

# Optimal Health Insurance

by Mark V. Pauly\*

## 1. Paper objective

My goal in this paper is to use neoclassical welfare economics to discuss the optimal form of insurance arrangements for medical services. My focus will be on insurance for services rather than on insurance to cover lost wages or other non-medical costs associated with illness, although I will consider these costs briefly at some points. Despite recent criticism of the use of neoclassical welfare economics by some (Evans, 1997; Rice, 1997), I will use it in this paper; my defense for doing so is that there is no obvious alternative normative criterion to use, and the arguments against this approach are substantially overstated (Pauly, 1997; Gaynor and Vogt, 1997). To avoid the most common objections, I will explicitly assume that other income has been redistributed to achieve what society regards as a fair distribution, and that there are no feelings of envy of one citizen toward another's use of medical care.

## 2. Optimal insurance in the simplest case

The simplest specification of an insurance contract specifies an event and an amount the insurance will pay if that event occurs. The event should be entirely outside the control of any party to the insurance transaction; it should be a random event. For instance, a person might buy a policy that will pay \$100,000 on the occasion of one's own death, as long as the death is not a suicide. The key elements in this insurance arrangement are the specification of the event and the amount to be paid. As long as the event can be accurately and inexpensively defined, a policy offered by a competitive insurer at a given premium will be optimal if it specifies an event–benefit pair that maximizes the individual's utility. Even though it is very simple, this case has two characteristics which are relevant to optimal health insurance: the event (or, more broadly defined, the “state of nature”) has to be defined, and the amount to be paid has to fit the individual's specific desires or demands. The difficulty of specifying the state of nature for health insurance and the need to link benefits to individuals' subjective (and varying) demands will be themes I will emphasize further in what follows. In particular, the need to think of optimal arrangements that take account of differences across citizens in the values they place on risk protection will be an important influence on judging the efficiency of private markets relative to collective choice.

---

\*The Health Care Systems Department, The Wharton School, University of Pennsylvania

Please send all comments and correspondence to: Mark V. Pauly, Ph.D. 208 Colonial Penn Center 3641 Locust Walk Philadelphia, PA 19104-6218 Phone: (215) 898-2837 Fax: (215) 573-7025 e-mail: pauly@wharton.upenn.edu

### 3. Compact descriptions of multiple states

Many insurances link the event that triggers benefit payments to a monetary loss. If people are risk averse (display diminishing marginal utility of income), they will prefer insurance at actuarially fair premiums to risking a loss. In these cases, the most efficient way to write a contract often is to link the benefit payment to an estimate of the amount of the loss in wealth. It will ordinarily not be efficient to pay benefits in excess of the loss (and may, as will be discussed below, cause additional moral hazard), but it is clearly possible to make the benefit payment less than the loss in wealth. If insurance is offered at actuarially fair premiums, the optimal benefit for each possible loss level will equal the loss in wealth: the insurance will be “full coverage first dollar”. If the premium is in excess of, but proportional to, the actuarially fair premium, the optimal insurance provides for full coverage above a deductible, with the size of the deductible growing as the deviation of premiums from fair premiums (the “administrative loading”) increases.

The intuition behind this result is reasonably simple. Relative to paying a fair premium, a risk averse person suffers relatively little utility reduction from risking rather than insuring a small loss, since the marginal value of money in the loss state will still be close to its value in the no-loss state. In contrast, this utility loss (or “risk premium”, if viewed as an amount the person would be willing to pay for coverage in excess of the fair premium) is disproportionately larger for a large loss. A deductible is equivalent to being without insurance coverage for small losses, those cases where the administrative cost is larger than the risk premium. In effect, where losses do not do that much harm, it is more efficient to avoid paying the insurer for administrative expenses to cover it.

In summary, the benchmark for optimal insurance against is full coverage above a deductible for easily measured losses occurring in response to exogenous and random events. The possibility that events might be exogenous but non-random (i.e. correlated) is negligible for health insurance in developed countries. However, given the assumptions made so far, and assuming no regulatory or artificial barriers to entry into health insurance markets, one can come to the informal conclusion that a competitive equilibrium is possible and, if buyers are fully informed, will be approximately efficient. The approximation arises because of the theoretical qualification, generally unimportant in practice, that complete and perfect efficiency requires pooling of all risks in a single pool. If this is not to involve a single insurance firm (and therefore rule out competitive markets), it requires perfectly functioning contingent claims markets (Arrow and Debreu, 1996), which is a strong assumption. However, as long as there can be many large insurance firms – and if reinsurance is possible – the extent of pooling in large economies can be great enough to drive any remaining efficiency gains from additional pooling close to zero (Pauly and Kihlstrom, 1971). In this sense, a competitive insurance market is “practically” efficient.

### 4. What makes medical services insurance different?

There are some important differences between the events covered by medical services insurance and the stylized insurance model already discussed which affect both the optimal design for such insurance and our conjectures about the optimal institutional arrangements in which to provide it.

There are four deviations from the stylized model that I will discuss.

### *Externalities/public concern*

In the stylized model, any changes in spending that might follow from the presence of insurance are assumed to affect only the utility of the insurance purchasing unit. In contrast, medical services use may affect the wellbeing of others, not only when the services treat contagious disease (the standard public health case), but also when the treatment generates no objective spillovers and yet others in the community care about the relievable but unrelieved suffering of others. This “altruistic externality”, which I have discussed in depth some time ago (Pauly, 1971), provides what in my view is the best positive and normative explanation for collective intervention in medical insurance markets.

### *Moral hazard*

The rate of use of medical services is to some extent under the control of agents (consumers, healthcare providers). Even if the occurrence of illness is exogenous, occurrence of the event “medical care spending” is not. Much of the optimal design of health insurance can be viewed as an attempt to control moral hazard, either by affecting consumers (demand side) or providers (supply side).

### *Information/action asymmetry*

Not all agents are well informed about the occurrence of illness and the properties of medical services; more importantly, some have different information from others. Sometimes consumers will be less well informed than suppliers of care or insurance; this can obviously lead to less desirable outcomes than if they were well informed. Second, sometimes consumers are better informed about their state of health than providers or insurers; this can lead to adverse selection in insurance. Importantly, even when insurers and patients have the same information, there may be differences in their ability to act. For example, if insurers are forbidden to adjust premiums based on known risks or to refuse applicants, buyers can decide whether or not to buy but sellers cannot refuse to sell. This can create adverse selection that could have been avoided, and it also creates incentives for insurers to “cream skim” as they try to avoid the limits on their behavior.

### *Non-competitive markets*

In the stylized case, as already noted, one assumes that markets for commodities are competitive, even markets for commodities whose use might be affected by the occurrence of losses or the presence of insurance. For example, the event of fires changes the demand for construction materials, but any changes associated with losses/insurance are such a small part of the overall market for such services that it is reasonable to assume that markets are and may remain competitive and the market price unaffected. In contrast, medical services may not be supplied competitively and are primarily used in connection with illness (medical losses) and (because of moral hazard) may have their demands affected by the extent and form of insurance.

In what follows I will initially ignore the question of altruistic externalities and government involvement, and discuss optimal insurance when insurance and medical services are considered to be private goods. Later I will show what modifications might be made when spillover effects are taken into account.

## 5. Optimal insurance in the presence of moral hazard

Moral hazard arises when people behave differently with insurance than without insurance in ways that do not depend solely on the shift of income from one state of the world to another. In the case of medical services, the primary type of moral hazard known to occur to a significant extent is the association of higher levels of expenditure on medical services with higher levels of insurance coverage. To give a classic example, the Rand health insurance experiment found that total spending increased by nearly 50 per cent when insurance coverage was complete, compared to the situation when the insurance had a large deductible. Economists explain such behavior as the response of a rational consumer the lowering of the price of care at the point of use (Newhouse, 1993).

Health insurance, public and private, tends to cause such moral hazard because of the difficulty of determining the true state of health. If it were easy to tell how large a reduction in health insureds had suffered, benefits could be provided in the form of health-state-specific pre-specified indemnity payments, and the only effects on spending would be income effects. However, because it is impossible to identify the state of health easily beforehand, insurance contracts tend to describe the state of the world by the amount of medical care spending. As long as medical services have positive value at the margin, insureds and providers paid a price above their marginal supply price will then supply larger quantities and qualities of care, the more generous is the insurance coverage. In the limit, if insurance coverage is full, all spending that will provide any positive benefit, however small and however costly, will be provided.

Several devices may be used to reduce the harm done by moral hazard. There are devices which operate on the demand side, and devices which operate on the supply side. The main demand side device is co-payment or co-insurance, both of which expose the insured to some portion of the expense in order to temper somewhat the effect of moral hazard. Since the last units of care consumed under full coverage have very low value, and since the fully insured person is protected against risk, we know that it will be optimal to have some positive level of cost-sharing – the exposure discourages the consumption of costly but virtually useless care, and exposes the insured to very little risk. More generally, there is a tradeoff between controlling this inefficiency or “welfare cost” of moral hazard and greater exposure to risk. The optimal level always involves positive cost-sharing but its extent varies across individuals if they vary by degree of risk aversion and/or demand for medical care and varies across medical services depending on their price elasticities of demand. Absent any other distortions, with a given technology, competitive insurance markets would supply policies that provide the optimal extent and form of cost-sharing.

Competitive equilibrium may fail to be optimal if technology can change in cost-increasing, but quality improving ways and insurers are not able to refuse to cover new technologies (Baumgardner, 1991). If they are able to rule out coverage of some technologies, given some exogenous supply of technology, the outcome will still be efficient (Ramsey and Pauly, 1997).

However, if the supply of technology is endogenous, because there are public good characteristics associated with innovation and because there are issues of optimal product variety, it is more difficult to come to conclusions about the efficiency of competitive markets compared to what an omniscient welfare-maximizing government might do. Even in this case, however, some cost-sharing will usually be optimal, but the optimal level may not be the one that will emerge in a competitive market.

The alternative type of control of moral hazard is through “supply-side” alterations of

two types. One type alters the marginal revenue received by providers of care. If providers have utility functions in which patient welfare, accurately measured, is included, and costs are constant, the optimal payment policy will be one which pays a marginal price below marginal cost but then pays a lump sum to make up the difference (Ellis and McGuire, 1990). If supply or marginal cost curves are upward sloping, there is a point on the supply curve (so called pure fee-for-service) which is optimal, but it may not be the competitive equilibrium price (Pauly, 1996), and may also therefore involve a lump sum payment to bring total provider compensation up to the competitive level.

Figure 1 shows an especially simple version of this argument. The quantity of some service  $Q^*$  is assumed to be the welfare maximizing quantity. If the insurer sets the marginal price at  $P^*$ , that quantity will be supplied, and it will be delivered to patients based on marginal benefit if the provider is at all concerned about patient welfare.

The other type of supply-side control is the use of a quantity limit, in the form of guidelines, benchmarks, and the like. The optimal level of such a benchmark trades off reduction in moral hazard in the case of patients less severely ill with rationing of cost beneficial care to severely ill patients. In both cases, “errors” occur because severity is imperfectly measured; if severity could be measured perfectly then indemnity benefits could be provided.

Both of these supply-side devices constitute the essence of so-called “managed care”, in which providers are subjected to varying rules or financial incentives to control moral hazard. The optimal use of these devices depends on variations in the demand for care, the welfare cost

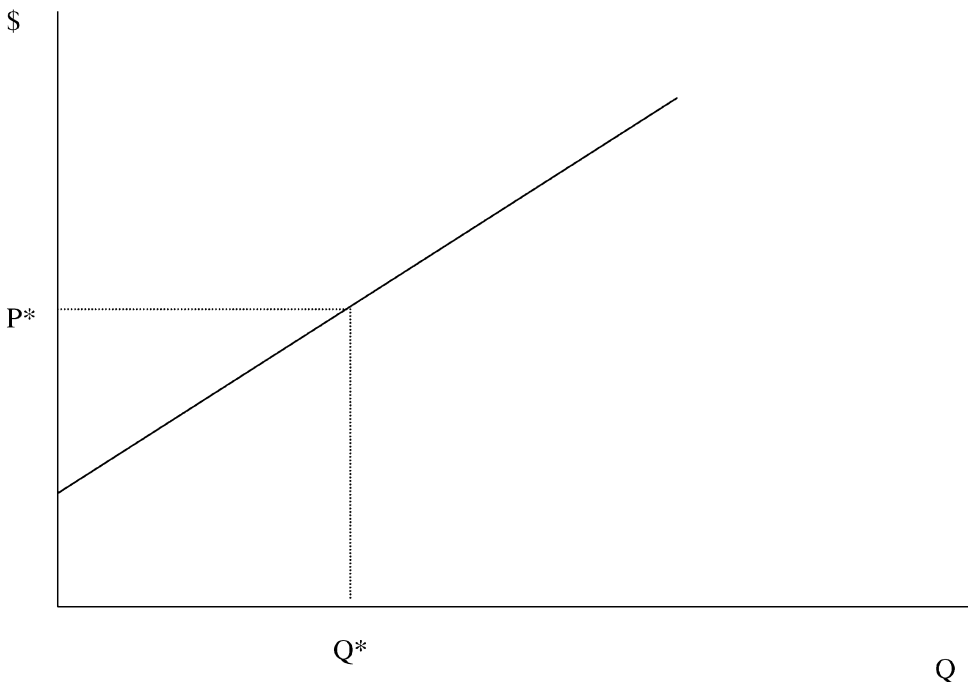


Figure 1

of errors under imperfect rationing rules or financial incentives, and any non-monetary costs associated with efforts to avoid rationing. The overall optimum, in a world of imperfect information, uses a mix of all three devices: patient cost-sharing, provider financial incentives, and rationing rules (Pauly and Ramsey, 1999). Here again, there is no single optimum, but rather the mix depends on variations in consumer demands, attitudes toward risks of different types, and the level of non-monetary costs.

## 6. Information/action asymmetry

A property of health insurance markets is that different people have different expected expenses in a given time period, based on such characteristics as age, gender, the presence of chronic conditions, or the continuation of a spell of acute illness. If insurers and insureds have identical information and are able to act on it, the result will be an efficient competitive equilibrium in which premiums vary with risks. We can distinguish two possible exceptions to this case. Insureds may know more about their expected expenses than insurers; this is the case of “essential” adverse selection. The other possibility is that insureds know which customers are higher risks, but are unable or unwilling to charge them higher premiums. In this case, it is permissible action, rather than information, which is asymmetric. One may then have “inessential” adverse selection and “cream skimming”.

I first consider the case of essential adverse selection. In this case, competitive markets will not necessarily achieve first best optimal outcomes, but the main question is one of second best. The equilibrium depends on how insurer behavior is modeled. No buyer is refused insurance, because insurers cannot distinguish among risks. However, in the simplest Nash equilibrium case, there cannot be a pooled equilibrium: either there is no equilibrium, or insureds of different risk levels end up buying different (separated) policies. Optimality judgments in this case are difficult to make, even in theory. In the case of no equilibrium, we know that there is no enforced equilibrium that will make all risks better off than they would be in one of the states through which the system cycles. In the case of separating equilibrium, we know that the low risks buy too little insurance compared to a world of perfect information, but it is very difficult to find a way to make all risks better off; even when such possibilities do exist, the resulting situation is incompatible with competitive markets, but usually requires some compulsory cross-subsidization.

There are few documented empirical examples in health insurance of adverse selection with separating equilibria. This is a matter of considerable uncertainty, but I remain skeptical that there would be substantial and important information asymmetries to be dealt with in a competitive health insurance market.

There are, however, many empirical examples of inessential adverse selection and resulting cream skimming. This occurs when private insurers are forced by regulation or by employers to ignore known risk-related characteristics. The market then behaves as one under adverse selection, though it does not have to do so. However, in this case, insurers do more than just design policies to appeal to low risks; they try to avoid selling any policies to high risks they can identify, something that does not happen in the case of essential adverse selection. In many ways, this situation achieves the worst of all possible worlds, with both risk sorting and refusal to insure.

Is it then ever optimal to charge the same premium to insureds who have different expected expenses? The answer is that it depends on the true source of the variation in expected expenses. If the reason is a characteristic not randomly assigned to a person, there is no efficiency gain from risk pooling. Averaging across the young and the old, or across people

who live in high-cost and low-cost areas, will only cause adverse selection and cream skimming.

However, if risk varies because of the presence of conditions whose period of treatment exceeds the period of coverage – for example, the case of a chronic illness such as diabetes or chronic obstructive pulmonary disease – it may be optimal to provide what Kenneth Arrow (1963) has called “insurance with a longer time perspective”, insurance that reduces the random fluctuation in lifetime premiums. Moderate year-to-year fluctuations in premiums are inconsequential from a lifetime perspective, but contraction of a condition which would lead to a long series of high single-period premiums is something a risk averse person would prefer to avoid.

In theory, competitive markets will provide protection against this risk. Methods using advance (but declining) payments, or adding a supplemental premium to pay higher future premiums, have been proposed (Pauly, Kunreuther and Hirth, 1996; Cochrane, 1996). An even simpler theoretical solution is to collect a single lifetime premium. However, all of these methods may have practical problems in a world of imperfect capital markets and imperfect foresight. Employment-based group insurance can provide an alternative vehicle for protection, but it and the other methods can be undercut if citizens eventually become eligible for community-rated social insurance, such as US Medicare (Pauly, Nickel and Kunreuther, 1998).

An alternative method of improving both efficiency and equity, consistent with our later, more extensive, discussion of collective intervention, is a system of credits or vouchers which vary with risk, matched with a competitive insurance market which collects higher premiums for higher risk persons. Such a system can provide protection to citizens against risk, and yet still furnish insurers an incentive to seek out and service people with chronic conditions. Such “risk adjustment” need not be perfect to provide a great deal of benefit. The key point for an optimal system is that two characteristics must hold: (1) the insurer must receive more net revenue when signing up high risk customers, but (2) the funds to pay for that risk adjustment must not be collected as an add-on to premiums charged to lower risks, but rather should be collected by non-distortive means such as general revenue taxation. The reason is that required community rating will discourage low risks from buying coverage or buying the right amount of coverage. (Of course, all of these considerations only matter in situations in which there is a voluntary insurance market, in which either the fact of coverage or the amount of coverage is decided by consumers. Where a given amount of social insurance is mandated, it does no harm if the payments for risk adjustment are tacked onto the compulsory payment for the insurance. But as soon as some element of market choice is added, such devices, common among many social insurance systems, begin to cause deviations from optimality even under ideal government management.)

Another argument for the collective risk adjustment system is that, especially for very high risks, it will doubtless prove difficult to administer and enforce either multi-period contracts or regulations which require competitive insurers to continue selling to risks on whom they are sure to lose money. At some point, both private contract enforcement and public regulation are likely to break down in the face of irresistible temptation to avoid bad risks.

## **7. Market imperfections**

Medical services insurance reimburses or pays for medical services. In all developed countries, insurance furnishes most of the funds for medical services. In addition, in all

developed countries, the supply of some types of medical services is regulated, either by government or by professional associations. Finally, buyers of medical services are often not as well informed as they should or could be about either prices or qualities.

All of these characteristics mean that, on the supply side, sellers of services may have some market power; they may be able to continue to sell some services even if they raise their price above the competitive level. In many cases as well, insurers function as buyers of services who may also have some market power, in the similar sense that if the insurer reduces price below the competitive level some services may still be supplied to it. In countries in which there is a single insurer or cartel of insurers, whether public or private, there can be monopoly in the sale of insurance *and* monopsony in the purchase of medical services.

The overall economic result is obvious: markets may not have the efficiency properties of competitive markets. The problem with provider market power or private insurer market power is also obvious: if prices are forced above the competitive level, quantity is reduced. The real cost of medical services is also reduced, but the transfer of monopoly rents to sellers means that monetary expenditures are likely to be higher than they would be under competition (or under more competition). Monopoly also can lead to lower quality, although it need not necessarily do so in economic theory, and especially may not do so when the seller is a private but non-profit monopolist. Quality can be too high, as well as too low, under monopoly. Either way, quantity is likely to be lower than under competition. In general, however, the problem with monopoly is the same as in other industries: consumers choose inefficiently low quantities and/or incorrect levels of quality.

The remedy for this problem is more competition, but for competition to work properly, the other conditions needed for optimality must be satisfied. The assumption made earlier – that the distribution of income over consumers is fair – must be satisfied, since competition directs output toward people with purchasing power and abhors cross subsidies. Thus competition may very well lead to less charity care or transfers to research or medical education, but the ideal solution is to increase public subsidy of the poor, of research, and of education, if society really agrees that more subsidy is desirable. A more serious conceptual problem is that buyers must be informed about price and quality. If they are poorly informed, monopoly will remain; if they are unevenly informed – for instance, having better knowledge of price than quality – competition may ignore quality. If they do not know prices, or are not exposed to price differentials because insurance coverage is nearly complete, competition will take place over quality. If the price level is then administered and set too high, quality may again be excessive, even under competition.

The perils of monopoly are generally well known, even though such institutions as professional licensure and patents, which create monopoly, are often tolerated in the medical care sector. Less well appreciated is the potential for harm from monopsony. Let us take the most counter-intuitive case first: suppose medical services consumers were to form a buyers' cartel (say, through a large consumer-controlled health plan). For those medical services with upward sloping supply curves, this buyer will try to exert market power. If the initial state had been one of monopoly, some reduction in unit price closer to the competitive level will be efficiency-increasing, but consumers can gain further by pushing price below the competitive level to the monopsony level. Quantity will again be too low; consumers will be better off, but suppliers worse off, and the loss to suppliers exceeds the gain to consumers (Pauly, 1998). From a single country perspective, the question may turn on whether suppliers are part of the population whose welfare is to be maximized. Nurses are usually citizens, but not necessarily owners of drug or device firms.

These observations also imply that medical care spending levels, in the presence of



monopoly, monopsony, or bilateral bargaining, may be poor indicators of real economic cost or of efficiency (Pauly, 1998). A national system that achieves low spending by pushing prices below the competitive level does not necessarily lower true cost (at least, not by as much as spending falls), and even low cost could be inefficiently and excessively reduced under monopsony.

## 8. Externalities

All countries provide some collective subsidization of health services for their citizens, and this subsidy takes the form of health insurance. One attractive way of understanding why this behavior occurs, and judging its efficiency, is to assume that this subsidy implements a collective demand for altruistic benefits. The assumption is that most people care about unrelieved but treatable illness on the part of their fellow citizens, sometimes (though relatively rarely) because they might contract a contagious illness, but more typically because they are bothered by suffering. In economics, we can think of this as a kind of externality, in which a good provides benefits to non-users as well as to users. My utility is reduced when you are ill, and it increases the more care you use to treat your illness, up to a point. The last phrase is important; the value anyone would attach to subsidizing an additional unit of care for a sick fellow citizen will decline the more care that person is using, and will go to zero at some point. (My assumption of no envy means that it never gets negative.) In effect, additional consumption of medical care by someone currently consuming at relatively low levels is a kind of public good, providing benefits to non-users as well as to the user. Ordinarily the benefit will be larger to the direct user than to others, but it can be positive for them.

The notion of optimality here takes into account these external benefits. It says that the optimal amount of care for a person to use at a given cost is that quantity at which the cost equals the *sum* of the valuations of *this person's care* over everyone in the community – the person and everyone else who is concerned about him. In the absence of public intervention, the actual level of use of many people (primarily the lower income or higher risk people) is likely to fall short of this optimal level.

There are several ways to get use to the optimal level for each person. In theory, an omniscient welfare-maximizing government might simply allocate to each citizen the optimal quantity (given their illness and, this is important, their tastes), and raise the money through non-distortive taxation. The market can never do better than this ideal government, even if there were no public goods. However, both the assumption of ideal government and the assumption of ideal taxation would be questioned. In such a case, we then need to think about second-best solutions that recognize some of the constraints on what governments can or will do. The other case would be for the ideal government to mandate that each citizen use the optimal quantity and pay for it themselves. Despite the substantial difference in tax burden, these two solutions are identical in terms of their medical sector efficiency; they differ only in the distribution of cost.

In my work, I have taken the modest compromise attempt of trying to specify what optimal insurance would look like in a situation where the amount of public subsidy is minimized and yet only voluntary behavior assumed. While a technical treatment is possible, here I primarily give the intuition. I make two key assumptions. First, the demand for medical care, given illness level and user price, is normal. Richer people will want to consume more (though perhaps as quality rather than quantity). Second, the optimal level for everyone is not the level demanded when all care is made free by insurance, but is lower than that level for much of the population.

Given these assumptions, we can consider two polar cases. First, we will want to fully subsidize the care of very low income families; we do this by offering them a full coverage insurance that is tax-financed. Second, the amount of care high income people will buy on their own (given the insurance they would choose) is so great that there are no external benefits *at the margin*. This is the case that Buchanan and Stubblebine (1962) called one of infra-marginal externalities; the rest of the community is better off than if the rich person had used no care, but the value of additional care is zero. It then follows that the level of subsidy to care should decline as income rises, going from its initial very high value down to zero. Probably this would mean that the subsidy to insurance should also decline with income, but the minimum level of coverage required in order to receive the subsidy should also decline with income. One simple version of such a scheme, suggested long ago by Martin Feldstein (1971) and myself (1991b), would be a set of catastrophic policies with maximum deductibles related inversely to income, and with the deductible falling to zero for very low income people. Note that such a scheme means that cost-sharing should certainly be allowed, but it should be limited to non-poor people.

One useful addition to this minimally optimal scheme would note that the insurance subsidy needed to induce the purchase of optimal coverage might have to be larger for higher risk persons (since their premium would be higher than for lower risks). The final model then is one where premium subsidies might vary with income and risk.

If we think instead of plans that rely on supply-side limitations to deal with moral hazard, the same type of income-related subsidy to insurance purchasing would apply. What about the tightness of the limits? This is a rather more speculative area, since we know much less about what affects the rate of use of care under insurance with given supply-side limits. There is some reason to believe that the actual rate of use of care is still positively related to income. For instance, in the US Medicare system that provides essentially complete and uniform coverage at administered prices to all, McClellan and Skinner (1997) have recently found a strong positive relationship between income and lifetime benefits. Part of the reason for this conclusion is that higher income people tend to live longer and therefore receive benefits for more years, but another reason is that they receive higher benefits per year. They make more frequent use of the system, they use care of higher intensity and, most importantly, they tend to live in areas where technology, access, use given access, and prices are all high. My impression is that the same relationship with income tends to hold in national health services, although the effect of income on spending is surely less strong than in market systems without universal subsidies. There is little evidence on the effect of income in a system with competing in a managed care program. In the US higher income people do tend to join more costly managed care plans, ones with fewer restrictions and more expensive providers.

The main alternative to income-related coverage would be provision to all of a uniform credit or voucher equal to the cost of the optimal managed care plan for the lowest income subgroup. If this cost is less than the desired cost for all people with higher income, then supplementation would be possible. However, I think it highly likely that the cost of optimal low income policy will be greater than the cost of the (unsubsidized) optimal policy for at least some higher income people.

All of this suggests, as before, that the managed care choices of the well off do not need to be subsidized; they are infra-marginal here as well. Almost surely the poor need to be subsidized to choose a managed care plan more generous than they would choose on their own. What about the middle class? If they were put in the same type of plan as the poor of equal health levels, they would use more care in some dimensions and perhaps the plan would become more costly.

If so, there is another reason for income-related subsidies. To the extent that managed care plans are introducing more patient cost-sharing (as they are), that cost-sharing should be subsidized away for the poor but not for the middle class. Beyond this there still remains some ambiguity in benefit design and supply responsiveness. One recent economic model of behavior under managed care with supply-side limits indicates that providers bargain with patients when there are limits, and arrive at a point that is intermediate between what the supplier prefers to supply and what the buyer would like to demand (Ellis and McGuire, 1996). If this is true, and if the better off do not have less bargaining power than the poor (surely plausible), then the conclusion again is that the optimal plan for the poor should have relatively weak supply-side limits, but those minimum limits can be tightened as income rises. Again, the better off with tastes to avoid strict limits can optimally buy more coverage than this minimum if they pay the full cost themselves, but there should be no subsidy to more gentle managed care limits.

### 9. Beyond the extremes: the optimal extent of government involvement

As noted, government could in theory regulate every citizen's use of medical care and get it right. Likewise, the point at which income is judged to be high enough for government to cease controlling, so that people are given private insurance market options, could be quite high if we thought that government was able and willing to prevent every small externality. In a world of less than perfect government, we might not want it to interfere so much. One factor which may help to define the ideal extent of involvement is the degree of variation in demand (heterogeneity of preferences) among citizens. Even when almost all citizens would demand the same quantity and quality of care, governments might not make the right choice. But if there is substantial heterogeneity, a government constrained to treat everyone the same will surely not do the optimal thing for everyone. The optimal extent of government involvement, perhaps proxied by the income cut-off as well as by the strength of regulation, should be greater in homogeneous communities than in heterogeneous ones. The great virtue of the market (despite its deficiencies) is not necessarily productive efficiency; some government systems are fairly efficient (US Medicare or Canadian insurance administration, for example). The real virtue of markets is in catering to diversity. Even here, buyers must be informed and market power must be absent.

### 10. Conclusion

This last thought allows us to conclude that different communities may choose different systems and yet they may all be optimal. As usual, the most important requirement for optimality is that the instrument fit the situation. Beyond this, there is a role for patient cost-sharing, provider incentives, and collective subsidies to offset income and risk variation.

#### REFERENCES

- ARROW, K.J., 1991, "Uncertainty and the Welfare Economics of Medical Care", *American Economic Review*, 53(5), pp. 339–347.
- ARROW, K.J. and DEBREU, G., 1996, "Existence of an Equilibrium for a Competitive Economy", in Debreu, Gerard (ed.), *General Equilibrium Theory. Volume 1*. Elgar Reference Collection, International Library of Critical Writings in Economics, no. 67, Cheltenham UK: Elgar, pp. 38–63.
- BAUMGARDNER, J., 1991, "The Interaction Between Forms of Insurance Contract and Types of Technical Change in Medical Care", *Rand Journal of Economics*, 22, pp. 36–53.

- BUCHANAN, J.M. and STUBBLEBINE, W.C., 1962, "Externality", *Economica*, 29(4), pp. 391–384.
- COCHRANE, J.H., 1995, "Time Consistent Health Insurance", *Journal of Political Economy*, 103(3), pp. 445–473.
- ELLIS, R.P. and McGUIRE, T.G., 1990, "Optimal Payment Systems for Health Services", *Journal of Health Economics*, 9(4), pp. 375–396.
- ELLIS, R.P. and McGUIRE, T.G., 1996, "Hospital Response to Prospective Payment: Moral Hazard, Selection and Practice-Style Effects", *Journal of Health Economics*, 15(3), pp. 257–277.
- EVANS, R.G., 1997, "Going for the Gold: The Redistributive Agenda Behind Market-Based Health Care Reform", *Journal of Health Politics, Policy & Law*, 22(2), pp. 427–465.
- FELDSTEIN, M.S., 1991, "A New Approach to National Health Insurance", *The Public Interest*, 4, pp. 93–105.
- GAYNOR, M. and VOGT, W.B., 1997, "What does Economics Have to Say About Health Policy Anyway? A Comment and Correction on Evans and Rice", *Journal of Health Politics, Policy & Law*, 22(2), pp. 475–496.
- KIHLSTROM, R. and PAULY, M.V., 1971, "The Role of Insurance in the Allocation of Risk", *American Economic Review*, pp. 371–379.
- McCLELLAN, M. and SKINNER, J., 1997, "The Incidence of Medicare", *NBER Working Paper 6013*, April.
- NEWHOUSE, J.P. and The Insurance Experiment Group, 1993, *Free for all? Lessons from the RAND Health Insurance Experiment*. Cambridge and London: Harvard University Press.
- PAULY, M.V., 1971, *Medical Care at Public Expense*. Praeger Publishers, Inc.
- PAULY, M.V., 1971b, *National Health Insurance: An Analysis*, Washington: American Enterprise Institute.
- PAULY, M.V., 1995, "When Does Curbing Health Costs Really Help the Economy?" *Health Affairs*, 14(2), pp. 68–82.
- PAULY, M.V., 1995b, "Paying Physicians as Agents: Fee-for-service, Capitation, or Hybrids?", in Abbott, T (ed), *Health Care Policy and Regulation*, Amsterdam: Kluwer Publishers.
- PAULY, M.V., 1997, "Who was that straw man anyway? A comment on Evans and Rice", *Journal of Health Politics, Policy & Law*, 22(2), pp. 467–473.
- PAULY, M.V., 1998, "Managed Care, Market Power, and Monopsony", *Health Services Research*, 33(5), pp.1439–1460.
- PAULY, M.V. and RAMSEY, S.D., 1999, "Would You Like Suspenders to Go with That Belt?", *Journal of Health Economics*, 18(4), pp. 443–458.
- PAULY, M.V., KUNREUTHER, H. and HIRTH, R., 1995, "Guaranteed Renewability in Insurance", *Journal of Risk and Uncertainty*, 10, pp.143–156.
- PAULY, M.V., NICKEL, A. and KUNREUTHER, H., 1998, "Guaranteed Renewability with Group Insurance", *Journal of Risk and Uncertainty*, 16, pp. 211–221.
- RAMSEY, S.D. and PAULY, M.V., 1997, "Structural Incentives and Adoption of Medical Technologies in HMO and Fee-for-service Health Insurance Plans", *Inquiry*, 34(3), pp.228–236.
- RILE, T., 1999, "Can Markets Give Us the Health System We Want?" *Journal of Health Politics, Policy and Law*, 22(2), pp. 383–426.