model, some set of assumed cause and effect relationships, describing the entity whose behaviour is predicted. This is true in particular of all policy analysis -- to formulate and predict the effects of a policy one must have, at least implicitly, models of the behaviour of the actors who will be affected thereby. The models may be loose, probabilistic, perhaps inconsistent, but without them one cannot make policy at all. One can only make blind stabs of unpredictable effect.

## **HOSPITAL REIMBURSEMENT: TAKE THE MONEY AND ...?**

As an illustration, we move from the abstract field of hospital modelling to the intensely practical problem of hospital reimbursement. For years, hospital managements, reimbursers, and students of health care have criticized prevailing modes of hospital reimbursement in both Canada and the United States as failing to provide incentives for efficient management of hospitals, or worse, for providing perverse incentives. Yet efforts to develop and apply alternative systems embodying incentive patterns consistent with more general social objectives have consistently been unsuccessful.<sup>11</sup>

Such alternatives have sometimes been referred to as "incentive reimbursement" systems, but of course all forms of reimbursement create incentives insofar as they make an individual's or organization's access to resources conditional on some form of behaviour. Moreover the behavioural response to any such incentive pattern will depend on the objectives of the individual/organization, and on the constraints, resource or otherwise, which bind it. An economic model of an organization is simply an explicit representation of these objectives and constraints. Accordingly any criticism (or advocacy) of a particular reimbursement system must logically rest on some implicit or explicit model of the hospital. Explicit models tend to be oversimplified and unrealistic; implicit ones are more commonly internally inconsistent.

At present, hospitals in Canada are reimbursed on some version of a negotiated budget; a system referred to in the United States as "prospective reimbursement". The process varies in detail from province to province, but in general the budget is based on a forecast of the hospital's patient work load and associated programs, which then serves as a basis for determining manpower and other input requirements. Combined with wage and price projections, these yield an estimate of the cost of total hospital operations during the year. A hospital which exceeds its budget may attempt to negotiate acceptance of this overrun for reimbursement; its success will depend on the general fiscal or political climate as well as the reimbursers' perceptions of the

reasons for the overrun. Unexpected and uncontrollable increases in work load, for example, might represent a justifiable overrun, though the allowed increase in budget would not in general be proportional to the load increase. Insofar as a significant proportion of a hospital's costs are believed to be fixed, invariant to work load, any budgetary adjustment would attempt to identify and reimburse only variable costs. Budgetary underruns are in general returned to the reimbursing agency. Some provinces permit the hospital to retain all or part of any surplus, but administrators tend to believe that such surpluses may be removed from next year's budget.

It is easy to see that almost any model of hospital objectives will, in this environment, lead to spending of all of the approved budget. Underruns will be a result of error. Overruns may or may not be encouraged, depending on the associated penalties; the hospital management will presumably have to predict the probability of an overrun's being accepted, and the cost to the institution or to management themselves if it is not. Negotiating effort will seek to maximize next year's budget, subject to the costs of negotiation itself.

Within the general pressure for "more," however, several different hospital responses are possible. The budgetary process per se encourages spending all one can get and getting all one can; it does not encourage efficient management of the resources available. Improved efficiency, as is frequently pointed out, will in the presence of given work-load targets lead to a loss of revenue. On the other hand, reimbursing agencies do try to keep in touch with actual hospital operations, so that egregiously inefficient management, if observed, might make future budget negotiations more difficult (or current overruns less acceptable). Indeed as noted above, quantity-maximizing models of hospital objectives, with or without "quality" (*i.e.*, service intensity) adjustment, predict that hospital management will still strive for maximum technical efficiency, minimum cost per service, in order to maximize output under the total expenditure constraint.<sup>12</sup> But they are silent as to the nature or effectiveness of the hospital services which will be produced. If, of course, utilization were exogenously determined by "medical need," and such need were finite as in panel (a) of Figure 1-3, then conceivably hospital output would eventually reach this limit, and any additional budget would be either given back to the reimbursing agency, eaten up in technical inefficiency or organizational slack, or passed through in workers' salaries. In fact, however, physicians' and hospitals' *perceptions* of need seem to follow more closely panel (b) of Figure 1-3,<sup>13</sup> and their direct influence over utilization enables them to keep it rising. Indeed, utilization increase, "unmet needs," is a standard and sometimes effective lever for negotiating budgetary increases. But it is difficult to separate the role of hospitals from that of physicians in this hospital utilization process.

Almost any behavioural model of a hospital, however, save one which postulated altruistic self-sacrificing administrators and Hippocratic physicians who had taken vows of poverty, will "predict" the obvious -- upward pressure on budgets -- in an environment of prospective budgetary reimbursement. The interesting questions are how reimbursing agencies acquire information and establish priorities to govern budget-setting. As noted in Tables 8-1 and 8-2, the per capita availability of hospital space and resources in Canada has remained relatively stable since 1971, with costs growing slightly faster than the overall economy. But the number of *physicians* has grown much more rapidly (Tables 1-2, 7-1). The average physician's access to hospital space has thus been significantly reduced. This in turn limits her billing capacity. If this dramatic expansion in physician supply had not occurred, pressure on hospital capacity might have been much reduced. We do not know to what extent hospitals *qua* hospitals, as distinct from the physicians who use them, can influence the use of their own services.<sup>14</sup>

## **PAYMENT BY UNIT OF SERVICE: LOWER UNIT COSTS, HIGHER TOTAL COSTS?**

At the other end of the spectrum, one could reimburse hospitals as private insurers in the United States do, on the basis of a fixed price per unit for each type of services provided. Hospitals would set charges separately for each of the components of a care episode, a *per diem* for ward care (accommodation and meals) plus a specific price for each diagnostic or therapeutic intervention such as an operation, a lab test, a CT scan, or an aspirin. In the United States setting, such charges often depart substantially from actual costs of production. Hospitals thus can, and do, cross-subsidize losses on some services with profits on others, while running an overall surplus on operations to finance new growth. The existence of such cross-subsidies indicates an absence of competitive market pressures. But a universal public insurance program could also, if it chose, reimburse hospitals on a fixed payment per unit of service basis, either individually itemized or as an inclusive cost *per diem*.

The usual argument against doing so is that it would encourage hospitals to expand output of services whose marginal cost, the cost of producing one more unit of output, is below the average charge or cost level at which reimbursement is usually set. (Variable reimbursement on a scale reflecting the dependence of cost on output would require enormous amounts of information.) A fixed payment *per diem* encourages hospitals to keep patients in longer so as to earn "profits" on late and less expensive days of stay; similarly, a fixed reimbursement per lab unit encourages additional testing at very low (so long as excess capacity is available) marginal cost. The "profits" thus earned can be used to finance expansion, or absorbed in other organizational objectives.

On the other hand, payment per service unit would presumably encourage efficiency in the narrow, technical sense of minimizing cost per unit of service, because the rate of reimbursement per test, procedure, or day of stay would be, or become, equalized across hospitals. Differences in the needs or severity of different hospitals' caseloads would be reflected in different patterns of service input per patient, but unless such services were inconsistently defined, such variation would not require different rates of reimbursement for the same service.<sup>15</sup>

In this system the "span of control" of efficient management is extended; over time more resources are assembled under their direction. High-cost hospitals, on the other hand, would have to cut back other programs to finance losses if their costs per unit of service exceeded the group rate of reimbursement. Eventually the whole group of hospitals would become more efficient, either because management was stimulated to improve, or because those whose performance remained sub-standard would see their institutions shrink in favour of growth by the more efficient. Hence the appeal of "incentive" reimbursement. It is analogous in its effects to profits in the private marketplace, which both reward the efficient directly, and provide them with additional resources at the expense of the inefficient.

The underlying implicit model of the hospital is one which assumes an unlimited appetite for more resources with which to expand services, and a willingness to trade off the effort required to minimize costs per service unit in order to free up resources for program expansion. It differs from the quantity-maximizing model, however, in removing the assumed exogenous demand curve constraint on output. Rather the hospital, perhaps acting through the medical staff, is considered able to influence the intensity of servicing per episode of care provided, independent of the out-of-pocket charges, if any, to patients. The model also assumes atomistic, not collusive, behaviour by groups of hospitals. Individual hospitals are assumed to be motivated to seek out and use ways of lowering unit costs of services; this information could then be used by the reimbursing agency in setting future reimbursement levels for groups of hospitals. The

long-term result would then be a relatively high degree of technical efficiency for the group as a whole in the production of particular services, but high and perhaps indefinitely expanding levels of servicing intensity per patient and associated cost increases.

The appropriateness of such a reimbursement program thus depends on the extent to which hospitals are, or are not, constrained in their determination of service intensity, and on the relative importance of service intensity and of technical inefficiency in contributing to expenditure growth.<sup>16</sup> Canadian reimbursing agencies, as well as government or non-profit (Blue-type) reimbursers in the United States, have generally tried to avoid such systems, presumably in the belief that service intensity is to a large extent under the control of the hospital or its medical staff, and/or that pure technical inefficiency is a less significant problem than (actual or potential) overservicing.

### **REIMBURSEMENT BY EPISODE OF CARE**

A popular suggestion, dealing to some degree with the overservicing issue, has been a shift to "case-based" or episodic reimbursement. The hospital would be paid a fixed sum for each inpatient with a particular type of problem or complaint, the amount being based on the average cost across a group of similar hospitals of treating that patient according to current standards. The hospital would thus have an incentive, not only to hold down the costs of the specific services used in treatment, but to work with the patient's physician to determine *less* service intensive ways of achieving a satisfactory outcome. Again, hospitals which were able to innovate in treatment, say by the substitution of day care for inpatient surgery, would earn a substantial surplus on some surgical cases which could be used to expand other programs (either capital or operating costs). High-cost hospitals, either technically inefficient or prone to excessive servicing, would see their resource base shrinking and might eventually disappear or be taken over by the more efficient. The approach is very attractive as a stimulus to more innovative and less costly forms of care in hospitals, and could lead to a great deal more attention being paid to the efficacy and effectiveness of the services now being carried out. The lab test which adds little or nothing to the diagnosis, the unnecessarily prolonged patient stay, would in this system subtract from the hospital's free resources for other programs, in contrast to the service-unit-based form of reimbursement where they add resources (so long as reimbursement rate exceeds marginal cost). In the negotiated budget system, improved control over service intensity frees up resources within the budget period, but these will usually be recaptured by the reimbursing agency if not spent in that period, and may also be removed from future budgets.

Episode- or case-based systems of hospital reimbursement rest on an implicit model of hospital behaviour in which perceived needs or demands for care arise external to the hospital, being generated by economic or medico-technical forces, but patterns of service within the hospital are controlled or greatly influenced by the hospital itself, or its medical staff. Otherwise, there would be no point in focussing financial incentives on the hospital as an organization. Such systems are radically inconsistent with models of the hospital which view it as constrained by some exogenous demand curve that relates its volume of utilization uniquely to the out-of-pocket payments of its patients.<sup>17</sup>

It is thus rather ironic that it is in the United States, where economic analysts (with some outstanding exceptions) seem to have had the greatest difficulty moving beyond the concept of an exogenous demand for health care, that the most extensive experiment with case-based hospital reimbursement is underway.<sup>18</sup>

The episode-based reimbursement system creates obvious and severe technical problems of classifying patients into reimbursement categories such that treatment protocols and costs are reasonably standardized *and* such that hospitals cannot manipulate the labelling system to increase revenue.<sup>19</sup> Procedural reclassification by physicians under fee-for-service reimbursement is a well-established phenomenon, and efforts to deal with it usually lead to fee schedules which are rather insensitive to differences among, for example, different types of office visits. Quality control problems may also become more severe, if one fears that hospitals will be encouraged to underservice patients, although the professional and public checks on that process probably outweigh any economic incentives. Definition of the episode is also a problem; for many chronic conditions discharge and readmission or transfer across hospitals may make it difficult to determine when one episode ends and another begins. And finally, the hospital is in effect encouraged to admit and process rapidly the simple, straight-forward and "cheap" cases, and to discourage admission of, or refer to others, patients whose age, complications, or other characteristics make a long or costly episode likely. The "cream-skimming" of which United States proprietary hospitals are often accused would be encouraged by episodic reimbursement. Yet these costly patients may have the greatest needs. The straight-forward surgical procedures, by contrast, which are frequently observed to vary dramatically in frequency across regions, and are often targeted as over-provided, would probably also be, at least at the margin, the most lucrative.

## **REIMBURSEMENT BY PERSON, NOT PATIENT, CARED FOR: CAPITATION**

This in turn leads to suggestions that reimbursement be based, not on episode, but on capitation, on persons potentially under care. The hospital could be assigned a panel of persons, based on its geographic catchment area, or it could enter into agreements to serve with individuals or groups. It would receive a fixed sum per year for each person for whom it was responsible, suitably adjusted for age, sex, and other objectively determinable personal characteristics affecting probability of hospital use. The specifics would obviously vary, depending on the form of overall hospital funding system into which capitation reimbursement was being inserted. Referral hospitals would receive smaller amounts defined over a wider population. Such a system creates maximum incentives to control unnecessary hospital use, as well as to limit excessive servicing within hospitals and to discourage waste of resources in producing services. It also, however, implies substantial reorganization of medical practice to link physicians, as well as persons, with specific home base hospitals. In smaller communities this might be no great change, but in larger cities "choice" of hospital frequently follows from choice of physician. Under a capitation-based system the potential patient would be locked in, for a time at least, to a particular hospital (or group of hospitals) and would thus have to use a physician with privileges there.<sup>20</sup>

Suggestions for capitation reimbursement implicitly assume that the hospital's influence extends to the patient admission or case generation process, while retaining the assumption that hospital behaviour will be motivated by opportunities to earn surpluses, or free resources, on some activities which can then be used for other and unspecified hospital purposes. Thus it is anticipated that the overall number of episodes of hospital care, in addition to the pattern of servicing provided for each, might be reduced (with no deleterious effects on the health of the empanelled population) if the hospital reimbursement system embodied incentives to do so.

## **WIDER STILL AND WIDER ... WHERE ARE THE BOUNDARIES OF THE HOSPITAL "FIRM"?**

At this stage, however, the implicit model of the hospital is becoming stretched out of shape. It may be plausible to think of hospital managements -- the administration -- as having a predominant influence (not total) over the production and costs of particular services. Extending management to include the medical staff collectively, it may be plausible that they can influence servicing patterns. When we begin to look at the admission process, however, we have effectively shifted our focus from management-as-administration to management-as-physicians, and it is not clear that the word "hospital" still refers to the same organization, group of people, or objectives.

If the reimbursement mechanism is to be used as an instrument for influencing the case generation process, the appropriate targets for such incentives are the physicians admitting to particular hospitals. Capitation-reimbursed organizations providing health care, community health centres or health service organizations, or, in the United States, prepaid group practices or health maintenance organizations, typically combine both physician services and hospital care within a single organizational structure. In industrial organization terms, capitation reimbursement is associated with vertical integration of production. This pattern of integration then internalizes all the appropriate efficiency incentives, to provide appropriate and effective care, at minimum cost, to a defined group of people.<sup>21</sup> The hospital as such ceases to be a self-contained firm in the conventional economic sense and becomes a component of a larger entity.<sup>22</sup>

The analysis of alternative reimbursement systems, and the combination of positive and perverse incentives with respect to efficiency embodied in each, thus points up a difficulty in our thinking about hospitals. The peculiar relationship between hospitals and physicians, at least in North America, can only be described as a form of incomplete vertical integration (Evans 1981). If the hospital-as-firm is defined to *exclude* (non-salaried) physicians, then it is a peculiar firm in which a significant share of the crucial resource-allocation decisions -- what to produce and how -- are made by people or groups who are not part of the firm's line management. If on the other hand the firm is defined so as to *include* admitting physicians, then, except for vertically integrated group practices, the management and financial structure of this broader assemblage is totally incoherent. It may be misleading, therefore, to think of the hospital as a firm at all, at least in the traditional economic sense of a set of well-defined production activities and managerial decisions under the control of a single transactor. To understand how hospitals behave, and how hospital utilization, patterns of care, and expenditures will react to different incentives, it may be as, or more, useful to study physician behaviour and objectives, as to try to model hospital behaviour directly.

The shift from organic to transaction models referred to above represents a move in this direction. Yet the transactions involved are not at arm's-length, nor price-mediated. "Rights of management," or authority over and responsibility for production decisions, are in a process of continuous negotiation.

In Canada the pattern of incomplete vertical integration is even more complex, because it must be extended to include provincial governments. These have responsibility, and power, to ensure the provision of hospital services throughout their jurisdictions. This includes both operating cost reimbursement, and separate capital funding for hospitals, giving governments a predominant influence in new investment and patterns of hospital capital formation. In the United States, by contrast, hospitals' operating cost reimbursements from public and private insurers include the costs of capital services; these depreciation and interest payments provide

funding for capital replacement and expansion. They also make public planning of hospital systems virtually impossible, as planning consists of trying to guide or prevent a hospital's spending of "its own" money. In Canada, funds for capital investment are provided, not through operating cost reimbursement, but (primarily) through separate government grants. This gives the public planning process financial "teeth." But it also means that yet another set of managerial decisions internal to the conventional economic firm, those over capital formation, are in Canadian hospitals shared with or taken over by an agency external to the hospital. From a province-wide perspective, hospitals could be thought of as operating units in a provincial public utility responsible for providing hospital care, like the individual generating plants of the provincial Hydro authority.

Of course this perspective is incomplete; hospitals are obviously owned and in theory at least managed by their own boards of trustees. But in some provinces it is explicit that the Lieutenant-Governor-in-Council may *at pleasure* appoint a public trustee to replace the board and assume any or all board functions -- or any other administrative functions. And even apart from such extreme forms of intervention, the predominant influence of provincial governments over the investment process, either within hospitals or through the construction of new hospitals, makes them, like physicians, part of hospital management.

Indeed a major problem with incentive reimbursement systems in the Canadian context is precisely that their major incentive, that of offering hospitals the opportunity to earn free surplus funds through improvements in operating efficiency, is implicitly an alternative way of guiding and funding organizational expansion, which would dilute the influence of the provincial government. Efficient hospitals, like efficient business firms, are enabled to expand and add more functions; the inefficient shrink. But suppose all the efficient managers turn out to be in regions of low growth or decline in population, and the inefficient in boom towns? Governments have a responsibility to match facilities to population needs, and no organizational framework exists for efficient hospitals to expand in areas of need. They grow where they are; an efficient hospital in Nelson, British Columbia could not use its surplus funds to buy out an inefficient one in Prince George.<sup>23</sup> Thus the fundamental concept of incentive reimbursement, of an arm's-length financial relationship which can be structured so that payments by one party induce desired forms of behaviour by another, cuts directly across the alternative framework of public authority and responsibility for the development of a hospital system. If government assumed direct ownership of hospitals, the firm structure would become clear; provinces would be running hospital public utilities on the Hydro model. In the same way, if hospitals placed all their physicians on salary, or physician groups established their own hospitals, or a private company or not-for-profit agency hired physicians and built hospitals, the resulting integrated firm would have well-defined managerial boundaries. But in the current situation, in which managerial responsibility is shared with groups outside the administrative structure, it may be inappropriate for most purposes to conceptualize a hospital as an economic firm. It may be a physical, legal, or organizational entity, but in economic terms it may be either part of a larger "firm," or a collection of smaller ones. The difficulty which economists have had in devising realistic and useful models of hospitals may reflect, not lack of sufficient ingenuity, but the nature of the industry itself.<sup>24</sup>

## NOTES

<sup>1</sup> The production function and input markets also form part of the constraint set.

<sup>2</sup> The relevance of such analyses outside the United States (or in it!) is thus questionable, but they are useful in providing a bridge between the conventional economic theory of the firm, which presupposes arm's-length output markets, and subsequent attempts to provide a more realistic description of hospital behaviour.

<sup>3</sup> The obvious weakness of this assumption may be shored up by assuming a "physician-patient pair" as the informed consuming unit. As noted in chapter 4, however, this dodge undercuts any independent theory of the provider; in this context it also assumes away the physician's role in hospital management.

<sup>4</sup> This opens the question as to why, if this interpretation is valid, insurance has become so over-extended, to which the reply is, "public provision plus tax subsidy to private provision." Presumably, in this framework, whatever level and form of insurance coverage, and associated hospital use, arose in an unsubsidized private insurance market would be optimal -- Dr. Pangloss again. There is one remaining possible wrinkle, in that the long-run equilibrium scale of quantity-maximizing firms may differ from that of profit-maximizers, with cost implications if the long-run average cost curve is not flat in the relevant range. But in fact (chapter 9) it appears to be flat.

<sup>5</sup> Since many hospitals manage to function at occupancy rates of 95 percent plus, it is unclear why 75 percent or 80 percent represents a capacity constraint. A probabilistic definition of capacity based on ensuring that the probability of hitting 100 percent and turning patients away never exceeds some minimal level, neglects the fact that most hospitals have others nearby.

<sup>6</sup> There is another alternative. One could accept the existence/validity of the clinical evidence on zero or negative health effects of "overservicing" (relative to need) yet assume that such overservicing either never occurs or more plausibly, occurs in some sense "optimally," *i.e.*, could not be reduced by any alternative institutional framework without excessive cost in some other dimension. In the complete absence of either supporting evidence, or the specification of any currently operational process yielding this result, such an assumption can be only a rather peculiar confession of faith -- Pangloss, but not economics.

<sup>7</sup> There will obviously be distributional differences. Rich people will generally get more care in a price system; poor people in a directly rationed system. Thus advocates of reduced insurance coverage and a greater role for direct charges to patients are also required to accept the existing income/wealth distribution -- which they appear prepared to do. They must also accept as equitable the pattern of distribution of burdens of ill-health, either pre- or post-insurance, a much more radical proposition which in most theoretical analysis is simply evaded or ignored. With good reason.

<sup>8</sup> While developed in the context of a North American hospital system, in which private medical practitioners treat their patients in, as well as out of, hospitals, the "physicians' co-operative" model appears to extend to systems with fully salaried hospital medical staffs. Insofar as physicians collectively "capture" the organization and manage it toward a set of objectives defined by themselves, the hospital becomes a co-op, although the definition of objectives is much more complex than simply practitioner net incomes. But of course practitioner net income is a wholly inadequate description of the objectives even of fee-for-service practitioners (see chapter 7).

<sup>9</sup> Except perhaps in the short run, if input supply functions are inelastic.

<sup>10</sup> For an excellent analysis of the historical evolution of this situation in the U.S., see Starr (1982), Chapter 4.

<sup>11</sup> The alert reader should note that in the process the implicit criteria for evaluation of hospital behaviour have shifted. The depiction of hospitals as simply firms with a rather peculiar set of objectives other than profit, but still constrained by an exogenous demand curve, carried with it the consumer sovereignty assumption that "appropriate" levels of output were to be judged in terms of their marginal production costs relative to consumer/patients'

willingness to pay. The hospital reimbursement literature embodies implicit or explicit models of the hospital in order to predict the effects of different reimbursement systems or formulae; but most such analyses judge the effect of such systems in terms of technical "need" criteria and emphasize hospital or physician, not patient, behaviour. Over- or underservicing in this literature, which may be a response to inappropriate reimbursement incentives, is defined relative to health status and need, not to willingness to pay.

<sup>12</sup> In an environment of universal insurance, however, the demand side of the utilization process in such models becomes a bit obscure. If the idea of an exogenous, price-dependent demand by independent decision-making consumers is preserved (for hospital care!?) presumably one must postulate permanent excess demand and some form of direct rationing. Otherwise, the hospital might run into a quantity constraint before the budget was exhausted. The few U.S. economists who glance sideways at Canada (as opposed to the much fewer who have actually studied it) do seem to take such consumer-driven excess demand for granted, apparently because their theoretical models require it to exist, not because it is observed.

<sup>13</sup> Or else panel (a) with  $N^*$  constantly shifting to the right as more resources become available, which comes to the same thing.

<sup>14</sup> U.S. experience with for-profit hospitals (Lewin *et al.* 1981; Pattison and Katz 1983) indicates that hospital ownership *does* affect patterns of medical practice, but this could be through medical staff selection.

<sup>15</sup> If each hospital could negotiate its own level of reimbursement for each individual service, of course, technical inefficiency (high cost production) in any one time period would contribute to future reimbursement levels. This would vitiate any efficiency incentives in the reimbursement process.

<sup>16</sup> Of course one might take the view that increased servicing was a social desideratum, at least over some range. The curve in Figure 1-3 might be perceived as having a relatively steep positive slope. In the earlier years of public intervention in hospital financing, this does seem to have been a dominant notion; the problem was viewed as one of getting more resources into the hospital sector to increase servicing. It is not so viewed now (except, of course, by physicians' associations). The shift in perspective only makes sense in the context of our health care system having reached a point on the Figure 1-3 curve which is flat, or almost so.

<sup>17</sup> One could, of course, have an exogenous, price-dependent demand for episodes of care (Stoddart and Barer 1981), but with the hospital controlling the service content of the episode. And it is generally accepted by all observers that in the U.S. it is the increase in servicing intensity, not in episodes of care per capita, which is driving hospital cost escalation.

<sup>18</sup> Since the designers and advocates of reimbursement policy rarely if ever make explicit their underlying analytic framework, and the articulators of formal analytic models rarely go beyond the academic journals to translate their work into public policy (other than those who advocate *no* public policy -- the easy out), "la conversation des muets" is likely to continue.

<sup>19</sup> This is the problem being addressed in the U.S. by the development of the system of Diagnosis-Related Groups (DRGs) which are being used as a basis for per case (episode) prospective hospital reimbursement under the U.S. federal Medicare program. This system defines a large number of categories of problems, on the basis of patient diagnosis, age, procedures undergone, and complications. Apart from the general difficulty of defining unambiguous sets of categories, the DRG system has two significant weaknesses.

First, within each DRG as presently defined it appears that the severity of patient problems, and associated costs, is highly variable. One can define a severity index, cutting across DRGs, which picks up much of this variation, but this index relies on a significant amount of subjective clinical judgement. Second, the DRG classification of a patient depends partly on the pattern of services received, enabling the hospital to influence its own level of reimbursement by its choice of interventions. (Use of a severity index in conjunction with DRGs has the same flaw; the process of index calculation is under the hospital's control, and much less well defined than patient age or diagnosis.)

On the other hand, the U.S. initiative marks an important step forward in reimbursement, which may also have

some powerful positive effects, and in any case should generate a good deal of new information about both hospital behaviour and the range of therapeutic possibilities.

<sup>20</sup> The intricacies of cross-institutional funding, and the location of payment responsibility when a patient uses services away from her "home" hospital, are critical to the actual functioning of such systems, but need not detain us here.

<sup>21</sup> Quality control, assurance that care is both adequate and appropriate, is not necessarily achieved by this reimbursement mechanism -- or any other.

<sup>22</sup> It would, of course, be possible for a group of physicians to accept reimbursement on a capitation basis for medical and hospital services, and then contract with an arm's-length hospital to provide services as directed by themselves. In this case, of course, the hospital might be reimbursed on a negotiated unit-of-service basis, since the hospital separate from the physician group would have relatively little direct influence over use.

<sup>23</sup> Over the long run, some variant of multi-unit management may modify this situation, but the problem of ownership remains.

<sup>24</sup> The development of for-profit or investor-owned hospitals in the U.S. may re-establish the hospital or hospital chain as a clear-cut economic firm. As a number of observers have pointed out, the consequences for the managerial role of physicians in such a structure will be profound. They may well find that they must either join the organization of the hospital and serve its objectives (profit) or deal with it at arm's-length as an independent economic transactor. See chapter 10 for further discussion.