

CHAPTER 4

LICENSURE, CONSUMER IGNORANCE, AND AGENCY

REGULATION OF HEALTH CARE AS PROTECTION OF UNINFORMED CONSUMERS

Public provision of health insurance is a relatively recent phenomenon in Canada, having developed almost entirely since 1945.¹ But the regulation of health care provision is much older, its basic structure having been established in most jurisdictions during the nineteenth century. In Ontario it pre-dates Confederation. The state either determines, or delegates to groups of providers power to determine, who may perform health care activities, who may provide them to the public, and how such providers must behave in the professional, economic, and often other spheres as well. There has never been a free market, or free enterprise in medical practice, in Canada, nor have private, for-profit hospitals ever played any significant role. And there has never been a serious proposal for dismantling the regulatory structure of health care provision and placing primary reliance on free competitive markets to allocate resources in this area. Only in rhetoric is there a conflict between state control and private organization of medical practice; in reality "private versus public" debates are over who shall direct the power of the state in regulating the health care market.

Present United States health care policy, by contrast, appears to take much more seriously the possibility of substituting competition for regulation in health care production and delivery, partly for ideological reasons quite separate from health care and partly because United States regulatory policies in the 1970s seem to have been much less successful than those in Canada. But it is too early to tell whether the policy rhetoric represents a real shift to competitive supply, which would require *inter alia* a dramatic reduction or elimination of the powers of the self-regulating professions, and of their influence over other providers. Alternatively, "competition" policy may serve merely as an excuse for relieving from public accountability the private organizations wielding state authority in this area (and for reduced public subsidy). So long as the state will prosecute those who contravene rules of market entry or conduct laid down by private organizations,² the industry is regulated; and references to free or private enterprise are (often deliberately) misleading.

Over the last decade there has been considerable discussion and analysis of the particular forms and extent of public and self-regulation, reflecting concerns that the balance between the interests of those regulated and of the wider society may not be appropriately struck. But the need for some forms of control over provider entry to and conduct in health care markets is (almost) universally accepted, on the ground that consumer/patients are insufficiently informed to protect their own interests. The technology of health care is sufficiently complex that consumers purchasing care in arm's-length transactions, responding only to relative prices, would make mistakes. Moreover such mistakes may be serious, and potentially irreversible -- the wrong care or care of poor quality (or no care) at a critical point can have permanently damaging or fatal results. And the possibilities for learning from the experience of others, or even from one's own, are distinctly limited. In cases of chronic illness, or of well-defined and fairly common episodic illnesses, one may be guided by past, or others', experience. But the same presenting

complaint or symptom may at different times or in different persons represent very different problems and require quite different responses.

In theory the consumer can always "buy" information with money or effort. But health problems may have a very significant time dimension, developing on a timetable which makes the acquisition of information not merely costly but impossible. At the same time, the mental state of the consumer may be such as to degrade sharply his capacity for information acquisition and informed choice. To characterize a person suddenly confronted with an acute abdomen or a severe fever as facing a conventional problem of optimal consumer choice subject to costly information is distinctly implausible. For the unconscious accident victim, it is bizarre.

Information is not only difficult and expensive (in time, trouble, and risk as well as money) to acquire, it is also costly to have. There is a genuine welfare loss involved in being aware of all the possibilities inherent in a situation. Most of us, if fully informed about health matters, would probably hesitate to get out of bed in the morning;³ we would certainly not ski, or ride bicycles in traffic.

ASYMMETRY OF INFORMATION BETWEEN PROVIDER AND CONSUMER

The significance of consumer ignorance for the optimum allocation of health care is not merely that it is extreme and pervasive, but that it is asymmetric. If both buyer and provider were equally ill-informed about the effects of a good or service, there would be no market failure and no case for public regulation.⁴ Buyers and sellers of economic or astrological forecasts are dealing in a highly uncertain commodity, but it is doubtful that the provider has any better idea of its value (to the user) than the user. Nor is there information about the value of the commodity available elsewhere in the economy which is being left unexploited in the market transaction itself. Imperfect information in these cases is part of the state of nature.

But in the case of health care, and of professional services generally, it is perceived by buyer and seller, and by the wider society, that the seller/provider of health care services has a very large informational advantage over the buyer. This asymmetry of information leaves open the possibility (or certainty) of severe exploitation of buyers by sellers in an arms-length, *caveat emptor* market environment.⁵ And private markets will "fail" in that they will give rise to a pattern of health care use -- too much of some things, not enough of others -- which yields a significantly lower level of consumer/patient satisfaction than could be achieved by alternative resource allocations. "Quackery" to the economist is a problem of resource mis-allocation, of inefficient use of human time and skills.

Of course consumers are constantly venturing into transactions, throughout the economy, with incomplete information, and the social response is not licensure of all producers (yet). But the informational asymmetry in health care, or professional services generally, takes a rather special form. Most producers know much more than most buyers about the technology of production, and the characteristics of the commodities produced. But the buyer is better informed as to what those characteristics mean *for her* than the seller is. She knows her own needs. Moreover, insofar as there is remaining uncertainty about the capabilities and reliability of a complex product, these uncertainties are specific to that product, not to the user. The buyer can thus accumulate information by questioning sellers or other users, or various sources of product-specific information. The seller can offer warranties. Or the state can establish minimum quality standards and grading systems, as for food products. There is no need to license sellers of canned beans, only to inspect the contents periodically for insect parts or rodent faeces.

But in health care the informational asymmetry extends beyond the process of production or specific characteristics of a good or service, to what its effects will be *on the user*. How will the care used interact with the patient's condition, whatever that is? The buyer is no longer the best judge of her own interests, but must rely on the seller's advice, which in turn implies that the seller accepts some responsibility for serving the buyer's interest. Their relationship cannot be arms-length.⁶

The informational asymmetry problem is conveniently captured in the health status/health care distinction. The consumer/patient values health status *per se*, not health care, but health status cannot be bought. Rather she buys health care in the expectation that it will contribute to health status. The normal consumer sovereignty assumption is that the consumer is the best judge of the value to herself of different valued commodities or states; this includes health status. But it does not extend to health care, because that is not itself of value. There is a technical relationship which is specific to each consumer and condition, by which health care affects health, and the expert provider is much better informed than the consumer/patient about the structure of this relationship.⁷

The asymmetry of information between provider and consumer is thus quite consistent with a general value postulate of consumer sovereignty as a guiding principle in the organization of economic activity. It locates the source of market failure, not in the inability of consumers to interpret the value to themselves of valued entities such as health, but in the obvious inability of consumers to be informed about the contribution of health care to the "production" of health, in specific situations. If health could be purchased directly, no difficulty would arise. Of course, patients who are unconscious, mentally disordered, in great emotional distress, or otherwise unable to protect their own interests may not be capable of making appropriate use of information even if they possess it. For these patients, who in fact account for a significant share of health care use, the application of the consumer sovereignty postulate itself is inappropriate, and someone else will have to make judgements about the value of health status as well as the efficacy of health care. It is not however obvious *a priori* that the same person is best equipped to make both types of decisions.

LIMITS ON THE PROFESSIONAL'S ROLE

For the consumer/patient who is not in some way disabled from making her own judgements, however, the health status/health care distinction has important implications for the limits of professional expertise. Superior information about the effects of health care does *not* imply a similar advantage with respect to the value of health *per se*. The dental patient may accept the dentist's expertise in deciding, on technical grounds, which teeth should be filled, which capped, and where endodontia would be successful, as well as her definition of how such services are best performed. But the patient is the best judge of whether the resulting oral condition is worth paying for. The different levels of dental care have cosmetic and perhaps comfort and convenience implications, but crowns versus fillings has no life-threatening implications and little if any in the way of external effects. For the professional to urge a particular course of treatment on a patient who is in fact informed of the relevant outcomes, is to step beyond the bounds of professional advice and into the area of marketing.⁸

In a graver setting, the choice of surgery versus radiation therapy for certain cancers turns on the values patients attach to different outcomes -- in particular to the trade-off between probabilities of short and long-term survival -- and to the quality of remaining life (McNeil *et al.*

1978). Surgery may have a higher short-term fatality rate, but a better five-year survival rate. Only the patient can judge which is to be preferred. Physicians who routinely advise surgery on the basis of the five-year survival rate may be leading some patients to the wrong choice.

Thus there remains an important role for patient information and choice of therapy in conjunction with professional advice, in selecting among available outcomes and perhaps taking account *inter alia* of their cost. But the prospects for elimination or significant reduction in informational asymmetry as to the effects of health care on health appear to be very small. Patient education is not a substitute for professional regulation.⁹ Indeed all of what presently passes for "health education," other than rather banal exhortations to eat a balanced diet and get more exercise and sleep, include recommendations to see one's doctor, dentist, or other provider more frequently, and to comply with their instructions. In any other context this would be easily recognized, not as education, but as marketing of professional services.

THE AGENCY RELATIONSHIP AND THE PROBLEM OF INCOMPLETENESS

The professional relationship can therefore be interpreted as a social response to situations in which the informational asymmetry between parties to certain types of transactions is perceived as sufficiently pronounced that independent, arm's-length transactions would permit one party to exploit the other, and to impose severe and possibly irreversible damage.¹⁰ Further, the mitigation of this asymmetry by other social mechanisms is viewed as impractical or impossible (Arrow 1963). Thus an occupational role is professionalized, in economic analysis, insofar as it involves a conflict of interest between acting as an economic principal, on one's own behalf, and acting as an agent, consciously and deliberately serving the interests of someone else even at the expense of one's own. The professional-as-agent assumes responsibility for directing the health care utilization of the patient, not as a profit-maximizing seller of care, but as an agent trying to choose what the patient would have chosen, had she been as well-informed as the professional. Thus the problem of asymmetrical information is to a degree at least circumvented, as the patient's "decisions" are based on the much better information available to the provider.

If this agency relationship were complete, the professional would take on entirely the patient's point of view and act as if she *were* the patient. All consumption choices made for the patient by the provider would be made so as to maximize the patient's well-being, subject to constraints imposed by income and prices. In this Hippocratic ideal case, there would be no remaining market failure due to incomplete information. The "physician-patient pair" would form the transacting unit on the demand side of the market, would be in fact the consumer, and this centaur-like creature would combine the information of the provider with the objectives and constraints of the patient. Assuming that providers were as close as is practical to fully informed, the market for health care would display no further informational problems.

It would, however, display some other problems. The perfect agent cannot at the same time be an economic principal -- unless she is also a perfect schizophrenic. The provider has interests of her own -- income, leisure, professional satisfaction, which are partially congruent and partly in conflict with those of the patient. The "perfect agent" would need a split brain, one half advising the patient solely in the patient's interest, the other half reacting to the patient's resulting consumption choices in a self-interested, own-welfare maximizing way. Economic analyses which assume self-interested, profit or income maximizing providers must either implicitly assume such schizophrenia as well, or else assume away the asymmetry of information problem

and the agency relationship entirely (thus removing any justification for regulation). Not surprisingly, such analyses rarely spell out their assumptions in detail.

In fact, however, the agency relationship is incomplete. There are several important sources of divergence between the objectives of the patient and those which will be sought by provider-agents advising or directing patient utilization. The primary public justification for regulation, particularly professional self-regulation, is to protect providers from competitive market pressures -- both competition among themselves and market entry by non-professionals -- which would tend to degrade or destroy the agency relationship. In the unregulated environment the consumer cannot in general tell which providers are genuinely acting in her interest and which are only pretending; and if the latter behaviour is more profitable it will prevail in the long run.¹¹ But protection from competitive forces does not remove all the sources of incompleteness in the agency relationship, and indeed provider regulation as currently applied in Canada generates additional forms of resource mis-allocation. Such regulation may still be preferable to none at all, but there is room for improvement.

In the first place, the provider is not trained to respond to, and cannot in fact know, the patient's more general interests. The idealized long-time family physician may have understood his patients as whole persons, but few professionals can hope, or care, to have this knowledge in the modern world. Thus the provider-as-agent will direct and provide care so as to improve his patients' health, not to maximize their well-being in a general sense. Moreover as noted above it is a frequent criticism of much health care practice that it focusses on illness not wellness, cure not prevention. This could be interpreted as saying that the professional provider seeks, not to maximize health status as such, but to provide any and all health care which will contribute to health for given levels of other activities and services.¹² The professional provider might reasonably retort that she cannot control all the other consumption choices which affect health or well-being generally; nevertheless the result will be over-provision of health care relative both to other factors affecting health status, and to other non-health sources of well-being. Costs of care, in the form of either direct disutility or other consumption opportunities foregone, are likely to be denigrated or ignored.¹³

Thus one source of incompleteness in the agency relationship is that the provider may respond to an incomplete or biased perception of the patient's interests. Related to this, the provider's perception of technical efficacy, the relationship between health care and health, may also be incomplete or biased. The "perfect agent" is assumed to be perfectly informed about health care technology and its application to a particular patient, the real life provider is not. And the combination of professional training with the perfectly natural human desire to "do good" (or more important, to have done good) for one's patients leads to an overestimate of the efficacy of interventions in general, relative to what can be scientifically substantiated. The urge to "do something" in the face of distress, and the self-limiting character of much illness, lead naturally to such bias in the clinical setting -- hence the long history of popular therapies later shown ineffective (Banta *et al.* 1981).

Information is, however, more usefully considered in relative rather than absolute terms. No real-life individual or institution ever attains "perfect" information about anything. A more reasonable standard or objective for social organization is that decisions should be taken, and resources allocated, on the basis of information which is "optimal" in terms of the balance between its cost and its expected usefulness. Resources can be wasted if effort and expense are devoted to gathering information which yields little or no improvement in decision-making; one can collect too much as well as too little information.

The agency relationship could therefore be considered "complete" if the professional-as-agent bases recommendations for care use on technical information on effectiveness which is

optimal in this more limited sense. The "perfect agent" is not "perfectly" informed, because that would represent an over-allocation of resources to information acquisition. A complete agency relationship would, and should, be based on some degree of incomplete information.

Unfortunately there is no a priori reason to assume that any particular organizational setting, whether competitive market or regulated environment, will generate this optimal level of information. Different forms of delivery system organization create different patterns of biases in practitioner information and behaviour. The possibilities for improving the level of provider information, and the efficiency or effectiveness of resultant patterns of health care utilization, by modifying the organization of health care delivery, are significant policy issues for discussion below.

Public support for continuing professional education, technology assessment, and research into and particularly dissemination of information about the effectiveness of current patterns of health care practice, all address perceived imperfections in the information available to providers, and reflect a belief that there is information available in, or accessible at reasonable cost to, the wider society which would improve physician decision-making and resource allocation.

PROFESSIONALISM AS A RESPONSE TO INCOMPLETE AGENCY

But the incompleteness of the agency relationship is not solely, or even primarily, a result of the imperfect information, on either effectiveness of care or patient preferences, available to agents. Even if the professional had all the information available in the mind of God, perfect agency would also require the use of that information to direct the patient's utilization, solely in the patient's interests, at complete disregard for her own. In fact, such complete selflessness is rarely found, among professionals or anyone else. The professional's influence on the utilization decision will in general respond to some blend of the patient's and the provider's interests, with the proportions in the blend varying according to the circumstances of each. Rarely will the professional seek to influence utilization in ways which she knows to be harmful to the patient, solely in her own interests. But practical health care situations are sufficiently complex and uncertain that the provider's perceptions of patient interests can readily adjust themselves to accommodate provider interests as well.

The existence of the balance is critical. The perfect agent has no independent existence as economic principal. The non-agent, arm's-length seller, however well-informed or however complex the commodity or service, looks only after her own interests and expects the buyer to do likewise. Only the professional as real-life incomplete agent has a foot in both camps and has to wrestle with the conflict of interest resulting from being a party to both sides of the health care transaction.

Nor is this role restricted to the self-employed professions. All who provide commodities or services in an environment of asymmetric information (or *a fortiori* restricted or absent patient ability to make rational decisions) are in a position to exploit this asymmetry to their own advantage and the consumer/patient's detriment. Professionalization is a process of trying to mould the objectives of providers, as well as to impose specific conduct regulations on them, so as to limit their willingness to exploit such situations. The marked segregation and peculiar socialization which is traditional in professional training cannot be explained by the technical content of the training itself; in fact the process often appears inimical to education as generally conceived. But it may be understood, in part at least, as a method of changing the objectives of

providers so as to enable them to balance the interests of patients against their own in a way not expected of, say, used car salesmen. The training is intended, not just to ensure competence, but to modify behaviour.¹⁴

In the same way much of professional self-regulation, particularly that of conduct after licensure, appears unrelated to issues of information and competence per se. Rather it is justified as discouraging conduct which would tend to break down the agency relationship. The argument is as old as Adam Smith (at least), that professionals must be protected to some degree from competitive market forces if they are to occupy positions of trust, to seek their patients'/clients' interests as well as, and sometimes at the expense of, their own.¹⁵

This is not, of course, to suggest that the sole or even the primary motivation for post-entry regulation, or the content of licensure requirements, is maintenance of the agency relationship in the public or consumer interest. Self-regulating professions collectively act as agents on behalf of the public generally, in assuring quality and ethical standards, in the same way as individual professionals act as agents for members of the public (Tuohy and Wolfson 1977, 1978). And this collective agency relation is also incomplete. There are numerous examples of self-regulatory bodies promulgating regulations whose primary or sole intent appears to be the economic or more general professional well-being of their members. Almost all students of the collective agency relationship (except for representatives of the self-governing professions themselves) have concluded that reform of the institutions of self-government is necessary to redress the balance of public and private interests served. Few, however, have suggested outright abolition.

Indeed Tuohy and Wolfson stress the parallel and mutually reinforcing nature of the individual and collective agency roles. Public perceptions of the professional-patient relationship provide the political support for self-government, just as self-government provides the institutional framework to protect the agency relation at the individual level, and they argue that neither can long persist without the other. The situation of pharmacy, whose individual agency role has largely disappeared, makes an interesting test case. Pharmacists' organizations have sought, thus far with very limited success, to recreate at least the impression of an agency role on behalf of buyers of drugs. And the self-regulatory privilege of pharmacy collectively has come under searching questioning. Nevertheless self-regulation persists.

There is yet another source of incompleteness in the agency relationship, which is for the future perhaps the most difficult of all to deal with. The perfect agent acts perfectly, but solely, on behalf of her patient or client. But when professionally directed services are collectively funded, whether by public or private insurance or public provision, the agent is simultaneously determining the level of others' resources to be devoted to the patient's well-being -- a political, distributional function as well as the exercise of technical expertise. If the health status curve (Figure 1-3) displays a very small positive slope throughout the relevant range, the professional's attempts to draw on collective resources will be unbounded (or bounded only by perceptions of the negative direct effects of health care on patient well-being).

Thus the institutional responses to the first two forms of market failure, uncertainty of illness incidence and external effects, create further problems of agency. Optimal resource allocation requires the physician to act as agent, not only of her patient, but of the wider society as well. Three sets of interests, not two, are involved. Insofar as collective financing mitigates the conflict between provider and user in the economic domain, it creates a new economic conflict between both together and the paying agency, or rather those who finance it. Efforts to train "cost-conscious physicians" represent attempts to extend the agency role to respond to this public dimension, but it is hard to be optimistic about the results.

Yet the alternatives, if providers do not accept this extended agency role, appear to be only three, and in the long run only two. One can accept, for a time, ever expanding costs, as the

United States has done. One can directly restrict the numbers of and facilities available to providers, as Canada has done. Or one can cut back on collective funding and ration access to services once again by ability to pay, as the United States appears to be in the process of doing.¹⁶ There do not seem to be any other choices. But there are, as will be touched on below, a number of different ways of expressing these choices in public policy.

THE IMPLICATIONS OF INFORMATIONAL ASYMMETRY FOR ECONOMIC ANALYSIS

The asymmetry of information between user and provider of health care is its most fundamental peculiarity as a commodity, and the source of the most serious failures of market processes as resource allocators in this field. Its implications for the theoretical analysis of resource allocation in health care, as well as for the organization of its delivery and finance, are profound.

The institutional responses to asymmetry -- professionalization, self-regulation, and the development of an agency relation between individual transactors and between the professions and society collectively -- make the conventional economic theory of supply inapplicable to these markets. Professionals themselves, and the institutions for which they act as gatekeepers, supply goods and services in response to diverse and complex motivations, under equally diverse and complex constraints which we shall attempt to analyse in more detail in subsequent chapters. What is clear, however, is that they do not behave as, and cannot sensibly be represented as, competitive for-profit firms, at arm's-length from their customers, supplying services in response to competitively determined input and product prices.¹⁷ A well-defined supply curve, showing amounts offered for sale by price-taking suppliers at each of a series of market prices, does not exist. Nor is the simple theory of monopoly applicable to the behaviour of large numbers of more or less co-ordinated suppliers, partly co-operating and partly competitive with each other.

Accordingly there is no tendency for market forces to lead to health care being offered at a price equal to its marginal cost. In the absence of insurance or subsidy, the monopoly power inherent in regulation will lead to prices above marginal cost, and hence *ceteris paribus* (which they are not) to underprovision of health care.¹⁸ This potential for monopoly exploitation is mitigated by the extension of the agency relation to the economic domain. Sliding-scale billing, *ex ante* anticipated uncollectable accounts, and overtly free care in the pre-insurance era indicated some provider concern for the patient's economic as well as physical well-being. But the extension of private insurance coverage led to steady increases in prices; United States unions frequently observed with some bitterness that whenever they negotiated insurance coverage for medical care, physicians' fees rose. And in Canada physician fees and incomes rose steadily relative to general levels of prices and incomes, all through the period of growth of private coverage (Barer and Evans 1983). Whether one interprets this phenomenon as direct agency behaviour -- physicians raising prices because the burden of prices to (individual) patients is reduced -- or indirect -- exploitation of the monopoly power conferred by self-regulation which in turn rests on agency -- the message is the same. Insurance coverage combined with non-competitive pricing behaviour leads to price escalation. Monopsonistic and compulsory insurance with negotiated prices can control this; private competitive insurance cannot.

Hence the asymmetry of information problem, and its associated institutional responses, provide powerful support for the Canadian decision to provide health insurance directly, not just

to subsidize private coverage (Evans 1983). Once the supply side of health care delivery has by regulation been largely exempted from control by market forces, direct control of prices and incomes through the insurance mechanism (or directly, as by Canada's Anti-Inflation Board or the United States Economic Stabilization Program) follows. Alternatively one must rely on a partial insurance mechanism which inflicts sufficient economic pain on users as to restrain suppliers through the agency relation itself. But a combination of self-regulation plus full or relatively generous insurance coverage without direct controls leaves price escalation unbounded.¹⁹

Even more significant, however, are the implications of asymmetry and agency for the "demand side" of the health care utilization process. The demand curve, as drawn in Figures 2-2 or 3-1, assumes that independent consumers of care are not directly influenced by suppliers in their decisions to use care, or alternatively that if such direct influence exists, its level is determined external to the market process itself. Non-agent suppliers simply offer care at the going market price; consumer/patients decide to buy, or not, after consulting their own tastes and wealth and the price they would have to pay. Perfect agents simply supply perfect, or best available, information to buyers. But real-life incomplete or imperfect agents supply information which will depend partly on technical considerations and partly on their own (and the buyer's) economic and professional circumstances. The demand curve is shifted by the advice suppliers give.

This direct influence over demand, sometimes referred to as supplier-induced demand, is precisely what the agency relation is supposed to achieve. Ill-informed buyers are protected, by provider advice, from consumption of unnecessary or harmful services (either inappropriate or poor quality) and also from failure to consume needed services. The provider directs the use of her own services, and of co-operant hospital, drug, prosthetic, and other services. Thus the quantity of care buyers wish to purchase, at any given price to themselves, depends on the advice, direction, permission they receive from suppliers.²⁰

The Naive Medico-Technical model, referred to above, recognizes this dependence, but assumes that provider advice is in turn determined by unambiguous externally set and uniform "need" standards and hence insensitive to price. Insofar as actual consumer choices are price-responsive, the reduction in use in response to prices represents "unmet need" due to compliance problems or economic barriers to care. These should be removed by subsidy or public insurance, and by consumer education. But the socially "right" level of provision is the need standard interpreted by the provider. The Naive Economic model simply ignores the existence of need, or of the professional agency relationship, and takes as socially "right" that level of care which consumers, however informed, value (at the margin) at or above its marginal resource cost.

PROVIDER INFLUENCE OVER DEMAND: NORMATIVE AND POSITIVE IMPLICATIONS

The direct influence of providers on use, however, has both normative and positive consequences. At the normative level, willingness-to-pay can no longer be interpreted as reflecting consumer preferences. Uninformed consumers may make choices which, if fully informed, they would regret. But so may those accepting professional advice. And to define patients responding to professional advice as fully informed, so as to give their utilization normative significance -- whatever the observed level of use, it must be right -- is a dodge worthy only of Pangloss.²¹ Even Pangloss would have a little difficulty in rationalizing the

dramatic swings that we observe inter-regionally and inter-temporally in specific patterns of utilization -- as for example when identifiable providers move into or out of an area. (But of course he could do it!) If providers in fact exercise significant or predominant influence over utilization levels or patterns we can no longer appeal to consumer sovereignty as a normative justification for accepting those patterns, but must instead judge them against more general criteria of what we, either collectively or as individuals, consider appropriate and are willing to pay for. In this judgement process, "need" standards have obvious appeal. The fully informed patient, whose behaviour we rarely if ever observe,²² might reasonably be expected, as a first approximation, to value care according to its contribution to health status, adjusted for its direct disutility. Consumption of *ex ante* ineffective or harmful care is a mistake, not a representation of consumer preferences for care per se, and optimal resource allocation processes should not respond to such mistakes.²³

At the positive level, the shifting of the "demand curve" in response to provider advice modifies or reverses the pattern of interactions among price, quantity, and capacity data predicted by "conventional" economic analysis. For example, an exogenous expansion in the supply of physicians or of hospital beds should, if one assumes a stable demand schedule as in Figures 2-2 or 3-1, lead to a drop in prices and an increase in quantities demanded and utilized.²⁴ Yet it is notorious in health care studies that increases in capacity tend to translate directly into utilization increase, with or without a corresponding price decrease.

Indeed in the case of physicians, there is usually a *positive* correlation between available supply and price, which can be interpreted as the result of physicians adjusting their behaviour to seek some sort of "target" income. When average workloads and incomes fall, due to exogenous increases in supply, physicians change their practice patterns to increase utilization. But if this expansion is insufficient to maintain income "targets," prices will be increased as well.²⁵

EMPIRICAL EVIDENCE OF "SUPPLIER-INDUCED DEMAND"

In the hospital sector, the direct influence of providers on use is reflected in the universal observation that bed availability is the principal determinant of bed use -- "A built bed is a filled bed" -- sometimes referred to as Roemer's Law (Roemer 1961). The direct effect of availability on use, long known to health care people, has also been demonstrated statistically in numerous jurisdictions, independently of price change (if any), demographic factors, or any other measured variables which might be expected to influence use. The relationship does not hold for all beds; paediatric and obstetric use in particular does not appear to respond to low observed occupancy rates. And even for medical and surgical beds, the "Law" is not literally true, additional capacity is not 100 percent occupied; and indeed statistical analyses suggest that occupancy rates do fall as capacity expands ("a built bed is only half filled"). But overall bed capacity emerges from study after study as the single most important factor influencing hospital inpatient utilization, and the level of bed capacity at which use would appear to stop responding to increases is double or triple current capacity or need estimates.²⁶

For facility planning purposes, this observation has the fundamental implication that there is no external "demand" standard, based on observed utilization, from which "needed" levels can be inferred. Providers will themselves determine use on the basis of, *inter alia*, available capacity. For our purposes, however, the significant point of the dependence of use on capacity is that it reflects the direct influence of providers on demand. When occupancy rates are low, physicians may find hospital access less costly to themselves, or administrators may in a variety of ways

encourage use, but whatever the linkages, physicians react by admitting more patients and/or keeping them longer. And patients accept the recommendations quite independently of any price shifts.

Hospital use is of course a reflection of patterns of medical practice. But one can also examine those patterns directly, to observe the role of provider influence independently of price. There are numerous examples of dramatic shifts over time in utilization of particular procedures or services which are traceable to provider, not patient, behaviour. During the 1970s the rate of performance of tonsillectomy, for example, dropped by half to two-thirds, all across Canada. Physician training patterns and attitudes had changed, and their criteria for recommending the operation became much tighter. But there was no change in prices faced by patients, and no evidence of independent patient choice at all. The agency relationship worked as intended -- new information or education led directly to new patterns of utilization without mediation by the price system.

Over a longer period, the steady reduction in frequency of house calls responded to some combination of physicians' professional concerns about providing adequate care in the home environment, and their economic concerns about the opportunity cost of travel and visit time. The exogenous demand model predicts that these concerns would lead to a rise in the relative cost of house calls, and a drop in the quantity demanded by consumers. But this did not happen. Relative prices of house calls did not increase; physicians simply refused to make them, and educated their patients not to ask for them. Yet another example, the dramatic fluctuation in hysterectomy rates in Saskatchewan in the early 1970s reported by Dyck *et al.* (1977) was apparently a response by surgeons to the threat of audit by the insurance commission -- of which patients were of course unaware.

Further examples could be multiplied almost endlessly, of individual procedures whose rate of performance has varied significantly in response to factors affecting providers, not patients, and without price adjustment. At the aggregate level this influence appears to underlie a sort of "Roemer's Law" for physicians as well, that more capacity leads to more use. During the period 1971-72 to 1980-81, physician supply per capita in Canada rose 2.7 percent per year, insurance coverage was universal throughout, and utilization per physician not only kept pace, but rose 1.7 percent per year (Table 7-5). Taking a longer view, from 1960 to 1981 the physician to population ratio rose 63.5 percent. And while increases in collection ratios make it difficult to identify true fee levels during the 1960s, it appears that physician workloads expanded by at least 1.5 percent per year throughout that twenty-one year period, and probably over 2 percent. There were very large swings in average physician relative incomes -- up very fast from 1960 to 1971, down equally fast thereafter -- but these were the result of adjustments in fee levels, not workloads (Barer and Evans 1983).

There may be some point at which saturation occurs, but we have yet to find it. And inter-regional comparisons suggest that there is still plenty of room for expansion. Cross-regional studies show physician use per capita varying almost directly in proportion to physician availability, even at levels of availability well above the cross-regional average.

Nor are these patterns solely the result of Canada's universal public insurance system. Similar patterns are observed in the United States, where about one-third of physician service costs are paid out of pocket. There, too, cross-sectional studies show utilization *and price* of medical services varying positively with physician availability. And the SOSSUS study of the mid-1970s (American College of Surgeons and American Surgical Association, 1975) made the point that despite an apparent surplus of surgeons (such that operative workloads were on average so low as to lead to questions of continuing competence), overall surgeon workloads, fees, and incomes steadfastly refused to fall. If anything, prices went up.

Similar behaviour is found in dentistry in Canada, despite the predominance of self-pay and private insurance relationships. A combination of community water fluoridation, (perhaps) better diet and oral hygiene, and increases in dental manpower and capacity has led to a reduction in the prevalence of dental disease and (in some quarters) allegations of an excess supply of dentists, at least relative to what the private market will support. But this surplus, if it exists, has placed no detectable downward pressure on dental fees. Rather, fees have continued to rise relative to the general price level (Table 7-3). Instead, the profession is responding with quite explicit promotion of care which is cosmetic as much as health-oriented, as well as of greatly expanded "preventive" care of undetermined efficacy. The evidence suggests, then, that in health care markets generally, prices do not appear to adjust so as to equate supply and demand, rather both price and utilization are directly influenced by providers. Part of the influence is through agency and advice to patients, and part through the control over their collective economic and professional conduct conferred by self-regulatory power.

PROVIDER INFLUENCE AND THE MEDICO-TECHNICAL MODEL: SOME DISCREPANCIES

One might anticipate that, as in the dentistry case, the role of providers in influencing utilization would be most prominent in those services which are both provider-directed and relatively discretionary on "need" grounds, and to some extent this is true. One also might expect that the initial decision to seek care for an illness episode would be beyond the provider's influence, while well-defined conditions for which care of a particular type was mandatory would leave little room for discretion. There is also some evidence to support this view. But in fact physicians and patients educate each other about appropriate cues for initiation of an episode, so that over time changing standards of medical practice can modify patient decisions. And significant inter-regional variations in treatment patterns are often found for care which might appear mandatory.

Indeed the wide variations across regions and providers in patterns of utilization or choice of technique of care (visit rates, rates of hospital admission and lengths of stay, rates of performance of diagnostic and therapeutic interventions), without corresponding observable variations in outcome, suggest that the Medico-Technical model is also unsatisfactory as an explanation of observed utilization patterns. While it is clear that providers exert predominant influence over levels and patterns of utilization, it is not at all clear what criteria guide them in this process. But on the basis of the often weak (or non-existent) links from utilization to outcome, we may presume that some patients, at least, would if fully informed choose different patterns and levels of care. There is considerable scope for improving the completeness of the agency relationship, or supplementing it with additional information or constraints.

While the influence of providers over utilization is reinforced by the regulatory structure of health care, it does not follow that an across-the-board deregulatory policy would necessarily yield superior economic performance. Such a blunt instrument would probably lead to significantly more competition in pricing behaviour. But providers' allegations that it would also lead to significant increases in provider-generated inappropriate utilization, at least within the present structure of health care delivery, cannot be lightly dismissed. The problem of informational asymmetry is real, and the agency relationship, while it may be in some respects unsatisfactory, incomplete, and inefficient, is a real response. The problem cannot be wished

away by wholesale "deregulation," much less by a phoney deregulation which would remove direct public oversight but leave the private regulatory structure of licensure and collusion intact.

Nor does the profound influence of providers on utilization imply that the result is wholly unresponsive either to economic forces or to objective "need" considerations. The fact that care patterns for similar patients vary considerably across regions, and from provider to provider, does not mean that a provider's reaction to a particular problem is arbitrary or random. The primary influence on utilization levels and patterns is obviously the provider's perception of patient health status and of the potential benefits from available diagnostic and therapeutic manoeuvres. And the degree of discretion in care permitted by prevailing best practice standards will vary greatly from one condition to another.

But there remains, particularly for diagnostic and monitoring activities, a broad zone of uncertainty in which optimal treatment and the limits of efficacy have not been scientifically established. In this zone, the provider can exercise considerable discretion before encountering ethical constraints. Economic considerations, conscious or otherwise, then can exert an effect on preferred practice patterns, and on advice to patients. Such considerations may be confounded, of course, with an ethic of "doing everything possible" for patients, subject to constraints on time and energy, so that when a physician sees fewer patients (because of a rise in the physician/population ratio, *e.g.*), she feels able to do more for each. Whether this is interpreted as more comprehensive care, or generation of utilization to maintain incomes, matters little to the outcome.

AGENCY AND THE EFFECT OF ECONOMIC FACTORS ON UTILIZATION

But patient initiative and compliance, as well as provider perceptions of patient ability to pay or comply, also affect utilization, and these too may respond to economic factors. Willingness to contact a provider, or to accept recommended treatment, is clearly related to out-of-pocket cost. The "demand curve" is not vertical. And this price sensitivity is likely to be greater for types of care about which patients feel themselves to be better informed, or less at risk of death or grave disability. Studies of ambulatory medical care use, in particular, have shown patient response to prices, although those which allow for health status differences as well usually show price responses swamped by health effects. As one might anticipate, the primary and dominant determinant of patient decisions to seek and use care, and provider decisions to recommend and provide it, is illness.

What does not follow from observed price sensitivity by individual patients, however, is either that higher out-of-pocket charges to patients will lower overall use, or that any changes in utilization which occur will be among the least needed forms of care.

On the first point, the argument from individual to group response rests on the assumption that provider recommendations will not change as patient-initiated contacts or compliance with recommendations fall. Such passivity in response to falling incomes and workloads is precisely what the incomplete agency relationship suggests will *not* occur, for any plausible formulation of provider objectives. Instead, providers will provide more care to those patients who come. As a result, overall use may not fall at all, and utilization will shift from more to less price sensitive patients or types of care. Such sensitivity is usually associated with income, and indeed studies in Canada have found that the principal effect of introducing or removing direct charges is to redirect care from poor to rich or rich to poor, whether or not the overall volume of use falls.²⁷

Whatever the response of utilization to economic factors, the relationship of this use to need is an entirely separate issue. Policy analysis using demand curves and willingness-to-pay as normative criteria for determining what levels and patterns of care ought or ought not (from a broader social perspective) to be provided, rests on a positive assumption of fully, or adequately, informed consumers as well as on individualist ideology. Asymmetry of information makes this assumption untenable, unless salvaged by the equally unattractive perfect agency model of professionals. So we cannot assume a priori that the utilization which is foregone in response to economic factors is either technically least necessary, or socially least valued.

Empirically, we find evidence that illness and low income are correlated, as are price sensitivity and low income. Moreover, several studies have shown that under universal insurance, low-income people now use more services per (age-adjusted) capita than average, whereas in the pre-insurance period they used less; and this use appears correlated with greater illness (Boulet and Henderson 1979; Siemiatycki *et al.* 1980; Broyles *et al.* 1983). On balance, then, it appears that unless prices faced by users are scaled proportionately to their economic resources *and* expected illness status, user charges will selectively deter low-income, not "frivolous," users. If such detailed scaling were feasible, we simply do not know whether any resulting impact on the mix or (if any) overall volume of health care use would be to increase or to reduce the proportion of "unnecessary" care.

The responsiveness of health care utilization to economic factors, as to any other, cannot be analyzed separately from provider objectives, constraints, and behaviour. To formulate models of "demand" for medical or hospital care, or pharmaceuticals, which exclude the direct influence of the physician, is to try to stage "Hamlet" without the Prince of Denmark. Equally, however, we shall see that one cannot explain the behaviour of providers without reference to at least their perceptions of patients' interests and conditions. The "supply" and "demand" sides of the health care "market" interpenetrate each other to an extent which renders invalid the traditional economic dichotomy of separate spheres of decision linked only by transactions at a given price. The fundamental importance for optimal resource allocation of the rich two-way flow of non-price information between provider and user is recognized by the extensive regulatory structure of health care delivery, which is intended *inter alia* to preserve and promote both the quantity and especially the quality of this flow. Resulting utilization patterns are neither "demand" nor "supply" in the conventional sense; thus we cannot use the conventional demand and supply apparatus in the conventional way, if at all. Asymmetry of information between provider and user, and the resulting professional agency relation, are the most fundamental sources of the "differentness" of health care as a commodity, in terms of the forms of economic analysis appropriate to its study, as well as the institutional framework which surrounds its organization and delivery. The implications of this asymmetry ramify as we shall see throughout the whole field of health care, prevention, insurance, education and training, investment and research. Absent informational asymmetry, and the uncertainty and externality problems could be dealt with by relatively limited public interventions in private markets.

NOTES

¹ Small-scale programs of subsidy to particular providers or patients, or direct delivery of certain services, have a much longer history.

² The public or private status of regulatory bodies in health care is ambiguous. In medicine there is a College of Physicians and Surgeons in each province, which is formally a statutory body responsible to the provincial

legislature for regulating medical practice in the public interest. The provincial medical association represents the collective interests of physicians. In practice, however, colleges appear to regard the public interest as an extension of the interests of the profession. Dentists have in some provinces simply amalgamated the two bodies, thus making explicit the "capture" of a nominally public regulatory body by a private association.

³ Were it not for the possibilities of decubitus ulcer or thrombosis!

⁴ Although there may be some justification for the banning of potentially dangerous commodities whose specific effects are unknown to both buyer and seller. Fortune-telling has at some times been banned on related grounds, though not, as far as I am aware, economic analysis.

⁵ Trebilcock and Shaul (1983) point out that in most parts of Canada the provision of *mental* health services is much less tightly regulated than that of physical health services, and that the background, training, and methods of providers are correspondingly much more diverse. They ascribe this difference to the absence of asymmetry of information in this field. Providers (collectively at least) have no more information about what "works" than patients do. There is no evidence that any one form of training or set of methods achieves superior therapeutic results.

Thus they argue, consistent with our discussion, that public regulation should *not* take the traditional form of licensure of self-regulating professions, but should focus on control of dangerous or exploitative practices. Subject to such control, the "market" for mental health services should be left open to a number of different types of competing therapists and methods.

⁶ The interaction between service and patient condition also introduces "first-mover advantages" (Williamson 1975) for the supplier, which undercut the potential competitiveness of supply. Before recommending care for a particular episode, each provider must acquire information about the patient which may be extensive and expensive. Thus patient "shopping" among providers involves substantial real costs (to patients, providers, or insurers) of replicated effort, and/or impaired effectiveness of care. Indeed the stress by providers on the desirability of a continuing relationship reflects the fact that the effectiveness of care in a particular case may be related to an extensive past history which is difficult, if not impossible, to communicate among providers. Nor, in a competitive, arms-length market, would providers have any incentive to do so -- quite the contrary. "Professionalism" is at least some check on proprietary control of patient-specific information, though it can also be used to control patient access to information.

⁷ As before, $U = U[X, \dots, HS(HC, \dots)]$ the consumer's level of well-being depends *inter alia* on health status. The partial impact of health status on well-being, $\partial U/\partial HS$, is (by consumer sovereignty) information privileged to the consumer. But the technical relation $HS/\partial HC$ is not. The informational advantage of the provider is with respect to the structure of the functional relation $HS(HC \dots)$.

There are in the literature analytical models of health care use which emphasize its relationship with health as the ultimately demanded entity, but which then assume that the consumer is fully informed (or at least as well informed as the provider) about the structure of the $HS(HC \dots)$ function itself. No justification for this counter-factual assumption, other than analytical tractability, comes to mind. If it were justified, it would undercut any rationale for regulation of the supply side of health care.

⁸ In chapter 1 this same distinction was discussed from the point of view of a society deciding how much to allocate to health care -- again the ultimate question is not a professional one.

⁹ The strong interest in information and self-care among certain groups of females in particular appears to be a response to predominantly male providers who are alleged to use expertise in the technical domain to try to impose their own preferences in the outcome domain, *i.e.*, to enforce by control of information and access to other facilities or commodities, choices which lead to the health outcomes providers think patients ought to want, whether or not they do.

¹⁰ The Ontario Professional Organizations Committee (Trebilcock *et al.* 1979) uses a more general definition in which professionalization is intended to protect "vulnerable interests," not defended in an arm's-length transaction, whether or not they are parties to the transaction. But the significance of this extension is primarily outside health care.

¹¹ Some theoretical analyses assume that informational asymmetry disappears in the long run, that patient information about specific services, or providers, progressively improves in such a way as to drive out or at least limit quackery. But this class of assumptions lacks any support beyond analytic convenience, and loses the essence of the patient's problem. For her there may not *be* a long run!

¹² The patient's well-being $U[X_i, \dots HS(HC, X_i)]$ depends on health and other things. The "perfect agent" in conjunction with the patient, would direct use of HC and all other X_i, X_j , so as to maximize U (for a given budget constraint and prices, but how would perfect agents also set the prices of their own services?) The real-life provider may try to direct health care use to a point where $\partial HS/\partial HC = 0$, taking little account of the internal structure of the $HS()$ function (prevention, in a general sense) or of the patient's marginal rate of substitution between HS and the various other X_i .

¹³ This is particularly apparent in the case of insured care, when providers recommend care of limited usefulness (diagnostic testing, *e.g.*) because it is "free." It is not, however, free to consumer/patients collectively, as it must be paid for in premiums or taxes.

¹⁴ There are some less idealistic interpretations, in terms of generating the collegiate loyalty necessary for a large group to function as a cartel, as well as psychological interpretations focussing on dealing with uncertainty and stress. Different interpretations are not mutually exclusive.

¹⁵ There is in the economic literature a discussion of the "principal-agent problem" which should not be confused (as its students occasionally do) with the professional relationship and the phenomenon of incomplete agency. In the principal-agent problem, the agent performs certain acts which affect the well-being of both principal and agent. The principal cannot directly control the agent's behaviour, but can within limits influence the way in which the agent's well-being depends on the agent's acts. The principal's problem is to influence the agent's payoff such that the act(s) which most benefit the principal are also those which most benefit the agent and hence will be chosen.

In this formulation the *objectives* of principal and agent are strictly separated; each intends to serve only her own interests. What distinguishes the professional agency relation is that the professional includes part at least of the patient's/client's interests in her own objectives. If, for example, there were some act which the agent could perform which would benefit the agent at the expense of the principal, but which the principal (or anyone else) could never detect, the professional-as-perfect-agent would not do it. The professional-as-imperfect-agent might, but not always, or some would and some would not. In the usual formulation of the principal-agent problem, the agent always does it. Such an agent is not a professional and no agency relationship, as the concept is used here, exists.

¹⁶ Whether this will contain overall costs, or merely shift the pattern of care use and the distribution of economic burden within an ever growing total, remains to be seen.

¹⁷ They can of course be so represented, and often are, but the resulting analyses are usually more misleading than enlightening. Elegance may be gained, but relevance is lost.

¹⁸ The case of the supplier as perfectly price-discriminating monopolist is a bit more involved, as the marginal unit of care sold will be priced at marginal cost, and hence apparently the level of provision is optimal. But it is a *different* optimum from the competitive one. The wealth transfer which results from perfect price discrimination will shift the *MRS* between income and leisure for self-employed providers, raising the opportunity cost of own-time, which is the principal component of the cost of such firms. So, if a non-discriminating self-employed monopolist begins to discriminate, output will indeed shift to where (marginal) $P = MC$, but the whole MC curve will rise. The result could be rising or falling output. This ambiguity reflects the flabbiness of the Pareto optimality criterion, and its sensitivity to wealth distribution. The non-discriminating monopolist is producing less than (Pareto) optimal output, but when she shifts to perfect price discrimination the result is optimal output even if output falls!

¹⁹ This conclusion would not necessarily apply in the event of a radical reconstruction of the supply side along the lines of the U.S. HMOs (Enthoven 1980), but this raises issues deferred to subsequent chapters.

²⁰ This approach to health care utilization most emphatically does *not* assume that the decision to utilize care is insensitive to price. *Ceteris paribus*, including among the *ceteris* the consumer's perceptions of health status and the

effectiveness of health care, one would expect use to respond negatively to price in the conventional way, through increased reluctance to initiate episodes of care or increased questioning of/non-compliance with provider advice (though the elasticity of response is unlikely to be large overall). The critical point is that the *ceteris* are *not paribus*, since the patient's perceptions of both health status and health care efficacy are directly influenced by the provider. And these influences, in the incomplete agency framework, will depend on the providers' circumstances, economic, professional, and personal. Thus an increase in provider supply, *e.g.*, lowers average incomes and workloads, and the response is to advise increased amounts of servicing. The "demand curve" shifts laterally (Stoddart and Barer 1981; Barer, Evans, and Stoddart 1979).

²¹ "All is for the best, in the best of all possible worlds" -- which follows from the assumption that Divine Providence is both beneficent and omnipotent. The significance of Pangloss is that, given his assumptions, he cannot be refuted. But he can be held up to ridicule, as Voltaire did.

²² In particular we do not observe her among physicians, or their families. They are particularly vulnerable to overestimates of the efficacy of therapy, for quite apparent reasons.

²³ Of course in an uncertain world some forms of care will turn out after the fact to be useless or even harmful -- some level of error is to be expected. But care which could reasonably be *forecast*, on the basis of presently available or reasonably accessible information, to be useless or harmful, clearly represents inappropriate care.

²⁴ Universal first dollar insurance is in theory no bar to this price adjustment; suppliers could always pay rebates, in kind if not in cash.

²⁵ The "target income" model is justly suspect among economists, because of its rather *ad hoc* flavour and its failure to explain the origins of targets. Further, physician real incomes are not as stable as it seems to suggest. It can, however, be recast in a more sophisticated version based on a general model of physician utility maximization (see below, chapter 7) which allows for provider discretion over both prices and quantities. Such a model predicts behaviour rather similar to the "target income" approach.

²⁶ Of course use can respond to capacity in the trivial sense that when there is a "shortage" -- desired use exceeds capacity available -- more capacity will permit proportionately more use. But "Roemer's Law" applies in situations of average bed occupancy from near 100 percent on down to 50 percent or below.

²⁷ Some studies show an overall fall, others do not (Beck and Horne 1978; Enterline *et al.* 1973). The difficulty in interpretation is that an increase in physician incomes would tend *ceteris paribus* to lead to less demand generation and lower overall use. So a pure demand-side, patient response can only be measured if physician incomes are held constant in real terms. The RAND experiment in the U.S. is designed such that the effects of patient responses to charges which are felt by any one provider should be trivial or undetectable; accordingly the *ceteris paribus* assumption should apply to provider behaviour, and the estimated responses (Newhouse *et al.* 1982) should represent the slope of a stable demand curve for part at least of the population. Unfortunately, for just that reason, such results give no information at all about what the effects would be of a general program of increased charges to patients, which *would* (if it affected use) lower each provider's income and workload. The experimental results, by design, cannot be generalized.