The basics

Health insurance

As with other markets, the existence of uncertainty and risk in the health care market has led to the development of insurance. Insurance allows people to reduce the impact that uncertainty has in their lives. People pay a fee to an insurer in exchange for the insurer's promise to pay them a certain amount of money if a specified event occurs. By purchasing medical insurance, people can lessen the financial cost to them of, say, suffering an unexpected heart attack or getting injured in an accident. In a setting where people attempt to maximize their happiness and are averse to taking risks, the purchase of an insurance policy improves their welfare, *i.e*. they are better off with the policy than they would be without it (Arrow 1963). As well, there is a social gain. Society as a whole benefits from the availability of insurance because risks are pooled or shared by many people, so that if a certain catastrophic event occurs, an individual is compensated for his or her loss out of the fees paid to the insurer by all of the people who insured themselves against this risk.[4] In addition to the benefits of pooling risk, the general interest in health insurance carries a moral aspect for many people. Individuals who do not possess adequate insurance may not be able to afford the medical care that is needed in an extreme situation and the end result may well be the difference between life and death (Zeckhauser 1969).

Insurance in general and health insurance in particular, however, can have distorting effects. One of these effects is moral hazard: insured patients demand more services than they would in the absence of insurance. By lowering the marginal cost of care to the individual, health insurance encourages the use of health care services (Pauly 1968). As well, individuals covered by insurance likely will consume more health care for an event than individuals who are not covered by medical insurance (Arrow 1963). For example, someone covered by insurance who suffers minor injuries in a car accident is more likely to engage in a series of tests and therapies regardless of medical need than someone without insurance.

Complete medical insurance, or free care, creates perverse incentives.[5] When individuals do not face any charges (*i.e.* a third party—the government or a private insurance company—covers their medical expenses), they have no incentive to restrain their use of medical services. This situation can produce excessive demand for care and waste resources, to the extent that the costs of producing these services exceed what individuals would be willing to pay for them. On the other hand, the absence of medical insurance may have the undesired effect of encouraging patients to delay seeking care, which may be more costly and harmful to their health than if they had received prompt treatment or medical advice. The goal is to seek a balance between the incentives to under-use and the incentives to over-use medical services.

The phenomenon of moral hazard is illustrated in figure 1. The segment *ab* represents the demand for medical care (*D*) and the supply of medical care or the cost of producing each additional unit of medical service (the marginal cost) is represented by the line "supply (*mc*)." Assuming that the market for medical care is perfectly competitive (no one provider is large enough to affect average prices) and that providers maximize their profits, individuals would choose to consume g^* units of medical care at price P_0 because, at this point, their marginal costs are equal to the quantity of medical care they wish to receive. At any point to the left of *d*, the cost of additional medical service is too great and individuals purchase less medical care than they would like. At any point to the right of *d*, individuals are willing to pay more for medical care. In economic terms, *d* is an equilibrium where supply equals demand; there is no waste of resources.

However, if individuals are fully insured so that the price or cost of their health care is zero, they will consume *b* units of medical care ($b > g^*$). They consume to the point where the benefit they receive from the last unit of medical care purchased is equal to the cost of consuming this last unit. Thus, with complete health care insurance coverage, where the direct price to the consumer is zero, resources are not efficiently allocated. Individuals are over-consuming and under-paying for medical care services.

This is not to say that the benefit from consuming the last unit of medical care is negative, just that its cost is

greater than its benefit. The welfare loss, then, is the area between points *bde*. If the demand for medical services were perfectly inelastic (*i.e*. a vertical demand curve)—if no matter what the price, people demanded exactly the same amount of medical services—there would be no welfare loss. The presence of elasticity in the demand curve implies that individuals will demand more medical services when the price to them is decreased, and vice versa. The more widespread the insurance coverage is, the more protected people are from the true costs of medical care, the greater the potential for over-consumption of medical services.

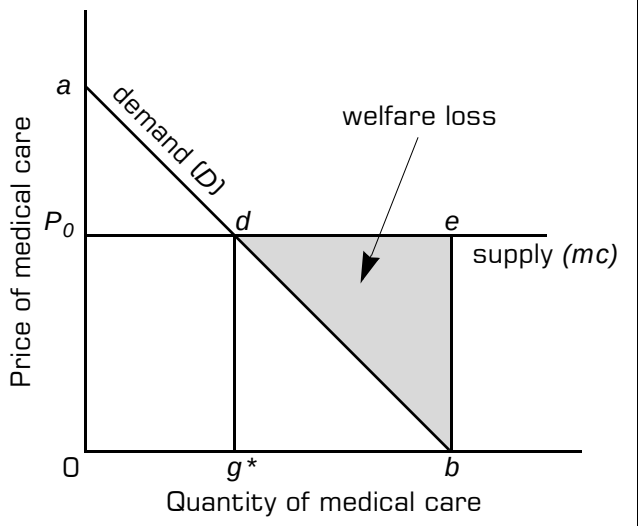
In the insurance literature, moral hazard is often seen as a moral or ethical problem. However, moral hazard is more a result of rational economic behaviour than of lower morality (Pauly 1968). Individuals may recognize that their excessive use of medical services will result in higher premiums or higher taxes but their increase in benefits from over-consumption is large while the incremental cost of their excessive use is small because it is paid by the entire insured population.

On the one hand, medical insurance increases social welfare because of risk-pooling while, on the other hand, it introduces perverse incentives towards excessive consumption of medical care. Richard Zeckhauser iterates this point well:

> In such a world [of uncertainty] you will be damned if you do not introduce a risk-spreading procedure, but you will be damned in another way if you do. (1969: 25)

It is because of these two conflicting aspects of insurance (risk pooling and moral hazard) that cost sharing has been introduced as a method of reducing the welfare loss due to moral hazard, while preserving most of the welfare gain from risk pooling.

Figure 1: Welfare loss of insurance



Note: For the long run marginal (*mc*) curve to be perfectly elastic, *i.e*. horizontal, one needs to assume that the insurance market is characterized by constant costs. However, this assumption can be relaxed. An increasing marginal cost does not invalidate the analysis; an increasing marginal cost curve would lead to an even larger welfare loss.

Traditional forms of cost sharing

Catastrophic insurance

Catastrophic insurance is simply an insurance policy that has a high deductible and covers all health care expenditures in excess of a pre-determined (high) level. An insurance policy that has, say, a $1,000 deductible per year, would cover all health care expenses in excess of a $1,000 per year. The insured individual or family, however, would be responsible for all of the costs incurred up to the $1,000 deductible amount. There are no deductibles in the current publicly funded Canadian health care system although some private insurance programs for services not covered by the public system make use of deductibles.

Deductibles affect individuals' behaviour in two ways. First, requiring people to pay for the first $*x* of their medical care, creates a financial incentive to restrain medical-care consumption. In other words, deductibles should diminish an individual's use of medical services because medical care is now more expensive as they have to pay for it out-of-pocket (price effect). Furthermore, as individuals spend more on medical services, they have less income to spend on other goods, including medical services (income effect). However, deductibles and catastrophic insurance also provide the conflicting incentive not to restrain health care consumption because expenses toward a deductible accumulate over time. If a person's first medical expense does not reach the deductible, his or her second medical expense need not be as large in order to reach the deductible. In other words, any medical expenses that do not reach the deductible have the effect of reducing the remaining deductible and thus diminishing the expected cost of medical consumption.

The rationale behind catastrophic insurance is threefold. (1) For individuals who are averse to risk, the greater the probability that they will incur a loss of some sort, the

greater the benefits will be to them of having insurance against this potential loss. (2) Insuring against the probability of an unlikely occurrence yields larger benefits than insuring against high-probability events (Evans 1984: 32–33). Less technically, relatively few individuals or families experience significantly large medical expenses in any year and, therefore, providing catastrophic insurance publicly need not be very costly. (3) Catastrophic insurance is an insurance policy with a high deductible and deductibles, high or low, exhibit three advantageous characteristics. First, since there seems to be little, if any, relationship between the size of an insurance claim and its administrative cost, higher deductibles can reduce the administrative costs of medical insurance. Second, the moral hazard of insurance may be less for larger claims than for smaller ones—*i.e*. as larger claims tend to involve more serious health conditions and treatments (such as a bypass operation after heart attack)—as there are fewer effective treatment options. There is, therefore, less incentive for patients to use more health care than is necessary (Keeler *et al*. 1977). Third, if individuals are averse to risk and want to maximize their overall happiness and wealth, they will choose an insurance policy that covers 100 percent of their medical expenses above a certain deductible if the insurer offers medical insurance at an actuarially fair premium (Arrow 1963).[6]

Martin Feldstein (1971a) was one of the first to propose a scheme of medical insurance that combined a high deductible with comprehensive insurance. Although he refers to it as major risk insurance (MRI), it is identical to the catastrophic insurance defined earlier. In Feldstein's proposal, the deductible increases with family income up to a maximum. Feldstein analyzes the potential benefits and costs of his proposal using six criteria: deprivation of care, financial hardship, cost inflation, tax burden, administrative simplicity, and general acceptability.[7]

Feldstein argued in the early 1970s that moving from the prevailing system in the United States to a system of high-deductible catastrophic insurance would reintroduce cost consciousness and provide individuals with incentives to use medical resources efficiently. This would occur because fewer individuals would be covered by extensive medical insurance. Thus, individuals would have a vested interest in shopping around for better prices and consuming medical care only to the point where incremental benefit equaled incremental cost. This, in turn, would apply downward pressure on medical-care inflation. The problem of moral hazard is reduced but not completely eliminated. Once the deductible has been used up, medical care becomes virtually free. At this point, incentives are once again distorted, though use may be restrained even if the deductible has been used up because the existence of a deductible reduces an individual's income (Pauly 1968).

Feldstein acknowledged that the introduction of catastrophic insurance, if provided by the government, would require that a fair amount of tax be collected. An increase in taxation is never without significant hidden cost but the welfare loss of increased taxes would be partially offset by the benefits of a less complex tax code (fewer income tax deductions) and smaller transfers towards the (American) national health programs of Medicare for seniors, and Medicaid for low income individuals and families. As well, with catastrophic insurance, only families which have exceeded the high deductible would make claims and, as a result, there would be fewer small claims and considerable administrative savings. It is also possible, however, that relating the size of the deductible to income would create an administrative nightmare (Newhouse 1993).

High deductibles can adversely affect the poor, who typically cannot afford much cost sharing. High deductibles may prevent or limit access to medical care and may, therefore, lead to the degradation of health for the poor. There is a valid argument for subsidizing the poor's consumption of medical care and it has been argued that more wealthy individuals may be willing to pay a portion, if not all, of the poor's medical expenses if wealthier individuals benefit from the subsidization for altruistic or paternalistic reasons or from reduced incidence of infectious disease.[8] It follows that the optimal deductible faced by the poor may be lower than that of other individuals, even zero.

As well, adjustments can be made to the size of the deductible faced by chronically ill individuals as the presence of a deductible potentially renders medical insurance pointless if one suffers from a chronic illness: the chronically ill end up paying the deductible every year for as long as they live (Arrow 1963). If a chronically ill person reaches the catastrophic insurance in some number of successive years, the deductible can be reduced or eliminated (Newhouse 1995).

Co-insurance

Co-insurance is commonly used, as are deductibles and user fees, to control the moral hazard of health insurance. Co-insurance requires individuals to pay some fraction of each dollar of cost (usually set as a percentage). For example, a health-insurance plan with 25 percent co-insurance rate requires individuals to pay for a quarter of all their medical care expenses. Similarly, an insurance plan with

zero percent co-insurance is equivalent to a free health care plan. With co-insurance, consumers of medical services pay a price for medical services but that price is lower than the market price.

Co-insurance can reduce the moral hazard associated with medical insurance though it does not completely eliminate the associated welfare loss. Individuals with a 25 percent co-insurance rate will increase their consumption of medical services until the last dollar spent on their medical care brings them a benefit of 25 cents. A 25 percent co-insurance rate decreases the welfare loss but since the cost of providing the last dollar of medical services is a dollar, the loss still amounts to 75 percent of the last dollar spent (Feldstein and Gruber 1994).

Figure 2 illustrates the impact of co-insurance on the size of the welfare loss. As seen earlier, with full insurance coverage, individuals will choose to consume *b* quantity of medical care with an associated welfare loss of the area within *bde*. If a positive co-insurance rate is introduced, individuals are faced with an effective price P_i, where P_i is less than P_0 but greater than zero. They will consume $g^{*\prime}$ units of medical care, which is more than they would if there were no insurance and individuals faced the higher price of P_0. The welfare loss with co-insurance is now the area enclosed by *dcf*. The welfare loss due to excessive insurance coverage has been reduced by the area *bfce*.

Mark V. Pauly (1968) argues that there is an optimal co-insurance rate for each individual, where the welfare gain from additional coverage equals the loss in welfare from higher premiums. Pauly, however, does not believe there is one single co-insurance rate that is optimal or "more efficient" for a whole population with diverse tastes. As well, the effectiveness of the co-insurance rate at reducing the consumption of medical services is directly related to the price elasticity of demand. The smaller the price elasticity of demand, the less responsive individuals are to changes in the price of a good or service. As a result, the effect of co-insurance will be smaller and, hence, welfare loss will also be smaller. It is also important to recognize that as the co-insurance rate increases, the amount of risk borne by individuals increases because the potential out-of-pocket costs increase with the co-insurance rate (Feldstein & Gruber 1994).

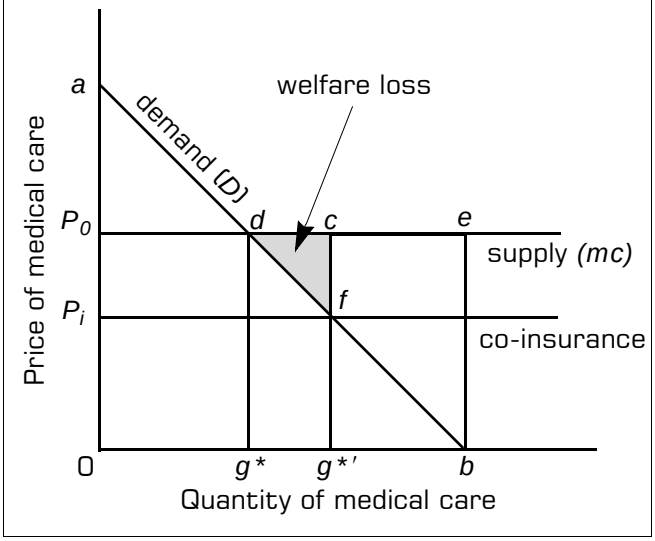
It is possible to combine deductibles with co-insurance. Feldstein (1971a) argues that high-deductible catastrophic insurance can be improved by introducing a co-insurance feature above the deductible. For example, you could have a basic deductible of 5 percent of income followed by a 50 percent co-insurance rate for an additional 10 percent of income. With such a scheme, it is possible to introduce co-insurance while keeping the maximum dollar expenditure of each family at the same level as it would be under the catastrophic insurance plan. Essentially, Feldstein's proposal attempts to find a compromise between the benefits of risk-spreading and the moral hazard of comprehensive medical insurance coverage. As with deductibles, the poor may not be able to afford a positive co-insurance rate but the co-insurance rate can be linked to income or the poor and the chronically ill can be made exempt from any cost sharing.

User fees

Opponents of cost sharing often use the literature on user fees to make a case against all forms of cost sharing (Evans 1993). User fees are a form of deductible applied to a given service—a payment, for example, of $5 for every visit to the emergency room or physician. The actual cost of the care given is greater than $5 but patients would pay the cost up to the $5 deductible while the additional care would be "free." In Canada, user fees are prohibited by the 1984 Canada Health Act.

Proponents of charging user fees for selected services argue that such fees increase efficiency in the health care sector and thus reduce costs: if required to bear a portion of their health care costs, individuals will curb their consumption of medical care and medical services of lesser value eventually will be eliminated. Thus, the welfare loss from insurance should shrink. As well, they contend that user fees can reduce the tax burden of Canadians because they redirect health care financing from taxpayers to users. Lastly,

Figure 2: Welfare loss of insurance with co-insurance



they believe that if the health care system is more efficient and more funding comes directly from users rather than from taxpayers in general, then governments will be able to increase the size of their health care budget.[9]

Opponents of user fees usually stress three drawbacks. First, user fees may increase administrative costs significantly because more human and capital resources will have to be devoted to their collection. Second, user fees erect a barrier to care that can have serious adverse health effects. Third, user fees may disproportionately shift the cost burden onto lower income individuals. This latter point will be examined in more detail in the following section.

Deficiencies with traditional forms of cost sharing

The main argument against the traditional forms of cost sharing is their distributional consequences.[10] Evans (1993) argues that the principal effect of introducing cost sharing in a tax-financed health care system like the Canadian system is cost shifting. If cost sharing reduces public expenditures on health care and the savings are used to reduce taxes, then it follows that taxpayers will pay less and users of health care will pay more in the form of deductibles, coinsurance, or user charges.

Evans believes that, as wealthier individuals tend to pay more taxes and less healthy individuals tend to consume more medical care, wealthier individuals pay a larger share of total health care costs in a publicly funded system, like that in Canada, with little or no cost sharing. In a system where cost sharing is more pervasive, like that in the United States, users of the health care system (*i.e*. the sick) tend to pay a larger share of the health care bill. It follows from this argument that the wealthy and healthy gain from cost sharing while the poor and the sick lose out. As well, since income and health tend to be closely related, this positive correlation reinforces the intensity of the cost shifting.

Evans contends that this pattern of income redistribution from the sick to the wealthy is true for all forms of cost sharing, even if some proposals exclude the very poor and the very sick. If cost sharing is linked with income, then the cost shifting is mitigated but does not disappear. If some segment of the population (such as those individuals below a certain income level) is exempted from cost sharing, cost shifting will still occur among the non-exempt population.

The argument that the wealthy and healthy benefit from cost sharing at the expense of the poor and sickly relies on the assumption that more cost sharing will result in lower taxes, which benefits the wealthy. It is not a certainty, however, that taxes will be reduced. Even if they were, the marginal tax rate of low income individuals could be reduced, and certain consumption taxes could be diminished. As well, any savings from greater efficiencies in the health care sector could be reinvested into the health care system or other social programs. Or, any savings could be left in the pockets of Canadians to spend as they wish on what they wish. Moreover, it is not clear that lower income and less healthy individuals lose more if cost sharing is introduced and taxes are reduced accordingly, since it is often the wealthier people in society who benefit more from social programs such as education and health care (Le Grand 1982; Horry and Walker 1994).

Evans' argument depends upon three assumptions: (1) wealthier individuals tend, on average, to be healthier than poorer individuals; (2) the sick use more health care, and (3) wealthier individuals pay more taxes than poorer individuals. With these assumptions, it seems reasonable that cost sharing would transfer income from the sick (and poor) to the healthy (and wealthy). However, this assertion may not be correct. The "sick" are not a homogenous group, and the income transfer is from the sick to the healthy (*i.e*. from the ill-poor and ill-wealthy to the healthy-poor and healthy-wealthy). Since the sick-poor outnumber the sick-wealthy, there could be a transfer from those with lower incomes to those with higher incomes. However, use of health care services tends to increase with income, not decrease. Individuals, sick or healthy, use more health care services as their income increases. This means that cost sharing does not necessarily mean cost shifting.

Consumer ignorance and supplier-induced demand

Opponents of cost sharing also point out that, because of ignorance, individuals may delay seeking care or forgo efficacious preventive care when faced with medical expenditures (*i.e*. when medical care is not free). This may subsequently result in higher medical expenditures if, for example, the illness has reached a more advanced stage (Roemer *et al*. 1975). As well, it is often argued that, due to the consumers' ignorance, physicians (suppliers) are able to induce demand (SID). For these reasons, they argue, publicly funded health care and government intervention in the market are not only justifiable but necessary. However, the hypothesis that suppliers of medical care control the demand for health care is a controversial topic in the literature about health economics.

The importance of supplier-induced demand (SID) stems from the fact that even if cost sharing reduces the demand for health care and decreases expenditures at the individual level, it may not result in an aggregate reduction in use and costs. When physicians and other health care professionals see their revenue dwindle because of the introduction of cost sharing (or MSAs), they will have an incentive to induce demand to restore their previous levels of income. In other words, the effects of cost sharing will be offset by SID. There is little doubt that the health care market is characterized by conditions conducive to SID. However, the relevant question is not whether the necessary conditions for SID are present but whether SID does occur in the health care market, and there is great uncertainty whether SID is a large problem in the health care sector.[11]

Those who believe that SID is problematic contend that the market for medical care is intrinsically different from other markets, and that medical services are unlike other commodities. To a certain extent, this is true. The medical market is different from the perfectly competitive model often used in economic theory, where every participant in the market is well-informed, there are many providers and consumers, and no one supplier can influence the average price of services. In the market for medical service, however, there is, first, a considerable amount of uncertainty about the effects of medical services upon health. Medical care is only one of several determinants of health and it is not at all clear how significant a role it plays (McKeown 1979). Second, there is the factor of consumer ignorance. It is often difficult for one who is sick to evaluate accurately the cause of the pain, or to determine how best to treat the pain, especially when information about treatment is scarce. Patients and physicians are not on equal footing when it comes to choosing an effective treatment.[12] Finally, risk is an important factor in the demand for medical care as most demand for health care is not predictable and the satisfaction one gets from medical care is not always easy to measure. As well, even if people are not completely satisfied with the care they receive, there are few, if any, relevant substitutes for medical care.

However, uncertainty and risk are not unique to the medical care market; it can be observed in many other areas of the economy such as automobile repairs and law. In addition, SID theorists generally assume that informed consumers (patients) would not be willing to pay for these supplier-induced services (Newhouse 1993), and that the induced services are less beneficial than the services that were foregone (because of the financial incentives of cost sharing to restrain total health care use). Neither assumption is necessarily valid (Newhouse 1993). Newhouse points out that if, for example, physicians induce demand by spending more time with their patients and billing for longer visits, it does not necessarily follow that patients will be worse off; they might well prefer it.

There are few economists and health-policy analysts willing to take the stand that the medical market is no different from other markets. Even the conservative British economic magazine, *The Economist*, acknowledges that the health care market is very complex:

> In some respect, admittedly, health is a more difficult and complicated issue than education. People have to rely on doctors to tell them what medical services they need, so it is harder for them to act as informed buyers. Health care is also an insurance product. (1997: 30)

Robert Evans (1984), among others, argues that the health care market is different from other markets because of the severity of market failures: uncertainty of incidence of illness, economies of scale, insufficient information for rate-making, adverse selection, and moral hazard. For the discussion of public policy, however, "market failure" is better used to describe instances in which the government can improve welfare in a way that the market cannot (Kennedy 1995): "the role for government intervention must be argued on the grounds of comparative institutional advantage over the market. The mere existence of externalities, market power and asymmetric information is not enough to justify government intervention."[13]

From traditional cost sharing to medical savings accounts

Traditional forms of cost sharing—deductibles, co-insurance and user fees—are used in the health-insurance market to alleviate moral hazard and thus diminish its welfare loss. As well, cost sharing is conducive to competition and should result in a more efficient health care system. However, cost sharing may entail regressive redistribution of income from the poor and sick to the wealthy and healthy, or it may impose a barrier to care that potentially endangers individuals' health status. Advocates of cost sharing have proposed ways to mitigate the effect of cost sharing on the poor but none seems to please its opponents. Advocates of medical savings accounts (MSAs) believe that MSAs can reduce the welfare loss of health insurance and induce competition in the medical market place without creating financial barriers to care.

Medical Savings Accounts

Medical savings accounts (MSAs) first emerged in the early 1990s during the national debate on health care in the United States. MSAs were initially proposed by the National Center for Policy Analysis (NCPA) and since have been embraced by many different organizations such as the Council for Affordable Health Insurance (Bunce 1996), the American Medical Association (1994), and the Family Research Council (Deeds 1995). As well, many research organizations such as the American Academy of Actuaries (1995), the National Bureau of Economic Research (Eichner, Wise and McClellan 1996), and the RAND Corporation (Keeler *et al*. 1996) have analyzed medical savings accounts as a measure for controlling health care expenditures.

What are MSAs?

Medical savings accounts (MSAs) are health accounts that are established in conjunction with catastrophic health insurance. They can be set up by individuals, by employers or by the government. The most common type is the American employer-funded MSA. In this type of plan, employers purchase a catastrophic insurance policy for their employees, which is much cheaper than a traditional insurance package. Employers then deposit a portion of the funds saved into MSAs for the employees, who are responsible for topping up their MSAs to the amount at which the catastrophic insurance begins, if they spend all the funds contributed by the employer. The MSA funds, including any funds remaining in the account after a specified period (usually a year), belong to the employee, although some restrictions may be placed on when and how the funds can be withdrawn and used.

Governments could easily take on the role played by these employers, providing individuals with catastrophic insurance and depositing funds into MSAs.[14] Depending upon the health status, age, and income level of those insured, the size of the government contribution could be all or a fraction of the MSA, that is, of the catastrophic insurance policy's deductible. Keeping government's health expenditures constant, all else being equal, the government contribution would be equal to the difference between the cost of an individual's or family's high-deductible catastrophic insurance policy and the cost of their current health plan.

The introduction of MSAs would encourage the consumer to enter the health care market as an active payer (American Academy of Actuaries 1995) and, it is hoped, provide sufficient incentives to motivate health care consumers to play a more prominent role in their consumption of medical care services and in their over-all health status. The promising characteristic of MSAs is that individuals purchase medical services with money they can otherwise keep because any funds remaining in the account at year end are the property of the individual. In effect, MSAs can indirectly establish a cost-sharing device without infringing on the most important philosophical cornerstones on which the Canadian health care system is built: universality, accessibility, portability and comprehensiveness.

Table 1 shows how a typical American employer-funded MSA can be seen as a combination of three different health insurance plans. First, individuals purchase medical services with funds made available by their employers (up to $857 in the individual coverage example). The first dollars of coverage are virtually free[15] since they are paid by the employers and not by the individuals. However, individuals have an incentive to use medical services prudently because every dollar not spent will eventually come back to them. Second, when the employer's contribution has been exhausted, individuals are responsible for the payment of medical care up to the cap where the catastrophic insurance kicks in (the next $643 in the example). This can be seen as equivalent to the deductible outlined earlier, with one important distinction; unlike regular deductibles, it comes into effect only after $857 has been spent (*i.e*. after the employer's contribution has been exhausted). Finally, once the insurance threshold ($1,500) has been reached, health care is free at the point of service for the individual.

Table 1: Example of American employer-funded medical savings accounts

	Individual coverage (insurance threshold = $1,500)	Family coverage (insurance threshold = $2,000)
Cost of catastrophic insurance	$ 877	$2,081
MSA contribution by employer	$ 857	$1,167
MSA contribution by individual (maximum)	$ 643	$ 833
Total cost	$2,377	$4,081

Source: Bond *et al*. 1996.

MSAs can be set up like a registered retirement savings plan (RRSP). They can be maintained administratively by the employer, a benefits administrator, a bank, or the government. During the first year, funds should be invested in low-risk, very liquid financial instruments so that funds that are meant to be available for medical care services are not quickly lost through risky speculation on the stock market. However, at the end of the year, individuals should be given more freedom to invest unspent funds in riskier instruments such as mutual funds, stocks, or bonds. This allows for capitalization in the health care market. Any funds invested today gain interest and can be used in the future for the more expensive care most individuals will need when they are older. This is contrary to the current system in which the tax dollars of today pay for people's health care today; because health care funds are spent immediately, there is no opportunity for them to be invested and to grow.

The unspent funds are one of the most important components of MSAs. By definition, the MSA funds that are not used during the year by the patient belong to the patient. Some of the options on how and when unused funds should be withdrawn are the following.

(1) Unused funds remain in the account until the age of retirement (and/or 65 years old); then all of the funds can be withdrawn without penalty and spent freely.

(2) Unused funds remain in the account until the age of retirement (and/or 65 years old); then all of the funds can be withdrawn without penalty and spent on medical care that is not covered by the catastrophic insurance plan.

(3) Unused funds remain in the account until the age of retirement (and/or 65 years old); then a portion of the unused funds can be withdrawn without penalty and spent on medical care that is not covered by the catastrophic insurance plan.

(4) Unused funds can be withdrawn at any time but are subject to a heavy (RRSP-style) penalty.

(5) Unused funds can be withdrawn at any time without any penalty.

These are only *examples* of possible scenarios. There are many other possible forms that an MSA system can take. Currently in the United States, individuals are free to use the funds that remain at the end of the year to purchase extra health care services (*i.e*. there are no extra taxes or financial penalties). However, funds that are withdrawn to purchase other goods and services are taxed and sometimes subject to financial penalties (Matthews 1997).

MSAs and incentives

There are three distinct sets of incentives at work in an MSA. First, the incentive to restrain use. The effectiveness of this incentive depends on individuals' perception of the employer's contribution. If individuals perceive the MSAs as insurance, then the incentive will be trivial; if the MSAs are perceived as potential savings, then there will be an incentive for individuals to constrain medical care consumption (AAA 1995). Second, for the individual's contribution, the incentive to restrain use is similar to that of a deductible except that it comes into effect only when the contribution from government or the employer has been exhausted. Third, once the insurance threshold, the deductible, has been reached, the incentives are akin to that of free care; that is, there will be little incentive to constrain use.

These three sets of incentives do not work independently. Rather, they work in sequence as medical expenses accumulate over the year. If a first medical expense does not consume all of the employer-provided funds, then a second medical expense need not be as large in order to reach the level at which the individual's contribution must be used. Any medical expenses that use but do not exhaust the employer's contribution have the effect of increasing the expected cost of medical consumption, because once the employer's contribution is spent, the individual's contribution begins. At the other end, any medical expenses that are paid from the individual's contribution but do not reach the deductible have the effect of reducing the expected cost of medical consumption for the individual as the amount that is needed to reach free care is less than before any expenses were incurred. It is also important to note that the sequence of employer's and individual's contributions can be inverted; that is, the individuals depicted in table 1 could be responsible for the first $643 of medical services (as they would be if paying a regular deductible), then the MSA funds ($857) would be used to purchase medical care.

Key issues

Savings or insurance? The success or failure of an MSA as a cost-sharing instrument depends heavily on the perception that each individual has of the funds (AAA 1995 and Keeler *et al*. 1996). If the MSA account is perceived as a contingency fund, then there is no financial barrier, no financial inducement, and thus no incentive to restrain use.

Individuals who adopt this view have no incentive to change their consumption of medical services. On the other hand, if the funds are perceived as potential savings, then the level of cost sharing is significant and is equivalent to a deductible. Individuals who perceive the MSAs funds as pure savings would have a stronger incentive to diminish their consumption of medical services. Whether an MSA balance is viewed as savings or insurance depends on several factors such as taxes, restrictions on the account balance, and the source of the contribution, and these factors will determine the effectiveness of MSAs at reducing the costs of health care.

Taxes Taxes distort incentives and often affect certain segments of the population disproportionately. An important question is whether contributions to MSAs by the individuals insured should be exempt from taxes (*i.e.* tax deductible). Exempting MSA contributions tends to favour individuals with higher incomes because wealthier individuals, on average, face a higher marginal tax rate (AAA 1995). A $500 contribution would only cost $333.33 in after-tax dollars to an individual who is facing a 33 percent marginal tax rate while it would cost $400 in after-tax dollars to a less wealthy individual who is in a 20 percent marginal income tax bracket. This is a standard problem in public finance. However, if MSA contributions are not exempt from taxes, there is little incentive to contribute to them.

MSA contributions can be compulsory. That is, individuals could be required by the government to deposit funds into an MSA. Certain provinces such as British Columbia and Alberta already use premiums to help finance their respective health care systems. Monies from premiums can easily be renamed and used as individuals' contributions. However, an alternative to compulsory contributions is the use of tax credits instead of tax deductible expenses. Tax credits can provide an incentive to contribute to MSAs without being regressive.

Another question is whether interest on positive MSA balances should be taxed. There are several possibilities: it can be taxed as ordinary income at the appropriate marginal tax rate; it can be taxed when withdrawn at the appropriate marginal tax rate; or taxes owed can be accumulated each year and be paid when the funds are withdrawn (AAA 1995). With any of these scenarios, the incentives associated with taxation of the interest on investment are simple: the smaller the effective tax rate, the greater the incentive to have funds in the MSAs at the end of the year, *i.e*. the larger the incentive to spend dollars on medical services wisely.

Individual's contribution: voluntary contributions or forced savings? Even if individuals have a strong tax incentive to contribute to their health account, everyone may not choose to do so. Individuals who decide not to contribute to their MSAs will essentially face a deductible with all its pros and cons which can create a barrier to care (albeit not a very large one). If individuals contribute fully to their MSAs, then there are no barriers to care, in the sense that there will never be a time when they do not have the necessary financial resources to purchase medical services. While voluntary contributions are preferable, mandatory contributions would achieve many of the same objectives.

Account balance The key issue here relates once again to how individuals will perceive the MSA funds: as savings or insurance. The perception will depend on the availability and accessibility of the unused funds. On the one hand, the more easily accessible the funds are, the closer they will be to savings. If the unused funds can be withdrawn without any penalty, then it follows that unused funds essentially become savings at year end; once the year has elapsed, the unused funds can be used by individuals as they wish. On the other hand, if the unused funds revert to the government once the year is over or once individuals reach 65, then individuals will most likely see these funds as nothing more than insurance and there will be an incentive to use all the funds available: "use it or lose it" (AAA 1995). The abundance of different arrangements possible is a strength of MSAs for it gives policy-makers more flexibility in achieving the desired level of cost sharing (Barchet 1995).

A major weakness of MSAs, however, is the uncertainty surrounding people's perceptions, and evaluating their perception of MSA funds is not a trivial exercise. In addition to the incentives already discussed in this section, unused funds may also create a wealth effect. Individuals who accumulate a positive balance may feel wealthier and this sense of wealth can, in turn, increase consumption of medical services, which are, after all, normal goods—that is, demand for them increases as an individual's income rises. As well, if unused funds can only be withdrawn at the age of retirement, individuals may simply reduce their contributions to a retirement plan. This redistribution of funds indirectly makes the unused funds totally accessible, which strengthens the incentives to shop wisely for medical services in order to buy even more health care or to increase savings.

In addition, the insurance threshold can be linked to the amount of the unused funds. That is, individuals who accumulate balances in one year would see their maximum

deductible increase in subsequent years. However, this mechanism diminishes the incentives to consume less health care use because it penalizes individuals who do so.

Setting a maximum permitted balance would also diminish the incentive to constrain use, and once the maximum balance has been reached, the incentives to restrain use are completely eliminated. Table 2 attempts to summarize some of the different key arrangements possible in an MSA scheme and their respective incentives to use the health care system prudently.

Other incentives

Asymmetric information A degree of consumer ignorance exists in health care and patients are, therefore, dependent on their physicians for guidance to a certain extent. In this context, MSAs may provide physicians with financial incentives to recommend unnecessary treatments to their patients. It has been argued that the amount of unnecessary services provided by physicians can be significant (Winslow *et al*. 1988; Roos and Sharp 1989). Evans (1987:173) goes even further and states: "Providers have ignored or resisted the external accountability implied by evaluation (based on evidence as opposed to professional opinion), or external comparison."

It has been argued by advocates of a "one-tier" health system that the Canadian system needs a third-party as "braking mechanism" to prevent an overuse of medical services. Neither the federal nor provincial governments have been able to ensure appropriate use of health and medical care services. Nor have they been able to ensure that patients who require care receive it in a timely fashion. Given their size and distance from the interaction between provider and patient, governments as "braking mechanisms" will never be able to intervene except with heavy-handed acts such as the closing of hospitals and a resistance to purchasing newer, more expensive medical technologies. While these measures may decrease use of the health care system, they do not encourage more appropriate use. One of the results expected from the use of MSAs is an incentive for patients to take a greater interest in their health and their consumption of medical services (Barchet 1996; McArthur *et al*. 1996). At the very least, MSAs likely will stimulate demand for information and this may well act as a "braking mechanism" (Barchet 1996).

Table 2: MSAs and incentives

Arrangement	Effect upon the incentive to reduce health care consumption
Threshold	
• A higher threshold strengthens the incentive to restrain use.	Positive
• Threshold as an increasing function of account balance diminishes the incentives to abate health care use as it penalizes individuals who consume less.	Negative
Balance	
• The longer one has to wait before withdrawing unspent funds, the weaker will be the incentive to restrain use.	Negative
• More restrictions on the use of unspent funds result in weaker incentives to restrain use.	Negative
• Setting a maximum permitted balance diminishes the incentive to abate use and, once the maximum is reached, eliminates it. A larger permitted balance provides stronger incentive.	Negative
Taxes	
• Taxing the contribution weakens the incentive to contribute to MSAs.	Negative
• Taxing interest income weakens the incentive to accumulate balances and, in turn, weakens the incentive to abate use.	Negative
Savings	
• If individuals contribute to MSAs as a substitute for other savings, then the incentive to restrain use will be strong.	Positive
• Compulsory savings weakens the incentive to reduce use.	Negative

Savings and intergenerational accountability In 1984, Singapore introduced a new health care system based on four basic principles: (1) free choice, (2) self-accountability and self-reliance, (3) free-market competition, and (4) restriction of the government's role to that of provider of last resort. With these tenets in mind, the government of Singapore implemented Medisave, Medishield, and Medifund accounts, schemes similar to medical savings accounts. Although some have questioned the success of MSAs in Singapore,[16] it cannot be disputed that the experience of the people of Singapore with MSAs has shown some of the broader potential benefits of adopting MSAs. First, Medisave plans are conducive to savings, which is favourable to economic growth and social welfare. Second, MSA-style plans increase intergenerational accountability; that is, individuals are required to accumulate funds in periods of low use (generally when young) in order to be able to purchase care in periods of high use (when older). Individuals face strong incentives to anticipate future risks and conserve monies. In Canada and the United States, the health care systems differ from that of Singapore in that for the most part they require younger generations to pay for the health care expenses of older generations. The MSA-type plans in Singapore have shown that MSAs are likely to encourage higher savings as well as to provide individuals with incentives to be intergenerationally accountable.

Individual responsibility and preventive medicine It is often argued that there is not a strong correlation between health care spending and health status (McKeown 1979; Evans 1982; McArthur *et al*. 1996). There is, however, one component of medical care that has shown promising results: preventive care. This is not to say that preventive care is the cure for all of the problems that ail the Canadian health care system and preventive care will most likely not reduce health care costs. It may even increase them because people will live longer to consume more health care services (Evans 1984).

With increasing pressures facing the Canadian health care system, legal, medical, and economic controls such as smoking bans and alcohol taxes have been offered as ways in which to "promote" healthier lifestyles, although these approaches are often worse than the vices of which they attempt to rid us (Morrein 1995). Irresponsible lifestyles do not justify government assault on individual freedoms unless a person's behaviour is harming another. If the societal objective is to promote healthier lifestyles, we must look more closely at medical savings accounts because, with MSAs, individuals have an incentive to give up a vice, maintain a healthy diet, use more preventive medicine, and act more responsibly with respect to their health care since, by living a healthier lifestyle and thus avoiding the medical care system, individuals will be able to pocket the savings.

Some critics have argued that MSAs may create a barrier to preventive medicine. They claim that individuals, because they can pocket any unused MSA funds at the end of the year, will forego preventive care. This, in turn, will have the effect of increasing health care costs because individuals will delay seeking care until they require very costly care. However, the RAND Health Insurance Experiment (see below, table 11) has demonstrated that, while cost sharing does reduce the use of medical services, including preventive care, it will not, for the most part, affect individuals' health. As well, if indeed the use of certain important preventive services, such as large scale immunization, is reduced by the introduction of MSAs, these services can always be provided free of charge to all by provincial governments or their health regions out of general revenues from taxation.

MSAs, the poor, and the chronically ill

Understandably, many people feel that health care should be available on the basis of need, and paid for on the basis of ability to pay (van Doorslaer *et a*l. 1993). J.P. Newhouse and his co-authors strongly reject cost sharing for the poor: "we start from the premise that substantial cost sharing for the poor is simply not an option" (Newhouse and The Insurance Experiment Group 1993: 352–53). It follows, therefore, that a special mechanism must be in place to deal with less wealthy individuals. MSAs can be adapted to limit the cost sharing faced by the poor while retaining the incentives to use health care efficiently. The maximum deductible can be set as a function of income, with lower deductibles for those people with lower incomes. Or the poor can be subsidized directly; that is, the government's contributions to less wealthy individuals can be increased. These two alternatives, however, come with some problems.

The latter option requires that the poor be identified in some manner, and this can be difficult to do and humiliating for the specified group. As well, a cut-off income level would have to be determined and this creates a strong incentive for individuals not to earn any income above the cutoff level or to hide any income above that level that they do earn. Nevertheless, MSAs allow policy makers to limit the degree of cost sharing imposed on the poor. Setting the maximum deductible as a function of income

may be a reasonable solution although this method may amplify the distorting effect of taxes.

Cost sharing with MSAs would also have to be revised for those who suffer from chronic illnesses. Imposing traditional cost sharing on the chronically-ill can be considered as a tax on the sick (Evans 1984). Those who are chronically ill have no other choice but to pay the cost-sharing portion of their medical bill year after year. However, with MSAs, cost sharing can be diminished or completely eliminated for those who suffer from specific chronic conditions or for those who repeatedly exceed the insurance threshold because of continued ill health.

Medical savings accounts: just another form of cost sharing?

Opponents and critics of MSAs, notably those from Canada, often argue that MSAs are nothing more than deductibles and co-insurance thinly disguised. However, there are important differences between MSAs and more traditional cost sharing devices such as deductibles, co-insurance, and user fees.

First, MSAs do not have a universal, co-insurance element. At most, only those individuals who can afford it are required to contribute to the costs of their health care, and even they will be partially subsidized. Health care is paid from general tax revenues for everyone once the insurance threshold has been reached and everyone will have at least some public funds that go towards the cost of the deductible placed in their MSAs.

Second, MSAs do not contain any user charges and individuals do not have to pay any extra charges when they use services. They pay the entire cost with funds from their MSAs or from the catastrophic portion of their insurance.

Third, the MSA maximum deductible will be equivalent to a traditional deductible only if the funds are fully perceived as savings by individuals. This can happen if individuals move money away from their retirement savings into their MSAs. How far MSAs are perceived as savings rather than insurance depends heavily on how they are designed. However, the most important difference between MSAs and deductibles is that MSAs can eliminate barriers to care. If all individuals contribute fully to their MSAs either because of strong tax incentives or because contributions are made compulsory, then there are no barriers to care—there will never be a time when individuals do not have the necessary financial resources to purchase medical services. Funds will be available in the MSAs, and it will be the decision of individuals if they use the funds to purchase health care services or save these funds for later consumption on non-medical goods and services.