

CHAPTER 5

MARKET FAILURE AND THE EVOLUTION OF HEALTH CARE INSTITUTIONS: A HISTORICAL PARABLE

MARKET FAILURE, POLICY "THERAPIES," AND SIDE EFFECTS: A FIGURATIVE SUMMARY

The analysis of the previous chapters can be summarized and synthesized into an "historical parable," a very much oversimplified description of how the Canadian health care system has developed through time to reach its current state. In this "parable," institutional evolution can be represented as a challenge and response. Each set of institutions arises in response to problems with the preceding set, and in turn generates its own problems.

The general theme of the story is that the nature of health care as a commodity, its intrinsic peculiarities discussed above, leads to certain distinctive forms of "market failure." Such "failure" means that the organization of health care production and distribution through unregulated private markets -- purely voluntary exchange processes -- governed by the price mechanism, leads to unsatisfactory outcomes. Resources are not allocated to or used in health care production, and/or the care produced is not distributed among users, in a way which most of the members of society find acceptable. Accordingly various forms of intervention, institutional responses, arise in both the public and the private sectors, which either supplement or supplant private market relationships. These interventions -- regulation, public subsidy, insurance, private charity, etc. -- are anticipated to lead to patterns of resource allocation and/or output distribution in the health care sector which are more generally acceptable to the wider society.¹

But these responses have the problems common to therapy in other fields -- they have harmful side effects. Furthermore, multiple therapies for multiple problems result in interactions which often accentuate these side effects. Side effects give rise, in turn, to further institutional responses which have their own strengths and weaknesses, and thus the system evolves through time.

There may or may not be a "final answer," an optimal health care system. Many such have been suggested, and located in hypothetical futures or imaginary pasts, but they tend to differ radically from each other. At present it seems fair to say that if a "best" way of organizing health care delivery exists, it has yet to be found -- in Canada or elsewhere.

But there do appear to be better and worse approaches to delivery system design. And while each country's experience is rooted in its own historical and cultural background, it does not appear that that background wholly determines the outcome. Reasonable people may, by taking thought and devoting effort, improve existing systems; conversely, by failure to do so, they can bring on, or permit, deterioration.² In making international comparisons of health care systems, Canadians tend to be rather smug, and perhaps with some justification. But reasonable satisfaction with the present is no guarantee of the future.

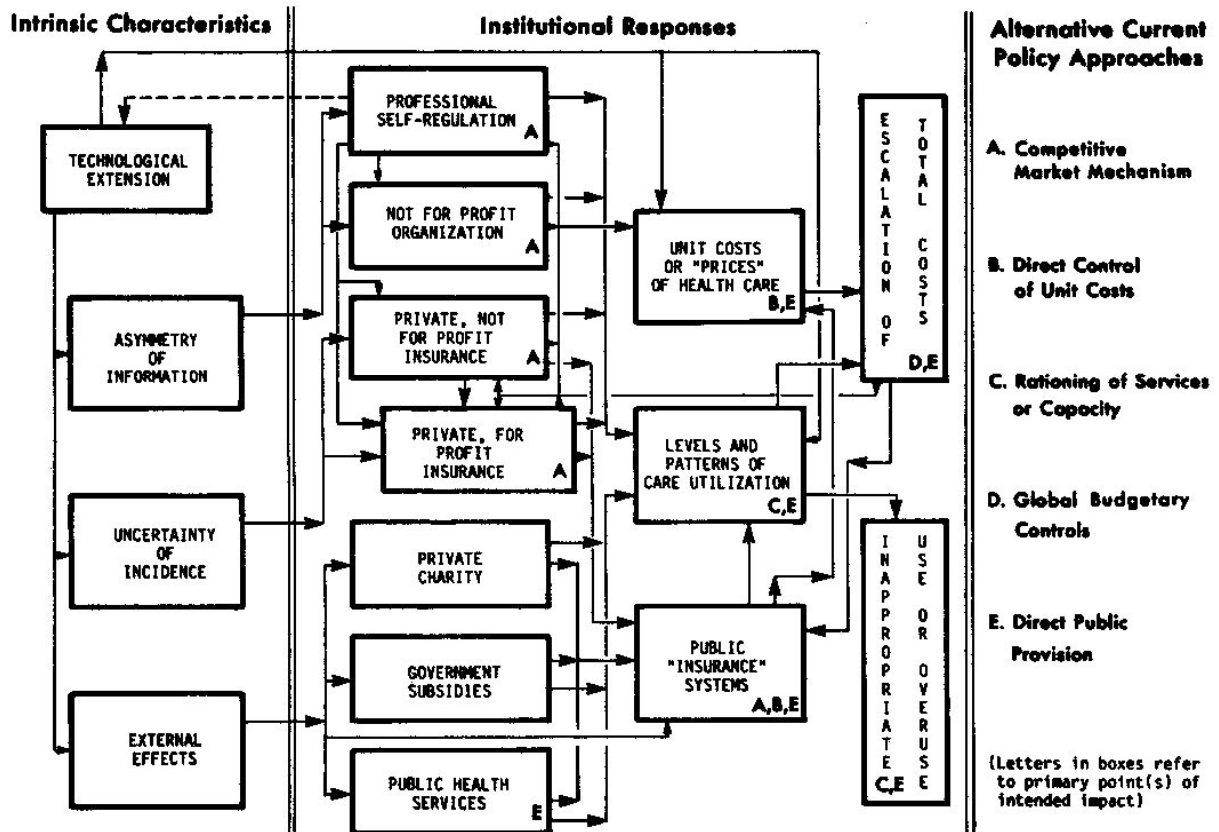
Looked at another way, while the intrinsic characteristics of health care analysed above may have dictated some pattern of regulatory intervention, they did not necessarily dictate the pattern we have. Nor is there any reason to believe that the pattern of regulation we have is the best which could have been established, even in Canada. Whatever is actual may indeed be rational, but is not necessarily optimal.

What does seem clear, however, is that the market failure problem and the process of institutional evolution are an interactive totality. Theoretically optimal "solutions" to the specific problems of uncertainty, externalities, or asymmetry of information, analysed in isolation from other sources of market failure whether intrinsic or derivative, provide few useful guides to policy. There is little point in controlling the patient's arthritis with a therapy which induces a bleeding stomach ulcer, or even of controlling that by contributing to kidney failure.

The process of institutional evolution may be traced out through time -- hence the "historical parable" -- so long as it is clearly understood that the tracing *is* a parable. The complex and messy realities of the historical record can (and will) be crammed into a preconceived analytic framework -- for which no apologies are offered. But the result does not, it is hoped, do undue violence to what actually occurred in Canada, while it provides a convenient way of interpreting that experience. Moreover the framework is general enough that, with only a bit more violence, it can be applied to other countries as well.

Figure 5-1 presents this framework. On the left-hand side are the various intrinsic characteristics of health care as a commodity which give rise to "market failure." In the middle are the various institutional responses to these, their interactions, and their consequences. On the right are a number of potential further policy responses, which are currently being tested or are under discussion in Canada and elsewhere. These are not so much specific policies as clusters of policies, themes, or policy stances. Each is then related back through its label -- A, B, ... E -- to the particular stage(s) in the analytic framework which would be its primary focal point.

FIGURE 5-1
Historical Evolution of Health Care Delivery in Canada



THE CAUSAL STRUCTURE OF INSTITUTIONAL EVOLUTION

To the upper left of Figure 5-1 is the process of technological extension.³ While this factor appears in the figure to be the underlying dynamic force in the whole evolutionary process, it is important to remember that it is public *perceptions* of the state of technology, not that technology itself, which give rise to social responses. Looking backward, we may believe that earlier perceptions were in error, and that many older technologies were useless or harmful. But the validity or otherwise of these perceptions, when judged against "absolute" standards of truth (*i.e.*, the perceptions of our own day), is quite separate from their institutional consequences.

Technological extension, over time, has tended to enhance all three of the intrinsic sources of market failure in health care. As the range and complexity of diagnostic and therapeutic interventions expands, and the subtleties of the human organism unfold, the information gap between provider and user of services becomes even larger. Providers specialize and sub-specialize, knowing "more and more about less and less," but, collectively, their information progressively expands relative to that of the patient. And the more powerful and specific their interventions, the more danger follows from misapplication.

By itself, such growth in provider knowledge need not lead to market failure; as noted in chapter 4, it is not the complexity of technology per se, but the difficulty for the user in determining the effects of its consumption on herself which creates problems. And one can think of specific changes in health care which have made, or could make, some forms of self-diagnosis

or self-treatment more feasible. But in practice the extension of technology has been such as to accentuate asymmetry of information about the consequences of use.

While the principal impact of technology is on informational asymmetry, it also bears upon uncertainty. Fluctuations in health status obviously generate corresponding fluctuations in well-being, and such uncertainties are per se undesirable. In the absence of effective interventions, however, health *care* expenditures will not show similar fluctuations. As noted above, neither colds nor sudden death represent a need for care. But as the range of possibly effective interventions expands, so does the variance of "needed" care and of expenditure. This, too, need not be so in principle; some technologies (polio vaccine, *e.g.*) may lower the variance of financial outlays. But again, the actual pattern of health technology has tended to increase such variance.

The external effects of illness may or may not be accentuated by technological change, but the policy responses to them certainly are. When effective technologies do not exist, externalities may be expressed through prayer and care, or isolation and quarantine. But their impact on health care use is small. As more is perceived to be possible, so more is "needed." Indeed, external effects may flow specifically from the discrepancy between the actual and the possible. The regret, and perhaps outrage, felt by others (as well as oneself!) over a potentially avoidable adverse consequence exceeds that over an uncontrollable "visitation of God." So "needs" are defined in large part by prevailing technology.⁴

The primary institutional responses to these characteristics are displayed in Figure 5-1. Delegation of public authority to the self-regulating professions to control both entry to occupations and service markets, and conduct of persons/firms in those markets, is a response to informational asymmetry, as is the promotion of not-for-profit modes of organization. Private insurance programs respond to uncertainty of illness/expenditure incidence, though as noted in chapter 2, the earlier private not-for-profit insurers responded also to issues of equity, and to professional objectives, in a way that the later for-profits did not. External effects lead to private charitable efforts, to public subsidies for specific individuals or organizations, and to direct public provision of services.

Each of these responses, however, leads to "side effects" (which may in some cases have been specifically intended). Self-regulation by providers confers upon them, individually and collectively, enormous power to control "markets" for their services by a very diverse collection of overt and subtle mechanisms which would be impossible and, in many cases, illegal (indeed, criminal offences) in "ordinary" industries. This has tended to drive up the costs per unit of such services, in a number of different ways, and thus to enhance the incomes of their suppliers.⁵

Non-profit organization is alleged to lead to the same result, permitting "organizational slack" in management, gold-plating of production to the satisfaction of providers rather than users, and rates of wage and salary payment which are above competitive market rates. For-profit firms in a competitive environment would be more strongly motivated to hold down costs, and would be forced to hold down prices.⁶

These unit-cost-inflating tendencies are then reinforced by insurance, of whatever type, which spreads the consequences over large groups and reduces or removes the connection between provider costs and market share. The structure of the insurance contract will determine whether the link is merely attenuated, or severed entirely. Furthermore, the self-regulatory power of providers can be, and has been, used to influence insurance markets so as to encourage the spread of forms of insurance which maximize provider discretion over pricing and patterns of servicing.

But the importance of the insurance process is easily (and frequently) over-emphasized. If providers collectively use their regulatory power to suppress competition over market share, then

the hypothetical significance of alternative forms of insurance for the behaviour of individual providers becomes irrelevant. What does seem clear, however, is that the total flow of resources into the health care industry is strongly affected by the level of insurance coverage. This effect shows up partly in increased utilization, but also, to a great extent, in increases in service prices/costs and provider incomes. The relative incomes of physicians and hospital workers rose dramatically throughout the 1950s and 1960s, as first private and then public hospital and medical insurance coverage extended across Canada; the growth of private dental insurance in the 1970s has done the same for dentists (see Table 7-4 below, and Barer and Evans 1983).

The interaction between self-regulation and insurance in their effects on utilization is less clear-cut, particularly since one must be careful about the choice of hypothetical alternatives against which to measure such effects. The role of providers in influencing ("inducing," "steering," "controlling") utilization is well established, and it appears that, as a result, the level and mix of health services used is more sensitive to levels of supply and to the professional and economic objectives of providers than it would be if determined by the free consumption choices of hypothetical fully informed buyers. But such buyers exist only in abstract economic models; in reality, the choice may be between professional control and for-profit control in a market of ill-informed users. As we shall develop in more detail below, both theory and experience with for-profit organization (the American clinical laboratory industry, *e.g.*, and, increasingly, the American hospital industry) suggest that professional control and not-for-profit organization may, on balance, *restrain* overall output, or at least lead to less rapid escalation, and may yield a mix more closely related to needs than does the for-profit alternative.

The institutional responses to external effects all tend to influence utilization of health care. In the short run all tend to stimulate use, though public provision of (effective) preventive services, such as immunization, may lower other forms of use in the longer run (including further prevention, as in the case of smallpox). They may also affect mix -- disease-specific charities, or public provision of specific services. But general public subsidies, or later, public insurance financed from general tax revenue (whether or not including compulsory "premiums" unrelated to risk) serve merely to influence overall utilization and unit costs.⁷

Indeed, the most general utilization effects probably arise from public subsidy of the training of health personnel, of new facilities construction, and of new technology acquisition. Insofar as utilization tends to be driven by capacity -- more doctors \Rightarrow more medical care use, more beds \Rightarrow more hospital use -- such subsidies probably have a greater long-run impact on utilization than any other factor. Once again, however, different institutions interlock. If the self-regulating professions could not set very high mandatory entrance standards, entry to the industry (as opposed to particular professions) would not be as dependent on large public subsidies. And if insurance, public or private, did not exist, utilization might not be so dominated by capacity.

FEEDBACK LOOPS IN INSTITUTIONAL RESPONSES: CIRCLES VICIOUS OR OTHERWISE

These effects generate some interesting feedback loops. Both increased use and increased unit cost increase (obviously) total expenditures. This, in turn, encourages further insurance coverage. But the increased cost of such coverage also stimulates the development of private for-profit insurance, offering experience-related coverage to low-risk groups and building on the methods of the non-profits. United States history suggests (Canada had not progressed as far

down this road when we "went public") that as insurance evolves from non-profit, community-rated to for-profit, experience-rated, self-governing providers will attempt to control the form of insurance. Efforts by whatever form of private insurer to influence the content or the costs of medical practice will, as far as possible, be suppressed (Goldberg and Greenberg 1977).⁸ Thus private insurance is also cost-expanding.

The combination of public subsidy programs, and private coverage which is both incomplete and progressively more expensive, creates pressure for public intervention. *Some* form of public insurance seems to follow automatically from the observation that private insurance cannot cover those most in need, or exercise any control over costs or use. At this point one can go to universal or to selective public coverage -- Canada or the United States. Both seem to be, or to have been, successful in supporting provision of care for a significant proportion of the population -- aged, poor, chronically ill -- who would otherwise receive inadequate care or do without. The partial approach does, however, provide much less comprehensive coverage, and drops a non-trivial proportion of the population "through the cracks" -- uninsurable by any public or private program (Aday *et al.* 1980).

The key difference seems to be that universal public insurance permits (though it does *not* guarantee) the collective control, both of unit costs of care, and of the overall level and mix of utilization. Thus, while all countries see themselves as facing "crises" of escalating health costs, the problem is measurably and significantly less severe in countries with a universal, sole-source-funded public insurance or delivery system (Canada, the U.K.) than in those with multiple sources of insurance and/or delivery, whether public or private (the United States, Germany).⁹ Indeed, provider spokesmen in Canada claim that public insurance leads to "underfunding" -- too low a level of costs -- because the share of national income devoted to health care has grown more slowly in the 1970s than in previous decades, or in other countries. The impact of the *introduction* of public insurance, however, was clearly to raise service costs, provider incomes, and health expenditures, though its subsequent application has tended, at least in relative terms, to hold them down.

Another feedback loop of considerable interest runs from the self-regulatory process to the evolution of technology. The ability to regulate access to, and conduct in, service markets confers a powerful influence over the forms of technology which can be deployed, and therefore over the types of research effort which are likely to be profitable. As we shall see later, this has, been a factor in inducing a bias toward cost -- and utilization -- expanding technologies, and away from cost reduction.

The effects of the various institutional responses to market failure in the health care sector have thus tended, singly and in combination, to promote increases in both utilization and costs per unit of service. Hence the (almost) universal concern in all developed countries with the overall costs of health care and their rate of expansion. Underlying that concern, however, is the concern that the level and mix of services being provided is inappropriate -- too much of the wrong things -- and that the process of provision is unnecessarily costly as a result of either or both of technically inefficient production or "too high" incomes of providers. In terms of Figure 1-3, we are both too far out on the curve, and below it, and Figure 5-1 displays some of the processes which have brought us there.¹⁰ This is not primarily a result of incompetence, stupidity, or venality; the present set of problems are a natural result of the social responses to earlier, and quite real, problems of market failure. But they are no less troubling for that.

CURRENT POSSIBILITIES FOR POLICY RESPONSE

An attempt to assess the prospects for alternative "solutions" or at least further responses, requires considerable *hubris* at any stage of discussion. It would also be quite inappropriate here because we have yet to explore the peculiar characteristics, objectives, and behaviour of the providers of health care, as well as of public programs for its regulation, subsidy, or provision. A synopsis is useful, however, because it is the institutional structure of Figure 5-1, responding to the peculiarities of health care as a commodity, which motivates the subsequent discussion. Otherwise health care providers would simply be the private for-profit firms of the micro-economic theory textbooks, or at least their real-world counterparts.

Among the policy themes or responses to the problems of cost escalation and utilization patterns which are displayed in the right-hand panel of Figure 5-1 is a set of proposals which would lead in just that direction. The current enthusiasm in the United States for competition, and for market or market-type institutions to control the production and pricing of health care, focusses attention on the roles of public and self-regulation and of insurance in promoting cost escalation and utilization. The competitive response would be to modify or dismantle the self-regulatory powers of present providers, making entry to the market easier for a variety of alternative providers. It would also remove public subsidies (through the tax system) to particular types of insurance coverage, and would promote, or at least not restrict, the development of combined insurance/service programs.¹¹

Much less complex in conception, and probably less far-reaching in effect, are the varieties of direct controls over unit costs, quantities of output, and their product, total cost. The Canadian provincial governments have a number of years of experience of direct fee bargaining with physicians, controlling prices but not (directly) quantities of services supplied. This appears to hold down overall costs, over a time horizon of several years at least, but tends to encourage increased servicing (see Table 7-5 below). American experience during the Economic Stabilization Program was similar, with an even more pronounced utilization response, possibly because measures of utilization were more open to manipulation. Experience with controls on unit costs in hospitals in some American states suggests even more difficult problems of quantity definition and control.

Direct rationing can take place by focussing on specific services -- who shall or shall not receive organ transplants or immunizations, *e.g.* -- or by limiting overall capacity. Canadian experience so far has been primarily with the latter; efforts have been made to slow the rate of growth of manpower (especially physicians), to lower acute bed-population ratios, and to limit the acquisition of new technology. The anticipated result is less rapid escalation of use rates, with the implicit assumption that providers will determine mix of use in an optimal or at least satisfactory manner. But little attention has been devoted to specific use patterns.

Paralleling capacity constraints, one can simply limit global budgets -- overall rates of reimbursement -- and let the institutions concerned allocate funds between incomes and outputs. Most effective and least demanding of information in the short run, such "freeze" policies raise questions of long-run viability.

Alternatively, governments could respond by direct takeover of some or all parts of the health services system, to run a public health care service. This would give direct control, at least in principle and subject to the complexities of management in this sector, over levels and patterns of output, and, subject to the problems of wage negotiation, over total outlays.

This last set of policy alternatives serves to highlight the underlying problem. A public health care service concentrates power and responsibility into identifiable hands, but does not tell those hands what to do. The market approach decentralizes decision-making, ostensibly to provide maximum scope for the expression of users' preferences. It will be noted that the letters indicating the intended point of application of different policy approaches in Figure 5-1 show the competitive market policies, A, influencing the farther left-hand side structural characteristics, rather than directly addressing outcomes. The assumption is that a "right" or "best" structure can be found which will automatically induce appropriate outcomes. But just as market failure undermines the appeal of competitive policies, and forces their (serious) advocates to seek other institutional mechanisms to shore up the market, so the imperfections of the political process generate concerns about a totally centralized scheme, both for users and for governments. It is no accident that Canadian provincial governments, while exercising great powers over hospitals, have so far refrained from direct takeover of their boards. The direct restriction alternatives provide greater scope for the more or less independent providers of care to work with the public sector -- or against it.

This discussion of the institutional evolution of health care has attempted to trace out the processes whereby cost escalation and utilization problems have emerged in Canada. Most other countries have reached the same set of problems and concerns, although some by rather different roads. Levels of expenditure seem to vary considerably across countries, depending on the road chosen. But in every country the conflict over share of national resources is being fought out between the health system and the rest of society, regardless of the level or share of resources currently devoted to health care. As Figure 5-1 suggests, there is no mechanism within the health care system itself to balance the pressures for expansion. "Equilibrium" or an "appropriate" funding level is viewed by that system as a share of national income which is expanding at each point in time -- always a bit more than we now have. The reasons for this perception will be developed in our discussion of the providers of health care, but its implication at this point is that a policy of "no policy," a public decision not to confront the health care system over funding, will merely lead to persistence of relative cost expansion, and postponement of confrontation to a later, more expensive stage.

FROM POSITIVE TO NORMATIVE, OR "IS" TO "OUGHT": LERNER'S RULE

But the discussion thus far, based on Figure 5-1, is incomplete. It has concentrated on the processes whereby the intrinsic characteristics of health care have led to institutional responses, all of which in turn tend to inflate both utilization and costs, and has tried to outline a range of different policy philosophies for responding to that expansion. But it has nowhere addressed explicitly the question of "How much, and of what services, is enough?" In fact, each policy alternative addresses that question implicitly. The public service approach presumably relies on epidemiological and technical data in a bureaucratic framework to decide what is to be done. The various forms of control over prices, capacity, or budgets leave a high degree of discretion over mix of services in the hands of providers and try to limit overall costs. The market-oriented policies try to find ways of introducing consumer choices into the decision-making process in contexts which are less vulnerable to informational asymmetry. The nature of the problem faced by all such approaches, however, can be represented compactly in a framework known as Lerner's Rule (Lerner 1944).

The problem of how much of its resources a society should allocate to particular activities can be represented in marginal terms as one of equating

$$MSV = MSC \quad (5-1)$$

the Marginal Social Value of the activity should equal its Marginal Social Cost. This is simply the balancing of benefits against opportunity costs as discussed in chapter 1, although with new labels. An expansion of health care activity, a marginal increment to care output, will, viewed by society as a whole, generate benefits labelled *MSV*, the value to that society of a marginal change in health care output. But the necessary resources, drawn from other lines of activity, have opportunity costs insofar as those other activities must be reduced in scale or scope. Other things are given up.

On the assumption that as the health care sector expands, it supplies the most valued outputs first, and then devotes additional resources to lower priority uses, the *MSV* of health care will fall as output expands (though changes in technology, needs, or preferences may cause it to shift up - or down -- at any given level of activity). This is just Figure 1-3, with the health status axis replaced by "value to society," and can be described as diminishing marginal utility of health care, for technical as well as preference reasons.¹² But by the same argument, a rational society (if there are such) would withdraw the resources required from their least-valued alternative uses. Thus (for a fixed overall resource endowment) an expansion of the health care system withdraws resources from progressively more valuable alternatives. The *MSC* rises, as health care output increases. But since *MSV* is falling, and *MSC* rising, there will be some point at which they are equal, and that represents the optimal size of the health care system.¹³

Formulated at this level of abstraction, the optimal allocation of resources to health care looks like a central planner's problem of daunting complexity -- but it is no different from the problem of resource allocation for goods in general. What Lerner's Rule displays is the way in which market mechanisms can be used to "solve" this problem, on a decentralized basis, but under very specific conditions. It thus enables us to pinpoint the various distortions created by market failure in the health care context, and to see their implications for optimal resource allocation in the abstract, as well as for the historical evolution of institutions and of cost and output patterns.

The market interposes between *MSV* and *MSC* a set of intermediate quantities or concepts, among which equality can be achieved by a chain of private decisions, like stepping stones.

$$MSV = MPV = P = MPC = MSC \quad (5-2)$$

The unit price of the commodity in question (or price vector, for commodity baskets) is *P*. *MPV* and *MPC* represent marginal private valuations and costs, respectively, as experienced by the buyer/user of the last (marginal) unit produced, and by its producer/seller. The equation between *MPV* and *P* will occur if informed consumers, knowing their own values, transact freely in smoothly functioning markets. It is represented by the point *P*^o, *Q*^o in Figure 2-2a, where purchasers of the commodity all value it at, or above, the price they must pay for it. For the buyer who values it least, *MPV* = *P*; those for whom *P* > *MPV* do not buy. Output and sales will be at a point where *P* = (society-wide) *MPV*. On the supply side, perfectly competitive profit-maximizing firms will choose to produce levels of output such that *P* = *MC*. If any firm can produce additional output at an incremental cost below the current selling price, it will, in a perfectly competitive market, choose to do so -- that will increase profits. So will cutting back

output when its marginal cost exceeds price. Equilibrium will occur when $P = MPC$, assuming perfectly price-competitive markets, including free entry and exit, served by for-profit firms. The price level itself, of course, responds (in a smoothly functioning market) such that the amount buyers want to purchase always equals the amount sellers want to supply, *at the going price*. Excess demand (supply) leads immediately to rising (falling) prices.

The linkages between MSV and MPV , and MSC and MPC , require both the satisfaction of specific technical relationships -- absence of external effects -- and the acceptance of the particular political or philosophical value judgements embodied in the consumer sovereignty postulate. At the technical level, it is necessary that no one person's consumption activity affects another's well-being, and that no firm's production affects either another's costs, or (except for its purchases/sales) consumers' well-being. All the costs and benefits associated with an activity (other than those reflected in changes in market prices) must be borne by the party or parties whose decisions govern that activity. But further, society must be willing to accept the principle that what people want (and are willing/able to pay for) is what they ought to get. In particular their "needs," as distinct from wants, have no normative significance for themselves or anyone else. The normative case for markets is ultimately rooted in this political judgement.¹⁴

In the health care system, almost every condition of Lerner's Rule is violated. An extended version of equation 5-2 makes this point clear.

$$MSV >? MPV^1 > \begin{cases} MPV^U \\ MPV^A \end{cases} = P^b < P^s > MPC < MSC \quad (5-3)$$

In this framework there are two prices, P^b paid by the buyer/user of services and P^s received by the seller, and three different sources of users' valuations. The uncertainty of incidence causes MPV to fluctuate at any given use level -- depending on whether or not one is ill -- which would in itself make it difficult (though not, in principle, impossible) to be sure that $MPV = P$. But insurance contracts, responding to that uncertainty, drive a wedge between prices paid and prices received. The difference $P^s - P^b$, which may equal P^s if care is "free," is made up by insurance.

But the MPV concept, in a world of asymmetric information, is difficult to formulate precisely. We have suggested three concepts: MPV^I , MPV^U , and MPV^A . The first is the valuation which would be placed on care by the hypothetical fully informed consumer of the textbooks, who would know the true structure of the relationship between health care and health. Less ambitiously, she might be assumed to share the best current knowledge about that structure. But she would not be restricted to the knowledge actually in the possession of her physician or other professional provider. On the other hand, MPV^U is the uninformed judgement of the patient confronting the health care system and initiating an episode of care. The intermediate case, MPV^A , is the valuation of care felt and expressed by the patient after being advised by her professional provider-agent.

None of these equates to MSV . The political/philosophical principle of consumer sovereignty presumably attaches to MPV^I , as one can hardly think of an uninformed or misinformed person making free choices. Indeed, deliberate misinformation is as much a form of coercion as physical intervention. But $MSV >? MPV^I$, because of the external effects involved; the ? is inserted to remind us that while true in general, this inequality does not apply to all forms of care, or independently of the state of health of the user.

On the other side, the patient/consumer will equate P^b to her view of MPV , subject to any direct restrictions on access. (One cannot, without professional approval, buy prescription drugs,

sign oneself into a hospital, or choose to receive treatment there). The shift from $MPV^U = P^b$ to $MPV^A = P^b$ occurs as the patient interacts with the provider and is informed or directed as to care use. This is the point at which the "demand curve" of the patient, as in Figure 2-2, is shifted right or left on the basis of professional advice, and control over utilization shifts from patient to

provider. In principle, $MPV^A > MPV^U$, provider advice can either raise or lower patients'

perceptions of the value of particular forms of care, as the provider will usually discourage or refuse to provide harmful or unnecessary care. But in general, $MPV^A > MPV^U$ as, for many forms of care, patients' preferences are totally undefined before they contact the care system. They are unaware of the existence of such services.

The problem of overuse then, in this framework, is that $MPV^J < MPV^A$. The agency role of the provider leads to care provision which the fully informed patient might not have chosen. In terms of costs, $MPV^J < MPV^A = P^b < P^s$ implies a dual discrepancy; the user of unnecessary care is led to overestimate its value and underestimate its price. On the other hand, the discrepancy between MPV^J and MPV^U , in detail, if not in total, is likely to be much greater. The justification for agency and professionalism, in the context of the consumer sovereignty value framework, is that the substitution of provider for patient judgement is presumed to lead to results closer to what the patient herself would have wanted, if fully informed, and that it is the latter preferences which society wishes to respect.

But the monopoly powers conferred by self-regulation, and the nonprofit environment, intended to support the agency relationship, also enable providers to hold selling prices above costs of production. This discrepancy shows up in both technically inefficient production, and above-market provider incomes. Thus $P^s > MPC$. Yet $MPC < MSC$ in general, because of the extensive system of public subsidies to private providers, as well as the possibility of negative externalities in production.

OVER TO THE SUPPLY SIDE

If the reader is left somewhat bewildered by the string of counteracting inequalities, that is as it should be. Reality is like that. While any one form of market failure (except for informational asymmetry) leads to a single inequality and a predictable bias in the $MSV = MSC$ relationship, the collection of all forms in equation 5-3 leaves us in total unable to predict, on a priori grounds alone, where we are relative to the optimum, above or below.¹⁵ And the various equilibrium mechanisms, which in theoretical market systems tend to lead us toward optimality, are clearly absent. Consequently, the normative significance of whatever levels and patterns of utilization and processes of production happen to emerge from these relationships is wholly undercut. Even accepting consumers' sovereignty as a value postulate, we are unable to infer "ought" from "is," or standards of optimal resource allocation from observed behaviour.

The pervasive "failure" of markets in health care delivery, even in situations where apparently voluntary exchange of money for services persists, leaves us with two classes of problems. The price system, in a theoretical perfect market system, performs both positive and normative functions. It determines, or guides, what will be produced, how, and for whom. And it generates, or at least elicits and records, the information from which (assuming the prior consumer sovereignty postulate) criteria of performance are developed. In health care, the price system is capable of playing neither role. Accordingly, in chapter 1, we introduced health status

as a proximate objective, an alternative source of criteria for good and bad performance. This, we suggested, might more closely approximate what fully informed patients, and others with an interest in their well-being, *would* choose as an objective. "Revealed preference" or willingness-to-use, in an environment of pervasive information failure and regulatory constraints on supply, will be an inadequate, and inaccurate, way of reflecting those hypothetical choices.

But the positive problem remains. If the market system does not govern patterns of production, delivery, and utilization in health care, or does so only to a limited extent, what fills the gap? The answer, of course, is that providers of health care have a degree of discretion, of direct influence over patterns of utilization and production of health care, which is highly unusual, if not unique, in a supposedly market-governed economy. If consumers' unaided preferences are not, and for good reason should not be, the primary data governing health care delivery, then the way is open for producers to make the key decisions -- and they do. To understand and to evaluate the quality of those decisions, and of their results, we must develop an explicit analysis of the behaviour of the different classes of health care providers, their objectives, constraints, and performance. In this performance, the automatic attainment of technical efficiency and opportunity-cost pricing can no more be taken for granted than can appropriate, efficacious, or optimal output patterns. No "market" enforces these. How providers choose what to do, and how to do it, occupies the central place in health care economics. And to this we now turn.

NOTES

¹ Such improvement is not "Paretian" improvement, in the sense of leaving everyone at least as well off as in the pure voluntary exchange case. It may be "potential Paretian" in the sense that where market failure exists, idealized regulation can lead to outcomes in which gainers could compensate losers (though they probably will not do so), but that criterion has well-known problems. "Improvement" probably really means according to some more general Social Welfare Function which the analyst believes she sees revealed by a society's actual behaviour, as well as in the public pronouncements of its opinion leaders or recorders.

² It is, of course, true that at a deeper level this is a pure statement of faith; the Free Will/Determinism question does not appear to be resolvable, at least by the intellect.

³ Frequently described as "technical advance," or "progress," but in the health care context a less judgemental label is called for.

⁴ This may even be entirely separated from efficacy, in the event that involved others wish, almost as a ritual, to ensure that "everything possible" is done -- regardless of its expected payoff, if any. This raises the interesting question of how the range of "everything possible" comes to be defined, and by whom.

⁵ Some argue that the income gains are all dissipated in "rent-seeking," in the long run at least. The case for full dissipation seems unsustainable, but in any case does not affect the discussion here.

⁶ Neither argument, of course, extends to for-profit firms in a *non-competitive* environment. Sir John Hicks' famous dictum, "The best of all monopoly profits is the quiet life" points out that for-profit firms sheltered from competitive pressures may also let costs rise for a number of reasons; and even if they do not, a for-profit monopolist will certainly seek to elevate *prices* above costs of production (including normal profit).

⁷ Whether up or down depends, as we shall note later, on how they are used.

⁸ As of 1983, there is some indication in the U.S. that this generalization may be breaking down and that private insurers *will* be able to influence medical practice. But it is too soon to tell, and in any case, leadership in this process has come from the U.S. public (Medicare and Medicaid) programs.

⁹ Sweden might seem an exception, with full public funding and very high cost. The difference appears to be that in Sweden funding sources are decentralized to the regional level, and are essentially "captured" by providers. Sole source funding must be at a level of government high enough to confront provider aspirations; the "political market" must be balanced (Marmor *et al.* 1976; Marmor and Bridges 1980).

¹⁰ If this view is not accepted, it is hard to see what meaning can be given to a "cost crisis." A health sector which was expanding its share of resources because it was producing a larger and larger quantity of highly valued outputs, in a highly efficient manner -- eliminating the common cold, *e.g.*, and developing ways of dissolving arterial plaques (safely) -- would hardly be "in crisis," any more than is the personal computer industry, whose costs/sales are expanding much faster than those in health care.

¹¹ The competitive approach in the U.S. is often, perhaps deliberately, confused with a pseudo-competitive alternative, which would neither modify the self-regulatory power of providers, nor promote alternatives to conventional "bill-paying" insurance, but would merely remove public subsidies to insurance and promote high-deductible, high-coinsurance rate coverage. Its proponents' objectives seem to be the removal of public oversight from a privately regulated and cartelized industry, so as to permit further cost escalation. The Canadian counterpart is a push for direct and extra billing by physicians, and "deterrent" charges in hospitals. But out-of-pocket payments by patients at point-of-service are neither necessary nor sufficient for a competitive approach. In this discussion and elsewhere, I assume that public health care policies are designed to address health care problems, not merely to limit government's liability and pass the problems back to patients. The latter interpretation is of course possible, but not very interesting.

¹² Those who are uncomfortable with an explicit representation in terms of utility are welcome to carry out the translation into the language of marginal rates of substitution. The argument goes through, but is, I believe, a good deal more opaque for non-economists.

¹³ Corner solutions, with health care output equal to zero or to the whole GNP, are theoretically possible but practically uninteresting. Furthermore, this relationship addresses only the issue of real resources to be allocated to health care, not their rate of reimbursement. In an idealized market system, the one follows from the other; but in the real world the relative income status of providers represents another major set of parameters in determining the cost of health care to the rest of society.

¹⁴ Which may have considerable appeal in our society, but is hardly "science"!

¹⁵ This is *not* a justification for assuming that the status quo is satisfactory! That would be a result of purest chance, and in any case we need not rely on a priori analysis.