

CHAPTER 2

RISK, UNCERTAINTY, AND THE LIMITS OF INSURABILITY

INSURANCE AND THE REIMBURSEMENT OF HEALTH CARE

Illness is unpredictable. Chronic conditions, or conditions which evolve through time, may have a more or less well-defined prognosis, and individuals of varying ages, circumstances, and pre-existing conditions may face very different probabilities of different states. But in general everyone's future health status is uncertain.

It follows that one's demand, or need, for health care is likewise uncertain; one cannot map out in advance an optimal pattern of health care use for the coming year as one might budget for food. And the costs associated with an unpredictable illness go well beyond the unexpected expenditures on care to which it may give rise -- being sick is uncomfortable, disabling, frightening, and sometimes fatal. Uncertainty with respect to health care expenditure is only a component, and not necessarily the most significant, of the overall uncertainty associated with illness.¹ As polar cases, both catching a cold and sudden death are unexpected events which reduce welfare, but neither generates a need for health care (though an uninformed consumer may in the first case demand it).

The institutional response to uncertainty is the development of insurance mechanisms, whereby covered individuals make regular payments to some risk-pooling agency in return for guarantees of some form of reimbursement in the event of illness. This agency might be a public body or a private firm, the payments might be premiums or taxes, and the benefits might be indemnities (fixed cash payments) varying across illness events, reimbursement of all or part of actual health care expenditures, or direct provision (public or private) of services as needed.

In practice, however, the evolution of health insurance in Canada has been marked not only by increasing numbers of people and types of care insured, but also by a shift from the private to the public sector. The share of health spending funded through the public sector has expanded steadily over time, with the principal jumps being the introduction of universal public hospital insurance during the late fifties, and medical insurance in the late sixties. (Dates of adoption varied by province.) These public plans built on a substantial base of private health insurance, provided both by non-profit and by for-profit insurers, which had been developing since the late 1930s (Canada, Department of National Health and Welfare 1954; Shillington 1972). In addition, some provinces had partial public service or payment programs with a long history.

Table 2-1 assembles data from various sources illustrating the development of public and private insurance coverage of health care expenditures. Hospital funding has always relied heavily on various public grants and funding programs: in 1932, when private hospital insurance was virtually non-existent, public funding already covered over half of hospital costs. Economic recovery led to increased reliance on self-paying patients; by 1940 these accounted for about 70 percent of revenues. But over the next twenty years hospital expenditures were rapidly taken over, first by private insurance, and then by the public sector.

TABLE 2 - 1
Measures of the Extent of Health Insurance Coverage, Canada, 1945-1981

	1945	1950	1955	1960	1965	1970	1975	1978	1981
% of Spending in Public Sector ¹									
-- All Health	n.a.	n.a.	n.a.	43	51	70.2	76.6	75.7	74.2
-- Hospitals	n.a.	n.a.	n.a.	72	89	93.9	94.6	92.7	90.6
-- Medical Care	n.a.	n.a.	n.a.	14	18	77.4	95.1	95.6	95.7
% of Population ² with Medical Care Insurance									
-- Private Comprehensive	n.a.	14.0	27.3	36.1	45.1	0.6	--	--	--
-- Private Limited	n.a.	4.8	9.6	13.8	15.5	--	--	--	--
-- Public	n.a.	n.a.	n.a.	7.9	12.1	96.3	100*	100*	100*
% of Hospital Expenses Paid by Third Parties ³	37	71	86	98	100*	100*	100*	100*	100*
% of Population With Some Medical Coverage ³	11	20	38	50	62	95	100*	100*	100*

SOURCES:

¹ Canada, Health and Welfare Canada (1979, n.d. [1984]).

² Canada, Department of National Health and Welfare (1954, 1963b); Fraser (1983); and unpublished data provided by Health and Welfare Canada.

³ Irazuzta (1979).

* The public programs nominally provide universal coverage, but a small proportion of the population may lose coverage through failure to pay premiums in provinces where these are still required. (*De facto*, coverage is rarely impaired.) Of the hospital expenditure not covered by the public programs, some is privately insured (preferred accommodation differentials covered by Extended Health Benefit plans), some is the responsibility of other agencies (Workers' Compensation, federal government), some is accounted for by non-residents of Canada. There remains some out-of-pocket charges, including the so-called "deterrents" in certain provinces, but these are not truly "private sector" expenditure, being rather an alternative form of public financing levy.

Medical care insurance developed later and grew more slowly; in the immediate post-war period such coverage was relatively rare. By 1960, over half the population had some form of medical insurance, but the degree of comprehensiveness was variable.

By 1975, as Table 2-1 shows, about 95 percent of all hospital and medical costs in Canada were funded through the universal public insurance programs. The public sector share of total health costs, through a combination of direct provision and partial insurance, had reached 75 percent. The remaining 25 percent was primarily out-of-pocket expenses for dentistry, drugs, and uninsured institutional services; private insurance (particularly for dental care) might amount to about 5 percent of total health expenditures.

Since 1975 the public share of hospital spending has declined somewhat, probably reflecting some combination of increased "deterrent" charges by provincial governments and growing preferred accommodation differentials. In addition there has been relatively rapid growth in dental and "homes for special care" (institutional care for the elderly) costs, for which public funding is relatively lower. Accordingly the public sector share of spending had slipped somewhat by 1981. But public programs for care for the elderly have grown significantly since 1975, and private insurance coverage of dental care has also expanded over this period. Total insurance coverage or public provision of health care is probably now close to 90 percent,

prescribed and non-prescribed drugs and dental care being the principal remaining exclusions. Unpublished Health and Welfare Canada reports based on Statistics Canada's Urban Family Expenditure Surveys showed out-of-pocket direct health expenditures per person of \$90.62 in 1978, 12.6 percent of the \$719.46 per capita total health spending.

Other countries have reached similar positions via different roads. In the U.K., public insurance of health care is combined with public provision through the National Health Service or other public authorities, which account for nearly 90 percent of all health spending, while a small private insurance sector covers well under 5 percent. European countries display a wide variety of institutional arrangements, but a similar predominance of payment through public and private insurance. Even the United States, which, in form and rhetoric at least, makes heavier reliance on private markets in health care than any other society, covered (in 1982) 88 percent of hospital costs, 63 percent of physician costs, and 72 percent of total health costs through insurance or direct public provision. Private insurance plays a much larger role, 34 percent of hospital and physician costs and 26 percent of the total, but out-of-pocket payment by patients for care purchased in "markets" is the exception, not the rule, even in the United States (Gibson *et al.* 1983).

The uncertainty of incidence of illness thus introduces two quite separate questions, How should insurance respond to illness, and Why *public* insurance? Opponents of public intervention in health care finance have pointed out, quite correctly, that the existence of (unwanted) uncertainty per se may be remedied by insurance, but that there are many forms of uncertainty other than illness, and the private insurance industry deals with them, apparently reasonably adequately.

We do not have public monopolies of fire, or life, or of the various forms of commercial insurance (although it should be noted that in Canada automobile insurance is in contention between the public and the private sectors). Thus to understand why Canada, and indeed every other society, has chosen to place the health insurance function in whole or in part in the public sector, we have to explore the reasons why, and to what extent, private markets might fail in the provision of health *insurance*, as well as in that of health *care*. Before doing so, however, it is necessary to analyse the relationship of insurance to the provision of health care.

MODELLING THE BENEFITS OF INSURANCE

For expositional purposes it is convenient, and common, to begin by assuming that any illness can be expressed in terms of a monetary equivalent, so that a dollar value can be assigned to the "loss" associated with a deterioration in health. This is not always true; the "health deterioration" of sudden death may not be compensable by any finite sum of money. However, it serves as a useful starting point.²

This monetary loss will not, of course, be equivalent to expenditure on health care. Only if care of a specific and well-defined amount were instantly and perfectly efficacious in relieving illness could one represent the consequences of illness for well-being by the dollar cost of care. In general, the money equivalent loss L of an illness will exceed any consequent (change in) health spending by some amount which allows for pain and suffering, anxiety, lost wages and/or leisure, and a risk premium for uncertainty of outcome.

We consider a consumer/patient contemplating an uncertain future in which a specific illness may or may not occur.³ Let the probability of occurrence be q , and of non-occurrence be $1 - q$, so that there are believed to be $100q$ chances out of 100 that the illness will occur. If it does, the consequences (expense plus misery) are judged by the consumer/patient as equivalent to a

loss of L dollars. Taking all other things into account, her well-being during the future period in question will be related (positively) to her wealth, which is expected in the absence of illness to be W .⁴ The occurrence of the illness is thus equivalent to a reduction in wealth to $W - L$.

If we express the dependence of the consumer's well-being or utility on wealth in the form $U(W)$, then we can describe well-being over the future period as $U(W)$ with a probability of $1 - q$, and the lower level of well-being $U(W - L)$ with probability q . If, *e.g.*, $q = 0.5$, there is a 50/50 chance of illness and, accordingly, a 50/50 chance of each of the levels of well-being associated with wealth W , and $W - L$. The expected level (in the sense of the mathematical expectation) is thus one-half, or 0.5, times $U(W)$ and 0.5 times $U(W - L)$. Representing this expected level as $E(U)$, we can write for the general case:

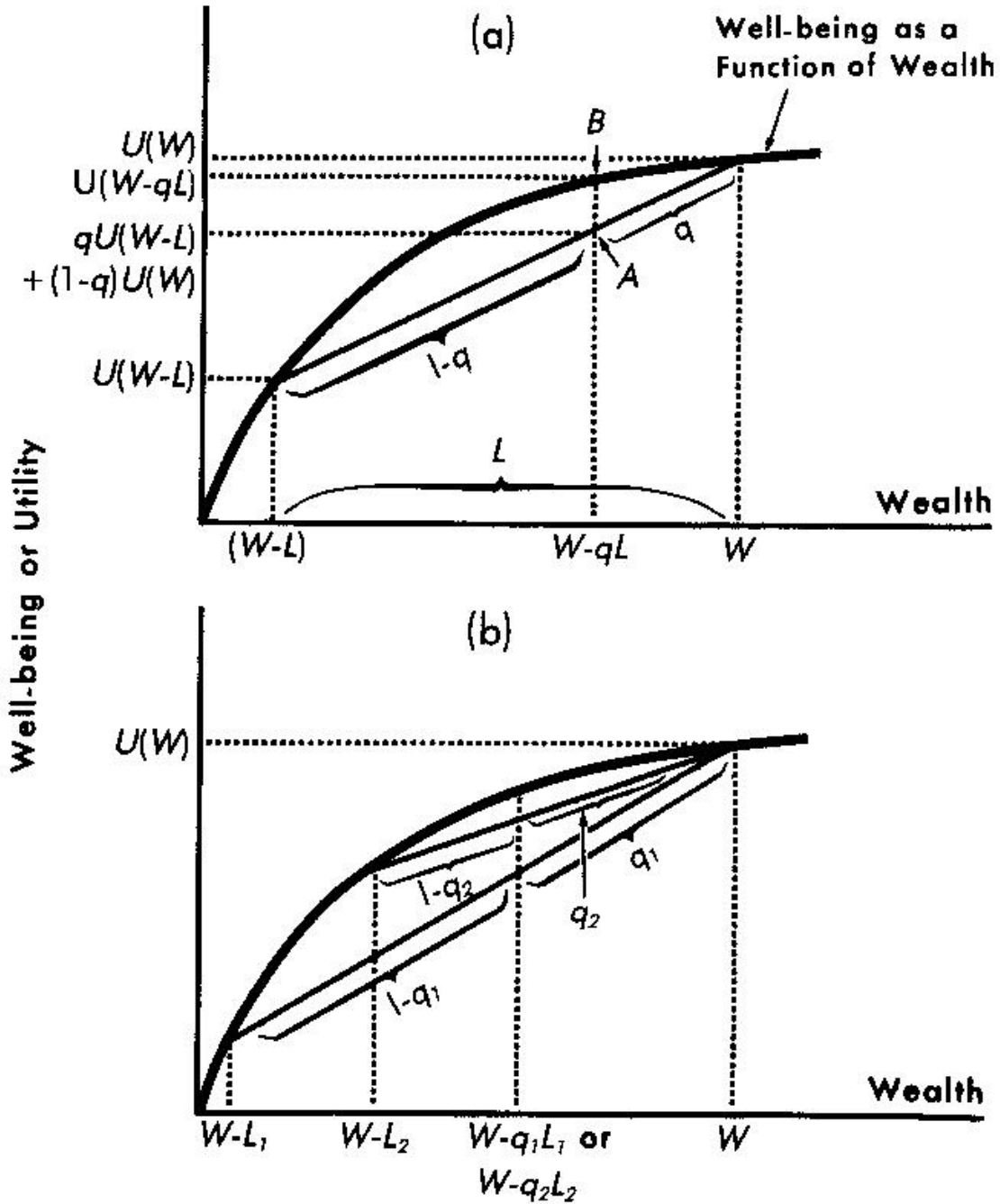
$$E(U) = (1 - q)U(W) + qU(W - L)$$

and since $U(W) > U(W - L)$, one is better off not to be ill, it follows that expected well-being is higher, for given W , as L or q are lower. The consumer is obviously better off if the future illness is less severe or less likely.

In Figure 2-1, panel (a), we represent an individual's well-being on the vertical axis, and wealth level on the horizontal. The relationship is drawn on the important assumption of diminishing marginal utility of wealth, *i.e.*, more wealth adds to well-being, but at a decreasing rate. The curve has positive, but declining slope. (People for whom this is not true, if they exist, do not buy insurance. They gamble a lot).⁵

FIGURE 2-1

The Welfare Gain From Insurance Coverage—Diagrammatic Exposition



An insurance contract offers a payment, say L , in the event that illness occurs, in return for a specified premium p which must be paid in either case. If the contract is actuarially fair, *i.e.*, the insurance agency makes no profit and is costless to run, then its payments over a large number of contracts will equal its receipts. The expected payment to any one insured will be the loss L

multiplied by its probability of occurrence, or qL , and this will equal the premium p . Thus, the insured individual will face a wealth level $W - p$ or $W - qL$, and associated well-being $U(W - qL)$, with certainty, rather than the uncertainty of $U(W)$ or $U(W - L)$. She is certain to be out of pocket the amount p , but certain to be reimbursed the loss L , if it occurs.

As can be seen from Figure 2-1 (a), however, while the expected *wealth* level is the same with or without insurance, $q(W - L) + (1 - q)W = W - qL$, the expected level of well-being is greater with insurance. $U(W - qL) > qU(W - L) + (1 - q)U(W)$, so long as $U(W)$ has the concave shape displayed, which for most people it appears to have. The gain in utility which results from complete insurance, coverage of the whole amount L , is the vertical distance from point A, the weighted average of the two possible uninsured utilities, to B, the certain utility when insured.

Panel (b) of Figure 2-1 represents two alternative losses, L_1 and L_2 , such that $L_1 > L_2$ but the associated $q_1 < q_2$ and $q_1L_1 = q_2L_2$. These represent a low-probability event with large loss, and a higher-probability event with smaller loss, such as a rare but severe illness with hospitalization and extensive care, and a more common minor illness involving one or two physician visits and a prescription. The expected loss in each case, size of loss multiplied by probability of occurrence, is the same. An actuarially fair insurance contract would charge the same amount to cover either loss, a premium equal to the expected loss qL in each case, and sellers of such contracts would then break even, on average, over a large number of contracts.

As Figure 2-1(b) illustrates, however, the benefit to the buyer of insurance, in terms of the gain in well-being from exchanging a risky situation for a riskless one, is substantially greater in the case of the large, low-probability loss than in that of the small, more probable one. And by extension of Figure 2-1(b), we can see that in the limit, if q goes to either zero or one, or if L goes to zero, the gain from insurance becomes zero, which is as it should be. If there is no risk (whether or not there is loss) there can be no gain from insurance. If the event is certain, a fair premium will equal the loss.

As Figure 2-1 demonstrates, at this level of abstraction there is benefit from insuring fully all risky events, however small the loss or close to zero or one the probability. Moreover, if insurance is to compensate for the loss associated with illness, it must pay amounts substantially larger than the cost of health care associated with illness. Since it is obvious that patients are not in general indifferent between a state of health, with no health care, and a state of illness with all necessary care fully reimbursed, it follows that even full coverage of health care expenditures would fall well short of compensating for the loss L .⁶

PRIVATE HEALTH INSURANCE IN PRACTICE: DEVELOPMENT AND LIMITATIONS

In the real world, however, full compensation of all risks is neither possible nor optimal. In the first place, the size of loss L can only be known by the sufferer; it is not directly observable by an external observer. Indeed, the illness itself is rarely monitored by the insurer; while possible in principle, it is usually costly and impractical. Expenditures on health care are taken as a signal that the illness has occurred. But insurance which compensated for L would then create incentives for false signalling. Patients would be encouraged to use unnecessary care, *i.e.*, care having no or even harmful effects on health, in order to collect the compensation for putative suffering which did not in fact occur. Accordingly, health insurance typically covers only actual health care expenditures, rather than losses due to illness,⁷ and in this sense all health insurance is partial and incomplete coverage.

This "moral hazard" problem, of changes in behaviour such that (real or apparent) losses rise in response to the existence of insurance, is a result of imperfect information about, or costly monitoring of, the true state of the world. An analogous problem would arise for the marketer of fire insurance if it were impossible to know how much damage a fire has done, or even if it has occurred. The insurer might then pay claims for construction services, relying on the contractor's statement that fire damage did in fact occur. The incentive problem is obvious.

Secondly, the actual process of providing insurance coverage is itself costly. Insurance agencies need staff, equipment, and buildings in order to set rates, write contracts, and adjudicate and pay claims. Competitive private companies must also allow for marketing expenses, and if for-profit, must, on average, earn a profit. Thus "actuarially fair" insurance contracts cannot exist. Premiums (or their equivalent in taxes or other payments) collected must exceed claims paid by some load factor large enough to meet the costs of the insurance function itself.

Consequently, insurance per se must add to the cost of the phenomenon being insured. The size of the load factor will vary, however, depending on how the insurance function is organized and marketed. It may run from under 5 percent of premiums for programs on the Canadian model, to 100 percent or more in the case of heavily marketed indemnity plans with very limited benefits. These mark-ups must be added to the actual cost of care for the insured group as a whole.

Later we shall see that the cost of health care for the insured group, both prices paid for care and amounts used, will itself vary, depending on the structure of the insurance system. The much higher cost of health care in the United States compared with Canada can be traced primarily to differences in their respective insurance systems. The overall relationship between health care costs and insurance coverage is rather complex. But an individual or small group contemplating buying health care insurance, and believing (rightly or wrongly) that the decision will not affect either prices of care or their own utilization rates, must trade off the proportion of the premium which is load factor, cost of insurance per se, against the gain in well-being from reduced uncertainty, as displayed in Figure 2-1. For small gains and/or large load factors, insurance is not worth buying.

In this trade-off, it will obviously be more advantageous to insure large and unlikely losses, because for a given premium, the gains in well-being are greater and (since few claims are processed) load factors will be smaller. Thus health insurance began historically with hospital care. Apart from public programs for the indigent or other special groups, private hospital insurance was introduced in Canada in the 1930s and spread rapidly during the 1940s and 1950s. A universal public program was first introduced in Saskatchewan in 1946, but most provinces waited until after the federal cost-sharing legislation of 1957. By 1961, every province provided universal public hospital insurance.

Medical care followed the same sequence, but somewhat later. Coverage for in-hospital surgical services, again the larger, less frequent claims, had the most rapid initial growth, but by 1960 medical and surgical coverage both reached about half the population. In 1962, Saskatchewan introduced universal public coverage of all medical expenses. By 1971, following the federal cost-sharing legislation of 1966, all provinces had universal public medical care insurance (Canada, Royal Commission on Health Services (Hall Commission) 1964, ch. 10; Le Clair, 1975).

It is important to note, however, that by the late 1950s private insurance companies were offering "major medical" policies whose market was growing very rapidly, from 228,700 persons covered in 1955, to 1,998,238 in 1961 (Canada, Department of National Health and Welfare 1963*b*). These plans reimbursed only expenditures per time period above a ceiling level or deductible. Small losses were thus not covered, unless associated in time with some larger loss.

On pure insurance principles, major medical or high deductible policies appear to be a superior way of purchasing protection from significant risks without the cost of processing large numbers of small claims. Moreover their relatively rapid growth in the 1950s suggested that they represented what consumers really wanted to buy.

Yet the Hall Commission, after reviewing the question of universal "first dollar" coverage of all medical claims versus "high deductible" coverage of expenses above some ceiling, and despite energetic advocacy of the latter by the private insurance industry and representatives of organized medicine, recommended full, first-dollar, universal coverage. Further, the Commission recommended public administration of insurance by a single agency in each province, rather than a division of the market between a number of private plans with public programs for those excluded from private coverage. There were good reasons for this decision, but they are not reasons based on insurance principles. If the only purpose to be served by a health insurance system, public or private, were optimal risk-bearing (and if health care expenditures were the only losses associated with illness), then the analysis thus far suggests that the major medical advocates were right. The continued advocacy of this form of health insurance by analysts in the United States reflects the same point.

But the argument for privately marketed, high deductible insurance is seriously incomplete. As pointed out in chapter 1, uncertainty of illness incidence is only one of a number of sources of market failure, intrinsic and institutionally induced, in health care services. The argument for high deductible coverage ignores these further complications and assumes, implicitly or explicitly, that health care is a "normal" commodity whose production and distribution not only can be, but *is* governed by conventional private markets. When proper account is taken of the other forms of market failure, the argument from insurance principles loses its force, as we shall see below, and the Hall Commission conclusion (and present Canadian policy) are supported (Evans 1983).

The evolution of dental and pharmaceutical insurance in Canada contrasts with hospital and medical, reflecting the point illustrated in Figure 2-1. Dental expenses tend to be predictable and controllable, more in the nature of maintenance costs than unexpected events. While they may be high in any one year, this is usually the result of accumulated need. Moreover, while often uncomfortably large, dental expenses are not ruinous in the way hospital and medical care can be. And very expensive dental work is generally to a significant degree cosmetic, not a response to a life-threatening or disabling condition. Expenditures for prescription drugs are likewise common and small, while in-hospital drug use is already covered by hospital insurance. Extensive use of prescription drugs on an ambulatory basis is generally confined to a well-identified group of the chronically ill. Such a group may deserve subsidy, but they are not "at risk" -- their condition is known.

Thus, the benefits of dental and drug insurance are relatively small, and private coverage has been relatively slow to develop (Evans and Williamson 1978). Traditionally, dentistry has been described as "uninsurable," reflecting its association with regular, small, predictable, and controllable "losses." Public programs have been established at the provincial level to cover particular groups, but these are more subsidy than risk-spreading programs.

The rapid spread of private dental insurance in the 1970s, on the other hand, would seem to belie its "uninsurability." But there are alternative explanations. One, heavily emphasized in the United States, is that the income tax system, which allows employer-paid health insurance premiums for employees as an expense deductible from the employer's taxable income, but does not tax them in the hands of the employee, encourages excessive insurance purchase (Mitchell and Vogel 1973). The argument applies to all health insurance in the United States; in Canada hospital and medical insurance premiums (where they exist) are, if employer-paid, taxable in the

hands of the employee. In November 1981, employer-paid private health insurance premiums in Canada were also made taxable to the employee, and insurers expressed concern for their markets, but this provision was reversed in October 1982. A critical "experiment" on tax effects was thus lost.

Alternatively, the dynamics of collective bargaining may be such as to encourage the spread of insurance, even though its benefits to the "representative employee" are minimal or non-existent. The bargaining process does not always model the behaviour of a fully informed rational fringe-benefit purchaser: it reflects political as well as economic motives. And individual buyers may be purchasing, through insurance, budgeting services and relief from uncomfortable consumption decisions. They do not want the anxiety of balancing prices and benefits at point of service (Evans and Williamson 1978; Conrad and Marmor 1980). Furthermore, there is evidence that even individual purchases of insurance in an open market are none too rational (Eisner and Strotz 1961).

SOURCES OF "FAILURE" IN PRIVATE INSURANCE MARKETS

The historical development of health insurance, with full coverage, first of hospital care, then of medical, and lagging coverage of dental care and pharmaceuticals is thus readily explicable in terms of differential trade-offs between benefits from pure risk-reduction, and load factor costs. What is *not* explained, as emphasized, is why Canadian coverage is universal, first-dollar, and public, rather than voluntary (universal in practice implies compulsory), high deductible, and private.

The sources of failure in private insurance markets, to which public insurance is a response, can be grouped under four heads:

- (i) economies of scale,
- (ii) insufficient information for rate-making,
- (iii) adverse selection,
- (iv) moral hazard.

Of these, the latter two have been the more intensively studied by economists, and seem to have the greatest theoretical interest.

(i) Economies of Scale

As noted above, the load factor component of premiums, the cost of operating the insurance program itself, varies greatly across different programs. It is very low in the Canadian public system, while in certain private for-profit plans the load factor may significantly exceed the actual claims paid. Overall, prepayment and administration costs represent 1.5 percent of health care costs in Canada, and 2.5 percent of the hospital and medical care costs which form most of the insurance load (Canada, Health and Welfare Canada 1979). Corresponding American proportions are 5.2 percent and 9.0 percent, but these include public Medicare and Medicaid programs, as well as self-payment. Private sector prepayment and administration costs are 13.4 percent of total private sector health care reimbursement (net of out-of-pocket and public sector payment), and 19.1 percent of private hospital and medical care reimbursement (Gibson *et al.*

1983). Load factors for private insurance plans in Ontario are in the 10 - 20 percent range (Ontario, Ministry of Treasury and Economics 1981).

These discrepancies are significant: an increase in Canadian load factors from 1.5 percent to 5 percent would raise health costs by over a billion dollars in 1982. Private sector load factors, based on American or Canadian experience, would imply extra costs of several billions. But the explanation of such discrepancies involves a number of factors: they cannot be interpreted as simply differences in efficiency, or profit versus non-profit.

One source of difference is that the insurance function is subject to inherent economies of scale. While some costs -- claims payment, contract administration -- depend on the volume of business done, others, such as information assembly and rate-setting, are the same regardless of the number of people insured. In general load factors will be lower for large insurers than for small. But this creates a classic public utility problem. If a large number of private insurers share a market, so as to maintain the discipline of competition to restrain costs and profits, the technical costs of doing business will be higher than necessary for each. But if a small number of private firms (or one) dominate (monopolize) the market, monopolistic exploitation is to be expected. In a market of 200 million people, like the United States, there may be room for numerous competitors of efficient scale, but in a Canadian province, the trade-off between efficiency and competition would be severe. A response to this, as to the similar problem of electricity generation, is public takeover.

Such a public utility approach creates its own problems of monitoring the public agency and holding it accountable to the public; these must be compared with those of scrutinizing private monopolies or oligopolies. But it has two additional advantages. First, multiplicity of carriers generates hidden diseconomies in the form of compliance costs for providers -- the administrative and clerical problems of dealing with dozens or hundreds of different insurers -- which add an indeterminate but allegedly significant amount to the costs of care itself. Part of reported hospital and medical care costs in the United States are indirect insurance costs; thus the United States-Canada discrepancy in insurance costs is even larger than reported. Second, a substantial part of private administrative costs are marketing expenses and agents' commissions, generated in the process of extending and competing for markets. These costs, too, are eliminated by the public universal system. "Premium collections" are piggybacked onto general tax collections; claims payments and data handling are centralized, and costs are much lower.⁸

Economies of scale in insurance provision are not only a source of additional costs in a competitive environment, they may lead also to "failure" of private insurance markets. A consumer might wish to buy insurance if it were available at a premium reflecting risk status (the actuarially fair premium) plus a load factor corresponding to technically efficient insurance administration. But if the only contracts available embody load factors inflated by the costs of inefficient small-scale operation and marketing expense, or supernormal profit from monopoly power, such a buyer might quite rationally choose not to buy private insurance, and yet to vote for a universal public program.

(ii) Insufficient Information for Rate-Making

The second source of private market failure, insufficient information for rate-making, arises in the case of health problems with a long time-horizon, such as occupational illnesses. In principle, it would be quite possible for employers or employees to purchase private insurance coverage which remained in force, at least for certain conditions, after the employee left or retired. Thus, the future health consequences of exposure to present hazardous conditions or substances would be insured by current premiums. In practice, the difficulty of determining risks

and assigning responsibilities for future illness states is such that these private contracts are not offered. Unless redress can be gained through the courts, a dubious and costly process, the worker or the state must bear these risks. If the risks of such future health costs are to be pooled at all, then, they are pooled through public programs. In a universal public program this happens automatically; in a private system it may occur through a public program for the elderly, or through some form of categorical assistance.

(iii) Adverse Selection

Market failure due to adverse selection is also a problem of imperfect information, arising from asymmetry of information between buyer and seller of insurance. The buyer of insurance may have better information about her risk status than the seller. In the extreme, if sellers were unable to discriminate among buyers at all, they might offer a single type of contract whose premium equalled the expected per capita loss averaged over the community as a whole, say, qL . But if the n potential buyers are distributed along a continuum of expected loss, such that:

$$q_1L_1 < q_2L_2 < \dots < q_nL_n$$

where L_1 is the loss contemplated by the i th potential buyer, and q_1 is the (buyer specific) probability of its occurrence, then all low-risk buyers, at the left end, are being overcharged, paying more than an actuarially fair premium (plus appropriate load factor). They are subsidizing high-risk buyers toward the right end, who pay less than their expected loss.

The practice of charging everyone in the community the same premium is known as "community rating," and was common among the early not-for-profit plans in Canada, as well as Blue Cross and Blue Shield in the United States. It has the effect of redistributing wealth from low- to high-risk individuals *ex ante*. An "ideal" insurance program, which charged each buyer a premium equal to her own risk status q_iL_i , would still redistribute *ex post* to those who actually experienced the losses L_i from those who did not -- all insurance redistributes after the fact. But "ideal" insurance does not redistribute *ex ante*; before the future is known it leaves each buyer's expected wealth unaffected (except for load factors) and merely relieves her of uncertainty. Community rating embodies an *ex ante* expected transfer of $q_iL_i - qL$ to each buyer, which will be positive or negative, depending on whether the individual's risk is above or below average. Whether or not such redistribution is socially desirable on other grounds, it is clearly unrelated to optimal risk-bearing.⁹

If buyers do not know their own risk status either, then community rating can persist. But if they do know it, either perfectly or to a sufficiently close approximation, then low-risk buyers at the left-hand end may conclude that the subsidy they are paying to high-risk buyers more than outweighs the benefits they gain from risk-pooling. If they do, and drop out of the market (self-insuring), then the insurer will find that average losses for the remaining insured group exceed average losses for the community as a whole. She faces adverse selection by buyers. Premiums will have to rise above qL . Depending on the pattern of the q_iL_i , and the shape of buyers' marginal utility of wealth curves, this can trigger further exits, until eventually the whole market disappears.

In practice, of course, adverse selection does not extinguish private insurance. It does, however, lead to the erosion of community rating, and to a level of coverage which falls well short of universality and comprehensiveness. Private insurance markets tend to evolve towards experience-rated group coverage, and to "major medical" -- high deductible -- and/or patient cost-sharing policies for individuals or groups.

The advantages of selling coverage to employee groups, conditional upon all or most of the group accepting coverage, are several. Administration costs are lower. The working population are generally healthier than the non-working. But most importantly, individuals cannot select into or out of coverage on the basis of their perceived risk status -- unless they are prepared to change jobs. So the insurer need not fear that she will be covering only the bad risks.

Furthermore, group experience can be monitored over time, and premium rates adjusted to that experience and to the group characteristics associated with it. But this in turn encourages competition by insurers to identify and "cream off" low-risk groups, offering lower premiums, with the result that coverage for high-risk groups becomes more expensive. Individuals not in employee groups, who are likely to be less healthy on average anyway, and of whom the less healthy are most likely to want to buy coverage, may be priced out of the market entirely.

Moreover, insurers will offer less complete coverage, in order to induce buyers to self-select into groups with differing risk status. Cost-sharing by patients, in the form of deductibles and/or coinsurance, are characteristic of the private, for-profit plans. Such features can be explained on the basis of pure insurance principles; the undesirability, in a world of positive load factors, of insuring small and/or frequent claims. But such contracts also serve to induce self-selection, as it can be shown theoretically that low-risk buyers will be more likely to purchase contracts with deductibles and coinsurance, whereas high-risk buyers will prefer more complete coverage. Accordingly, lower premia can be offered for less comprehensive contracts, not only because less is covered, but also because buyers of such contracts will, on average, be a lower-risk group.¹⁰

One cannot, therefore, infer preferences from purchases. A low-risk buyer or group might prefer full first-dollar coverage, *if* she could buy it at an actuarially fair premium (plus appropriate load). But if premiums for all such contracts are inflated to cover the claims of high-risk individuals, she may, as a second-best, choose a partial-coverage policy with out-of-pocket charges large enough to screen out high-risk buyers. Indeed, a low-risk buyer might be willing to buy full coverage at the community average rate, *i.e.*, one based on all low- as well as high-risk people. But under competition and adverse selection, that cannot be offered by private firms.

Competitive marketing of private insurance thus leads to a wide range of different rates, offered to groups on the basis of their own estimated expected loss. This rate differentiation is not in any sense a market "failure," rather it represents a reversal of the *ex ante* wealth redistribution from low to high risks which occurs under community rating. That reversal is costly in terms of information assembly, rate-making, and marketing. Whether or not one regards the resources as well spent depends on ones' distributional priorities, and is not a risk-pooling issue.

As identifiable good risks are creamed off, however, some of the remaining buyers will be priced out of the market. these, some will be simply unable to pay an actuarially fair premium because their expected risk is too large for their resources: these do not represent any failure of insurance markets per se. Their plight may be a social problem on other grounds, to be discussed in chapter 3 below, but their inability to buy insurance reflects an inability to afford the underlying (expected) care use and has nothing to do with uncertainty of incidence. They cannot afford Mercedes-Benz's either, but that is no failure of automobile markets.

The market does fail, however, with respect to lower risk individuals who *are* able and willing to pay a premium appropriate to their risk status, but who cannot communicate that status to an insurer. Lost amid the non-group pool, which includes the chronically ill and the elderly, they can buy coverage only at a premium price well above their own expected losses, if at all. (Private insurers may simply withdraw from the individual market, to avoid the administrative and political problems of the high-risk group as a whole¹¹). The position of such individuals is

analogous to that of low-risk buyers who choose partial over full coverage as a second-best policy which screens out high-risk buyers. In both cases, buyers and sellers of insurance would be willing to enter into mutually advantageous full-coverage contracts, but incomplete and asymmetric information, leading to adverse selection, makes it impossible to offer such contracts without drawing in high-risk buyers who cannot be identified in advance.

Such failure of private markets raises the possibility that compulsory public insurance could extend risk pooling in a welfare-improving way, if private individuals would prefer to buy comprehensive coverage but private markets cannot offer it at a fair price. And, in fact, one of the principal arguments used by advocates of a public program was the observation that the growth of private coverage in Canada seemed to have stopped well short of 100 percent of the population, and was in the late 1950s beginning to shift towards partial and away from full, first-dollar coverage. The Hall Commission concluded that the private voluntary insurance market was incapable of extending coverage to the whole population and, in particular, tended to exclude the poorest and least healthy groups in the population. On the assumption that universal and complete coverage was an appropriate objective, the Commission recommended a universal public plan. In the United States, the same observation about inherent limitations on the coverage of private insurance led to the creation of partial public insurance programs -- for the aged (Medicare), the poor (Medicaid), and for specific illnesses. While such partial or backstop programs supplement the private market significantly, they are apparently incapable of covering the whole population or of covering against all costs. About 10 percent of the United States population still has no health insurance coverage; out-of-pocket payments are about one-third of medical and 10 percent of hospital costs; and "catastrophic" expenses still strike some families who have limited or no coverage (Aday *et al.* 1980; Birnbaum *et al.* 1977). If health insurance for everyone is an appropriate social objective, then contrasting American and Canadian experiences suggest that partial public programs to supplement private markets are not an effective substitute for a universal public program.

Considerations of optimal risk pooling, however, do not necessarily imply universal first-dollar coverage as a social objective. Its advocates viewed such coverage as a way of promoting greater use of "needed" health care, and of redistributing wealth *ex ante* from those of lower to those of higher health risk status, a transfer which in general was from high to low income as well. A significant part of the population uninsured by the private sector were simply too poor to afford fair coverage. If their consequent inability to purchase health care was viewed as a social problem, and it was, that view had to be based on considerations other than incompleteness in risk-bearing markets.

Market failure due to adverse selection justifies public intervention only insofar as the uninsured or incompletely insured are willing and able to pay for additional actuarially fair coverage, but unable to buy it. We do not know how large a proportion of those excluded from or limited in private coverage in Canada prior to the public hospital and medical plans, or in the United States, then and now, fell into this category. But adverse selection considerations and current United States experience (Aday *et al.* 1980) suggest that such a group existed. And the converse implication of adverse selection is perhaps of greater importance. As noted above, the fact that people or groups buy limited (or no) private coverage does *not* indicate that they prefer this to full coverage at appropriate premium rates. Buying such partial private coverage but voting for full public coverage may be perfectly rational.

(iv) Moral Hazard

The last source of market failure in markets for private health insurance has received the most attention from economists. Moral hazard refers to a tendency for the existence of insurance coverage of any form to raise the expected losses insured against, as a result of either or both of greater loss or increased probability of occurrence. Since health insurance compensates, not for loss of health status, but for expenditure on health care services, the moral hazard problem in this context refers to a tendency for such expenditures to be larger *ceteris paribus* if a given individual or group is insured.¹² A setter of insurance contracts would thus have to set the premium above the expected loss qL for the uninsured individual or group by an amount large enough to cover this increase in loss (plus appropriate load factor), and the insurance buyer's wealth after premium payment, $W - p$, will be less than his expected wealth if uninsured, $W - qL$.

Moral hazard is quite distinct from adverse selection, although both may have similar implications for a private seller of insurance contracts who has a small share of the market. Moral hazard arises because the size of loss varies with coverage -- someone's behaviour changes -- and over the society as a whole total losses are increased by insurance coverage. It can occur in an insured group, all of the same risk status. Adverse selection, however, arises because people of different risk status are more or less likely to buy insurance. Over the society as a whole, adverse selection will lead to expected losses being redistributed by insurance, but total losses will not be increased.

Moral hazard may lead to market failure if potential insurance buyers predict that the rise in expected losses due to insurance coverage will more than offset the gains from insurance. If it were possible to purchase coverage *without* a change in losses, they would choose to do so, but such contracts are unavailable. One might hypothesize, for example, that dental insurance will lead dentists to raise their prices and/or patients to utilize more dental care, such that the necessary premium would substantially exceed one's current expected expenditures. Since the risk associated with dental care use is relatively small, moral hazard effects might make purchase of coverage on such terms economically irrational.¹³

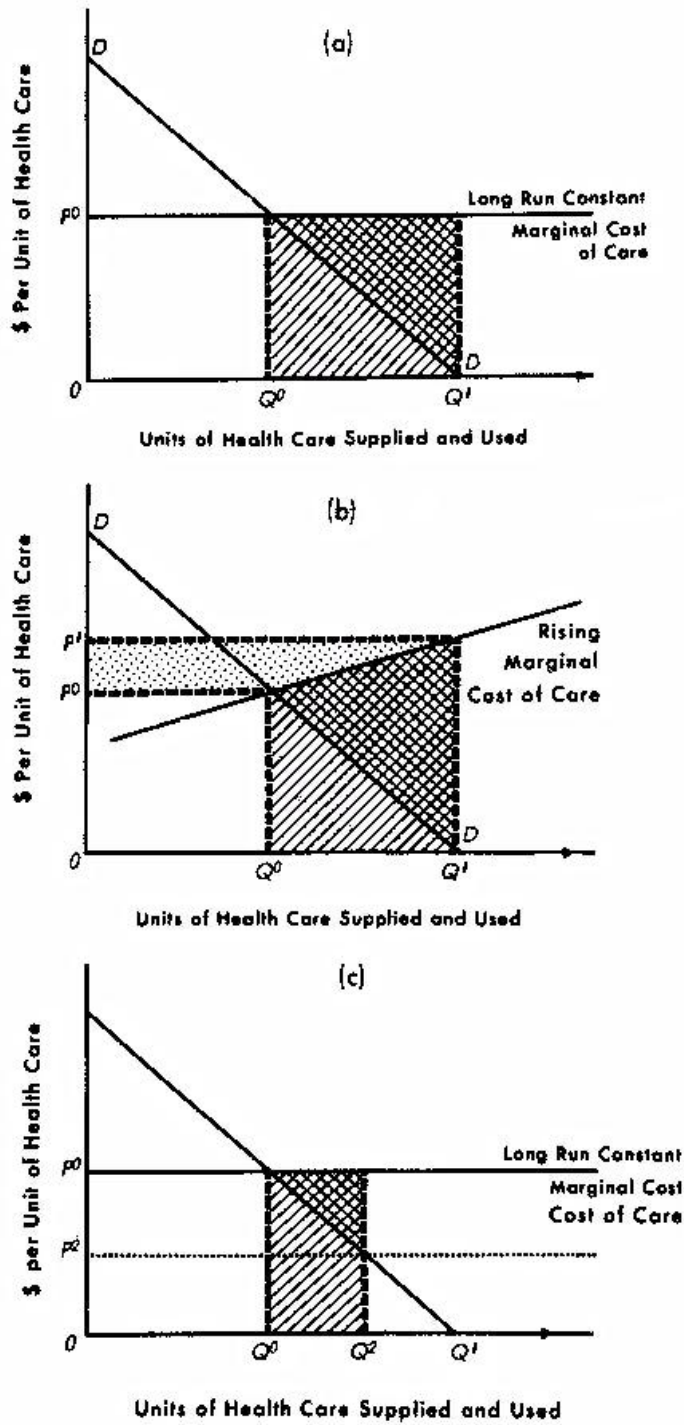
THE "WELFARE BURDEN" OF HEALTH INSURANCE

The existence of moral hazard, the appropriate response by the private insurance market, and the possibility of its amelioration by public intervention in private insurance markets, all depend critically on how the "market" for health care is believed to function. Many economic analyses of this question impose, either explicitly or implicitly, the assumption that health care is a "normal" commodity, supplied by competitive profit-maximizing firms and purchased by informed buyers whose demand is a well-defined function of price, and which just happens to be insured. If one thus assumes away the intrinsic peculiarities of health care as a commodity, as well as the extensive institutional structure, other than insurance, which modifies or supplants market institutions in organizing the production and distribution of health care, then "moral hazard" becomes identifiable with elasticity of demand. Insurance coverage which reimburses health care expenditures lowers the effective price of care to consumers, and they respond to this by increasing their consumption. Depending on conditions on the supply side of the health care "marketplace," this increase in demand may drive up prices as well, but the critical behavioural response to insurance is assumed to be that of buyers. Suppliers of care play a purely passive role, responding to the market.

The result is the so-called "welfare burden" of health insurance (Pauly 1969; Feldstein 1973), illustrated in Figure 2-2. In Panel (a) of Figure 2-2, health care is assumed to be produced by a perfectly competitive, profit-maximizing, constant cost industry, supplying care at a price equal to its (long-run) marginal cost. This price accurately reflects the resource cost of an additional unit of care, in terms of the foregone opportunities, the "other things" which could have been produced with those resources, and is assumed for simplicity to be the same at each health care output level. Consumer/patients are assumed to have a well-defined demand for health care, derived from some process of utility maximization under budget constraint, such that they choose how much care to purchase and use at each price level. The demand curve DD represents the amount Q they will choose to buy at any given price P . Being fully (or at least adequately) informed, they make this choice on the basis of their own preferences and income/wealth constraints, and relative prices, but *not* professional advice.¹⁴

FIGURE 2-2

The "Welfare Burden" and Wealth Transfer Effects of Health Insurance



In Figure 2-2a, P° and Q° represent the market equilibrium price and quantity of health care, pre-insurance. At this level of output, each unit of health care bought/utilized is valued by buyers as worth more than (or no less than) its resource cost. The vertical distance to the demand curve at any level of output Q can be interpreted as the value, in terms of willingness-to-pay, placed on the Q th unit of health care by the consumer who values it most highly. If one thinks of units of care as being produced and offered sequentially to users, the first unit(s) offered are highly valued, while successive additional units are assumed to be of less and less value. The last unit bought, at any given price, is valued at that price; additional units, being valued less, are not bought. And since buyers are informed, their judgements of the value of care are assumed to define the appropriate level of care provision (the consumer sovereignty assumption).¹⁵

If, on the other hand, an insurance program (public or private) were brought in which reimbursed all costs of care, then according to the demand curve in 2-2a, utilization would increase to Q' -- the point of utilization desired by consumers at zero prices.¹⁶ At Q' , the marginal or last unit of care bought/used is valued by consumers at very close to zero, but the resource cost, the foregone opportunities for other production, represented by that last unit, is still P° . Thus, resources worth P° are converted to health care worth (almost) nothing -- a clear misallocation and loss of welfare.

If we divide the distance between Q° and Q' into incremental units of care, the demand curve implies that some user values the $Q^\circ + 1$ unit at almost P° . But the $Q^\circ + 2$ unit is worth less, and so on, down to the Q' unit. The total value attached to the additional output after insurance is thus the striped triangle under the demand curve whose area is $\frac{1}{2}P^\circ(Q' - Q^\circ)$. But its resource cost is $P^\circ(Q' - Q^\circ)$. Thus, the total loss of welfare, the value of opportunities foregone less that of new care received, is the cross-hatched triangle with area also $\frac{1}{2}P^\circ(Q' - Q^\circ)$.¹⁷ Total health care expenditures rise from $P^\circ Q^\circ$ to $P^\circ Q'$.

In Figure 2-2b, the same analysis is carried out on the assumption of a rising supply curve. Higher prices are required to induce present providers to increase output and/or to draw new providers into the market. Insurance induces an increase in quantity demanded, which drives up price to P' . Total health expenditures increase by $P'Q' - P^\circ Q^\circ$, but now part of the increase is a wealth transfer to suppliers who were willing to supply their services/products at P° , or some price between P° and P' , and are in fact receiving P' . The increase in total expenditures divides into the value (to consumers) of increased care, $\frac{1}{2}P^\circ(Q' - Q^\circ)$ (striped), the welfare burden of increased care (resource cost less value to consumers) (cross-hatched) $\frac{1}{2}P'(Q' - Q^\circ)$, and the wealth transfer to suppliers, $\frac{1}{2}(Q^\circ + Q')(P' - P^\circ)$ (shaded).

The "welfare burden" of resources misallocated away from their most valuable use¹⁸ can be mitigated by limiting the extent of coverage. Figure 2-2c demonstrates the effect of an insurance program paying 50 percent of all health care costs, in the context of prices equal to constant unit costs of production. From the buyer's viewpoint, this lowers the price of care to P^2 , or $\frac{1}{2}P^\circ$. Quantity demanded, or utilization, will rise to Q^2 , where (if the demand curve is linear) $Q^2 = \frac{1}{2}(Q^\circ + Q')$. Now the value of new health care output to consumers is $\frac{1}{2}(P^\circ + P^2)(Q^2 - Q^\circ)$ (striped) and its resource cost is as before $P^\circ(Q^2 - Q^\circ)$, so the welfare burden is $\frac{1}{2}(P^\circ - P^2)(Q^2 - Q^\circ)$ (cross-hatched). But since $P^2 = \frac{1}{2}P^\circ$ and $Q^2 = \frac{1}{2}(Q^\circ + Q')$, it is easy to show algebraically that the welfare burden in Figure 2-2c is only $\frac{1}{8}P^\circ(Q' - Q^\circ)$ or one-quarter of that under full insurance.

INADEQUACIES OF THE "WELFARE BURDEN" ANALYSIS

It is even easier, however, to show that the welfare burden is minimized when there is no insurance at all. This result contrasts sharply with the earlier demonstration (Figure 2-1) that welfare was increased by the insurance of all risky events, no matter how small. To reconcile these, recall that the former analysis focussed on an uncertain event whose expected loss qL was independent of whether or not one was insured, and which occurred to a transactor with diminishing marginal utility of wealth. The "moral hazard" analysis of Figure 2-2 in fact embodies no uncertainty at all; the transactor is in full control of her consumption patterns, and risk, loss, or changing utility of wealth play no role. The "welfare burden" of insurance does, of course, vary with elasticity of demand; if demand is wholly unresponsive to price (if DD in Figure 2-2 were vertical) then there is no welfare burden from insurance. But there is no benefit either. In the absence of risk, it should not be surprising that insurance is not worth buying.

But both analyses also leave out of account, as emphasized above, any other peculiarities, intrinsic or derived, of the commodity "health care" itself. The pure insurance argument focusses on a "loss" which has a monetary equivalent, but leaves the relation between this loss and health care expenditures unspecified. The "moral hazard" as elasticity of demand approach sidesteps the whole issue of loss in a general sense, restricting the cost of illness to its subcomponent, the cost of health care. That in turn is treated as an ordinary commodity, supplied under perfectly competitive conditions to informed consumers, whose tastes for that commodity have, for some unexplained reason, a random component.

In this framework, which underlies a large part of the formal economic analysis of health insurance, the design of optimal insurance policies or programs is treated as a problem of trading off benefits from risk reduction, as in Figure 2-1, against costs from "excessive" utilization, as in Figure 2-2. Since utilization is assumed to be a result of decisions by informed consumers, the mechanisms usually suggested are various forms of direct charges to consumers -- deductibles, coinsurance, specific charges, which serve to lower the "welfare burden" as in Figure 2-2c. And since the only form of failure in insurance markets recognized by this analysis is that of "moral hazard" in this very narrow sense, there is no particular argument for public as against private insurance. Optimal coverage by either agency would require some form of cost-sharing.

The "excessive utilization" underlying the welfare burden argument bears a superficial resemblance to the "frivolous use" which is often alleged by physician advocates of direct charges to patients. But they are, in fact, quite different. "Frivolous use" refers to care which is unnecessary, in the sense that it does not contribute to health status; in the welfare burden case consumers are using care which they value at less than its true resource cost. Presumably, if they were fully informed, their values would reflect needs, but they may reflect other things as well, and particularly, incomes. Use of unnecessary care is medically frivolous, regardless of the income status of the user. But it is not "excessive utilization" in the sense of Figure 2-2, if it was used by someone who, for whatever reason, was willing to pay its true resource cost. On the other hand, effective care, which contributed significantly to the user's health status, would represent excessive use (in the sense of the "welfare burden" argument of Figure 2-2) if received by a poor person who could not buy it in the absence of insurance. It would be appropriate use if received by a wealthy person willing to pay its full price.

The distinction between health status and health care introduced in chapter 1 enables us to avoid these semantic confusions. If consumers value health care for its contribution to improved health status, then *informed* consumers will not buy ineffective care. If paying for care out of pocket, they will balance effectiveness against cost, and the medical concept of frivolous or unnecessary care will match¹⁹ the economic concept of excessive utilization. But the fact that

health status depends, in a technical way, on health care, simultaneously demolishes the basis for the assumption that consumers are, in fact, informed. If they are not, and their ignorance is remedied by provider-supplied information, then the demand curve *DD* in Figure 2-2 may not exist as a stable, negative relationship between prices paid by users and quantities consumed. And even if it does exist, it loses its normative significance as a guide to "appropriate" resource allocation.

Both the concept of moral hazard, and the analysis of health insurance generally, become richer and more complex when they are extended to take account of the external effects and asymmetry of information aspects of health care utilization, as well as the resulting extensive regulatory framework governing its production and distribution. The potential sources of "moral hazard" in response to insurance coverage include independent forms of supplier behaviour -- raising prices or changing recommendations about care use. The assumptions of perfectly informed consumers and perfectly competitive suppliers ruled out such responses except as a result of prior shifts in consumer behaviour; all effects of insurance were restricted to flow through the single channel of the consumer's utilization decision. In the real world, however, suppliers of health care are neither perfectly competitive nor profit-maximizers, and they exercise considerable independent power over both prices and quantities utilized in a "private" health care market. Consumers may control the initial decision to seek care for a particular problem, and in this decision their perceptions of, *inter alia*, costs to them of care may play the role envisioned in the demand curve of Figure 2-2. Thus, as Stoddart and Barer (1981) have shown, the demand for *episodes* of care may respond to prices in the conventional way. But the service content per episode is strongly influenced by providers, with greater control the more inherently costly (severe) the episode. And the most expensive forms of care are for chronic conditions in which the patient's entire remaining life becomes one episode. Possessing the (necessary and appropriate) power to influence patients' utilization patterns directly, providers can, and do, shift the demand curve, and vary utilization quite independently of prices paid by patients. As we shall see in chapter 4, this is the primary reason for the special regulatory framework surrounding health care supply.

Insofar as "moral hazard" arises from independent supplier behaviour, it is difficult or impossible to control directly in a framework of competitive, private insurance plans dealing with suppliers at arm's length. Attempts by American insurers, even those with apparently significant market power, to influence supplier behaviour have thus far been unsuccessful. Consumer/patient cost sharing, or incomplete insurance coverage, seems to have had some effect in the short run as an indirect mode of control over suppliers, though at significant cost in terms both of exposure of consumers to risk and of limiting the redistribution of cost burdens from ill to well. But it is not clear whether incomplete insurance serves to control overall levels of utilization and cost, or merely to redistribute care from more to less price-sensitive (usually poorer to richer) users. In any case, the longer term health care cost experience of the United States over the past thirty years suggests that present levels of insurance coverage in that country (coverage of about two-thirds of all health costs) are sufficient to permit "moral hazard" on the supply side to generate steadily escalating prices and levels of utilization.

By contrast, the public programs in Canada have, as shown in chapter 1, been quite successful in controlling utilization, and particularly price escalation, during the 1970s, without any significant exposure of consumers to cost-sharing. In effect, monopolizing the insurance function permits the control of moral hazard with comprehensive coverage; competitive insurance does not. It is sometimes claimed that much more extensive patient cost-sharing, *i.e.*, much less insurance, would eventually limit moral hazard in the United States setting. But the theoretical framework of such arguments is essentially that of Figure 2-2, so one cannot be too

optimistic. It is also suggested that integration of the insurance and care delivery functions, either insurers opening clinics or providers selling insurance in the form of capitation-based reimbursement, will limit "moral hazard" on the supply side. There is some evidence for this, but its generalizability is still questionable. We will return to these issues below.

In Canada the debates over public versus private health insurance were (and are) over much more than alternative mechanisms for pooling of risks. Universal public health insurance may in some respects be superior to private forms of risk-bearing, as discussed above, but it is even more important as a framework for policy intervention in the "market for" or resource allocation process of, health care itself. Such programs are a response to market failure in that market as well as, or far more than, in the market for risk-bearing. Accordingly, issues of risk-bearing per se provide only a very limited explanation for the public role in health insurance which we observe in every developed society (although limited is not non-existent). Analyses of health insurance which focus only on these issues, and which are frequently used to support private sector provision of insurance with significant patient cost-sharing, are seriously incomplete and, consequently, frequently misleading. If uncertainty were the only characteristic of health care which distinguished it from other commodities, there would be no justification for the extensive structure of licensure, regulation, subsidy, and public or non-profit provision which characterizes the supply of health care. But it is not, and there is. Whether the justification is adequate, or the structure appropriate, is a more difficult question.

NOTES

¹ The consumer's utility function may be extended thus: $U = U[X_i, HS(HC, E)]$ maximized subject to a budget constraint. Health status HS depends on health care HC and a random variable E indicating exogenous fluctuations in health (accident, infection, etc.). If E shifts so as to lower HS , U clearly falls, and HC may rise if it is "needed" (if $\partial HS / \partial HC$ increases), at the expense of other X_i . But the consumer's problem is the fluctuation in maximum U attainable with a given budget, due to the unpredictability of E , whether or not health care expenditure is increased.

² The derivation of this money equivalent requires the explicit or implicit recognition of health status in the utility function. If a consumer's utility is a function of commodities X_i , and health status HS , to which one commodity HC is an argument, then for a given exogenous vector of prices we can express (maximized) utility as a function of income or wealth and health status, $U = U(W, HS)$. If HS° represents perfect health, and HS' a specific illness, then we define L such that $U(W - L, HS^\circ) \approx U(W, HS')$, and L is the monetary equivalent loss corresponding to the health deterioration from H° to H' . The same point can be expressed equivalently in terms of (health) state dependent utility functions, $U(W - L) \approx V(W)$, where U is the utility from wealth when one is healthy, and V when one is ill with the specified condition.

³ The analysis generalizes easily to a range of illness possibilities, but becomes more involved.

⁴ Income is sometimes used here instead of wealth, but losses may exceed income in any one period. Moreover, the same income translates into very different levels of well-being for people at different wealth levels.

⁵ Shoenmaker (1982) surveys the theoretical and empirical underpinnings of this analysis.

⁶ Optimal insurance for an expected-utility-maximizing consumer might not, however, be fully compensating. In Figure 2-1 we assume not only that $U(W, HS') \approx U(W - L, HS^\circ)$ (see note 2, *supra*) but also that $\partial U / \partial W$ at (W, HS') is equal to $\partial U / \partial W$ at $(W - L, HS^\circ)$, which latter is (given diminishing marginal utility of wealth) greater than $\partial U / \partial W$ at (W, HS°) . This assumption, though common, is nonetheless somewhat suspect, and the expected-utility-maximizing consumer, trying to allocate her wealth so as to equalize across states the ratio of expected utility of

wealth to its "price", might choose to over- or under-compensate *L*. But this consideration raises issues well beyond the present discussion; Dionne (1982) and Evans (1983) provide more detail.

⁷ Disability insurance goes farther, to cover part of lost income, in situations where the loss is significant enough to justify the necessary monitoring costs.

⁸ There is a question as to whether the "scale economies" may not mask a loss to consumers from the reduced range of contracts available; the higher cost private system also provides a richer menu of choices of premiums and coverage. This issue cannot be dealt with adequately until after the discussion of adverse selection and moral hazard, below.

⁹ A tax-financed system will also transfer from high-income to low-income people within each risk category, as well as from low-risk to high-risk people within each income category. If well-being depends on income *and* risk status, such a two-indicator redistribution system may have much to commend it. But it cannot be justified on risk-pooling grounds *per se*; see chapter 3.

¹⁰ No equilibrium contract structure may exist in such markets, however; and if it does, it will not in general be Pareto-optimal. The theoretical analysis is developed from Akerlof (1970) by Rothschild and Stiglitz (1976) and Wilson (1977).

¹¹ "Of the eleven largest medical-care schemes under the sponsorship of medical groups, nine follow the same practice that is in effect in Windsor. Individual subscribers are not accepted (Associated Medical Services is one of the exceptions and that explains its higher overhead)." (Katz 1952).

¹² It was pointed out above that insurance compensates for expenditures on health care, not for all losses due to illness, because of the difficulty of monitoring such losses and the incentives for false signalling. Moral hazard in the context of health care use arises from the same problem of imperfect information about actual health status. If illness events and consequent health care "needs" were easily identified, moral hazard would not arise, because insurance payments could be based on the event, not the cost of care.

¹³ Though it might still occur, for several other reasons noted above.

¹⁴ Professional advice can be admitted if the professional is a perfect agent of the buyer (see chapter 4), but apart from the unrealism of this Hippocratic assumption, it completely undercuts the perfectly competitive profit-maximizing supply side, or indeed any other model of supply based on self-regarding transactors. The suppliers would have to be Lange-Lerner bureaucrats with rather complex instructions!

¹⁵ Of course, the distribution of wealth in the society will affect the shape of the demand curve, as well as who gets what. All positive welfare economics depends on a prior judgement about the social acceptability of the income/wealth distribution. Furthermore, the interpretation of the demand curve as representing the marginal utility of care at each consumption level requires certain very specific restrictions on the form of the consumer's utility function.

¹⁶ The non-satiation postulate of demand analysis would have utilization go to infinity at zero price, which, of course, it does not. One can remedy this by introducing time and access costs, which are not insured, but the resulting analysis gets progressively messier without providing further enlightenment. The more elaborate applications of neoclassical demand theory are, however, thus protected from refutation by reliance on unobserved quantities.

¹⁷ This analysis neglects the premium payment, which would lower wealth and, hence, might shift *DD* to the left. *Q'* would be reduced, as would the welfare burden. It also neglects a number of other complexities surrounding the application of consumer surplus, in the belief that these would add confusion, rather than enlightenment, without changing the basic story.

¹⁸ As judged by consumers, given a particular pattern of income/wealth distribution and of illness experience.

¹⁹ More or less -- there are some additional sources of slippage.