

# **The Demand for Health Insurance: Insights from Theory and Voluntary Markets in Less-Developed Countries**

*Mark V. Pauly*

*August 2004*

## **Introduction**

The basic empirical fact that motivates this look at the theoretical foundations for potential voluntary insurance purchasing in developing countries is this: In most of those countries, there is a substantial proportion of total market-level medical care costs that is paid out of pocket by citizens. While the fraction paid in this fashion varies to some extent across countries at similar levels of income depending on the form of public programs, it is almost always relatively large. For example, the percentage of national health expenditures paid out of pocket is estimated to be 80 percent in Vietnam, 47% in Indonesia and 26% in Columbia. In contrast, even in the United States which has the least ostensibly extensive program for mandatory health insurance among developed countries, the percentage paid out of pocket by citizens in 2002 was less than 14%.<sup>1</sup> As Cutler and Zeckhauser put it, “health insurance is common to all developed countries,”<sup>2</sup> but, beyond nominal public provision of limited amounts of care, it is uncommon in many developing countries.

An important positive question which I shall not address directly here is why the demand for insurance coverage in developing countries is so low. Instead, I will focus on the policy question

of whether, based on the economic theory of the demand for insurance, there *could and should* in principle be voluntary demand for insurance in these settings. I especially want to explore whether there could be a voluntary market for insurance for people below the highest income stratum, although I will not expect that there could be significant voluntary demand among the very poor. But, in between these two extremes, a large fraction of the population of developing countries might be able and willing to obtain at least some welfare-improving insurance; the theoretical basis for that possibility is what will be explored here. I will make some assumptions about the form and characteristics of insurance supply (to be discussed in a companion paper), and then ask whether, given those assumptions, one might expect to see voluntary insurance demand in positive and potentially significant amounts. I will also comment on alternatives to voluntary insurance, such as mandated coverage.

There have been many excellent studies of the insurance of various types—private, community based, and voluntary government-supplied—in developing countries. Virtually without exception, when those studies use the theory of insurance, they use it to comment on specific aspects of a system or design that is already in place. In this paper I intentionally take a different approach; I reverse the process. That is, I begin with theory and ask what could or should be possible to put in place. The theory of insurance demand (and the theory of insurance supply in a companion paper) is spelled out, with the intent of isolating in the theory those elements that should be present in insurance markets and insurance arrangements regardless of institutional structure, and (within broad limits) regardless of the level and distribution of household incomes. I address the question of what characteristics on the demand side are required for voluntary insurance markets to exist. Only after this exercise do I turn to consider aspects of institutional

structure in developing countries that may foster or inhibit what is in theory possible. In short, rather than using theory to comment (often inconclusively) on something already developed, I develop the framework for good design based on theory, and then use it to comment on what is or could be put in place.

### **Toward an Applicable Theory of Medical Insurance Demand**

Based on insurance theory, one of the most important necessary conditions for there to be demand for voluntary health insurance is known to be present in many developing countries: there exists a relatively high level of unpredictable out-of-pocket payments for medical services. This fact alone suggests both that the medical services voluntary insurance would cover are, at some positive quantity, worth more to consumer than their prices, and that there is a financial risk to be protected against. It suggests, virtually by definition, that such insurance would be widespread in the sense of covering a large fraction of total medical care spending (though, of course, it does not guarantee that the spending will be “adequate” by some normative definition).

This fact also immediately rules out one commonly mentioned explanation for the absence of voluntary medical insurance: that consumers (in general) cannot “afford” the insurance. While “affordability” (sometimes called “unfavorable economic and social conditions”<sup>3</sup>) does not have a precise economic meaning in any case, in this “behavioral” situation we can be sure that, at least for some consumers, insurance is “affordable,” precisely because the alternative to insurance—payment out of pocket—is voluntarily made. If the high out-of-pocket payments are “affordable,” and they must be if consumers are willing to make them, then insurance is in

principle even more affordable. Of course, this argument also allows for the possibility that even those who would not have had sufficient income to make voluntary out of pocket payments would demand insurance. A household unwilling to pay a high but rare out of pocket expense may still be willing to pay (to “afford”) the lower annual premium to cover that expense, as the work of Nyman discussed further below shows.<sup>4</sup> (This is related to the “truncated elasticity” point made by Vate and Dror.<sup>5</sup>)

The empirical evidence from developing countries on the distribution of out of pocket spending across the population tells us more. It shows that such spending is indeed fairly widespread, and is definitely not concentrated exclusively on a small sliver of very well off people. Data from Jamaica,<sup>6</sup> for example, show that, although average out of pocket expenses to supplement public medical care are eight times greater in the highest quartile than in the lowest, surprisingly they are positive even in the lowest quartile. More importantly, they are still high in the middle quartiles. Data from Rwanda also show positive amounts of out of pocket spending across the income distribution, with spending in the top quartile (which still represents a very low level of income) about four times as great as spending in the next lower quartile. Another kind of data describes spending during rare but severe episodes of illness. For example, Arhin<sup>7</sup> found that actual expenditures for 70 percent of hospital episodes exceeded 4-6% of income in Africa..

This information is usually presented as evidence for the high burden of out of pocket spending and the need for insurance, and it certainly does provide eloquent testimony to that. But because the spending is actually being made by consumers, that fact alone provides a bedrock foundation for demand for health insurance to smooth and spread that spending. Consumers, the theory says,

ought at least to be willing to pay for insurance as much as they would expect to pay out of pocket for covered services.

Since the initial valuation question—whether formal purchased medical care is of sufficient value to induce many consumers in developing countries to sacrifice other types of consumption to be able to obtain it—is answered in the affirmative, that naturally leads to the next question of whether using insurance to make these payments rather than risking out of pocket spending could also be of sufficient value to make voluntary (and presumably welfare-improving) insurance feasible.

### **The Theory of Insurance Demand**

Consider a set of non-poor households that are identical in terms of the beginning-of-period health status of members, income, education, and all other factors that might affect the demand for health or the demand for medical care. We know that some, and probably many, such households will over the period be relatively free from illness and therefore have low medical care demand and low potential out of pocket expenses. At the other extreme, the random nature of illness means that a small minority of households will have potentially high levels of medical care spending. That is, a few households will after the fact have much above average medical spending and many households will have below-average spending. (The rule of thumb here is that about 20 percent of households will make about 80 percent of this population's medical care spending.) This proposition presumably holds even if we are describing medical spending in developing countries which is quite low on average relative to developed countries.

Nevertheless, spending still varies substantially around that lower mean; some people in developing countries make out of pocket payments which are large relative to their incomes and even large relative to the developed country mean. There is no definitive evidence on the shape of the distribution of out of pocket spending in developing countries among the non-poor (as measured, say, by the variance), other things equal.

Assume for the present that the level of spending will be the same whether covered by insurance or not. If people attach value to health and to other items of consumption, and if utility is increasing in consumption but at a declining rate (diminishing marginal utility of consumption or “income”), it then follows that households will be willing to pay for insurance that covered all of their (given) out of pocket spending an amount that exceeds the expected or average value of that spending. That is, if there was an insurance arrangement that could make insurance available at a premium  $P$  that equaled the expected value of medical spending  $m$  (or  $pm$ , where  $p$  is the [average] probability of illness), the members of these households would prefer to obtain that insurance than to face the risk of out of pocket payments of varying and potentially large (but rare) amounts. By giving up moderate amounts of consumption in the “lucky” states in which no one in the household becomes seriously ill, people who buy insurance can avoid drastic cuts in present or future consumption of other things in the rarer but possible “unlucky” event of a serious illness that needs to be treated in a costly way.

This is an important point. Although (as we will see) there are some side effects of insurance which may lead to inefficiency (or inequity), in a most fundamental sense insurance is desirable for people to have because they desire to reduce the chance of impact of unexpected shocks to

their levels of overall consumption. While the utility gain from purchasing insurance to cover formerly out of pocket medical expenses can be quite substantial, that gain is usually not matched by substantially higher national output measured in money terms, and therefore is missed by the usual crude indicators of economic wellbeing, such as GDP. For example, if consumers in Indonesia were able to convert the approximately 4 percent of GDP represented by out of pocket medical payments into insurance, and that insurance was ideally efficient (in the sense of not stimulating inappropriate consumption of medical care), GDP would rise only by the administrative cost of the insurance.

But that measure could substantially understate the value to citizens of insurance. The value of insurance can in theory be measured by the “risk premium,” the amount in excess of the actuarially fair premium that people would be willing to pay for coverage rather than go without; this amount could be much higher than the administrative cost actually paid, but it is only the increased spending, not the gain to consumers (in the form of consumers’ surplus) that is observed. While we do not have precise measures of the risk premium even in developed countries, based on willingness to buy insurance at positive loadings we might conclude that the risk premium on average could be half or more of the expected expense. Thus the utility gain in the Indonesia example could be equivalent to 2 percent of GDP or more.

To the extent that it has addressed this question, research on developing countries has generally tracked the other (observable and desirable) correlates of the availability of insurance—such as enhanced consumption opportunities, less need for inter-family transfers or other insurance substitutes.<sup>8</sup> Here it is important to note that providing more self-financed insurance is quite

unlikely to increase aggregate total consumption. It does of course permit greater consumption of non-medical goods by those who, if uninsured, would have had high medical expenses. But the premiums or taxes needed to pay for it reduce such consumption, as does the necessary administrative loading. The mix of consumption would be changed: high medical bills presumably cut into consumer durables and housing expenditures (and investment too), whereas premiums probably affect routine spending like food and clothing. Only externally-funded (e.g., by donations or taxes on the wealthy) insurance can increase aggregate consumption.

Even these measures, valuable as they are, would tell much less than the full story relative to a correct measure of welfare that was tied to wellbeing rather than only to purchased consumption. Moreover, some of the consequences of the absence of insurance are behaviors which actually tend to increase measured GDP. Uninsured households save more (as a precautionary insurance substitute), and they probably have greater labor supply. The main point here is that, although the theoretical welfare gains from insurance are potentially enormous, they will not generally be well manifested in the economic aggregates policymakers typically monitor.

There is one important empirical implication of the theory of insurance demand: it defines a point on the insurance demand curve. If premiums are actuarially fair, the theory says that all risk averse people should buy insurance whose benefits just equal the amount of loss. Thus if insurance is subsidized sufficiently to reduce its explicit premiums to the actuarially fair level or lower, there should in theory be 100% takeup and 100% coverage. In reality this may not happen because consumers are poorly informed about the value of insurance, because they believe that their expected expenses are lower than those on which the premium is based, or

because moral hazard makes the premium much higher than expected expenses without insurance. .

How much *more* than the expected value or “fair” premium risk averse consumers would be willing to pay depends on how rapidly the marginal utility of income diminishes. If the utility function takes the form  $U(y, h)$  where  $y$  is consumption and  $h$  is health, the coefficient of absolute risk aversion (CARA) is the rate at which the marginal utility of income diminishes, or  $-U''/U'$ . If this coefficient is positive, a consumer will be willing to pay an amount in excess of the expected value of benefits, the so-called “consumer risk premium.” The fragmentary evidence on this question definitely shows that people in some developing countries are risk averse and would seek insurance,<sup>9</sup> but that not all may be so eager to obtain it.<sup>10</sup>

Obviously, in the real world where supplying insurance entails some real administrative costs, for insurance to be demanded in positive quantities the price consumers are willing to pay must exceed the price suppliers charge, and the price suppliers charge (in the absence of subsidies) must exceed the expected value of losses because it must cover those administrative costs. In the case of insurance, the price of insurance is not the total premium. Rather, it is the so-called “loading”: the excess of the price over a premium equal to the expected value of benefits, usually expressed as a percentage of premium or benefits. In market insurance, even in unsubsidized competitive markets, this price will be positive; insurance will not be free.

We can say more about willingness-to-pay this price for insurance than we can about other products, since one of the incentives for buying insurance is to reduce risk. Thus we can say that

a positive voluntary demand for insurance requires that consumers be risk averse enough to pay a premium in excess of the fair value. In short, the strength of risk aversion determines consumers' willingness to pay a premium in excess of the fair premium. In developed countries, judging from revealed preferences, consumers generally seem willing to pay the loading of 50 to 80 percent of expected benefits that is typical for individual insurance. There is little evidence on the strength of risk aversion in developing countries per se. However, we can ask whether, for a given monetary amount of loss, risk aversion is related to wealth.

On the one hand, insurance premiums cut a larger proportion of consumption for lower wealth people but, on the other hand, a given loss without insurance does the same thing. If people have constant *relative* risk aversion (CRRA) utility functions as is commonly assumed, the loading a person is willing to pay against a loss which is a given proportion of wealth remains constant. (CRRA is approximately  $[CARA]/U$ )

We do have some idea of the strength of risk aversion for risky health expenditures in developed countries because we can observe what levels of net loading people seem to be willing to pay. We know that when tax advantages reduce the net loading for non-poor people to negative levels, as they do for employment based health insurance in the United States, the take-up rate is very high, probably on the order of 90 percent or more. (We also know that about 30 percent of eligible poor people do not sign up for free insurance, which either reflects behavior inconsistent with theory or a high implicit price for enrolling.) At the other extreme, for people not eligible for tax subsidies who would have to buy in the individual health insurance market with loadings on the order of 30 to 40 percent of premiums, we know that the take-up rate is in the range of

about 25 percent, but that it increases with income, chronic conditions, and age (even when premiums are risk rated).<sup>11</sup>

I am not aware of any direct studies of the strength and distribution of risk aversion, and resulting maximum risk premiums people would be willing to add to fair premiums, in developing countries. We do know that some kinds of consumer insurances (such as life insurance) are successfully sold to a non-negligible minority of their population, even at loadings in the range for individual health insurance. So there is some reason to be optimistic, but clearly more empirical work on the average level and distribution across households of the risk premium is needed. At a minimum, it does seem plausible that consumers in developing countries would be risk averse. Thus we can conclude that *there will be positive demand for insurance at a non-zero loading*. But will that demand turn into actual positive purchases in the market? The answer to that question obviously depends as well on what the market equilibrium level of loading will be. If it is “too high” relative to consumers’ degree of risk aversion, no or few purchases may occur. Thus *a voluntary insurance market may fail to exist if markets cannot supply insurance at loadings which are low relative to consumers’ risk aversion*. Loading is primarily a supply phenomenon, but it is clearly important in determining whether there will be observed demand for insurance. If risk aversion is low relative to minimum feasible loading, there is no chance for voluntary insurance that is not subsidized. The possibility of markets thus depends crucially on this important (but relatively straightforward) empirical question. Trying to get some approximate answer to it would seem to be necessary even to judge whether it is worthwhile to explore further.

This discussion also has implications for the role of voluntary insurance in of “resource mobilization” for medical care. As Arkin-Tenkorang<sup>12</sup> has noted, new resources for health care are “mobilized” to the extent that insurance premiums exceed benefits paid out plus administrative costs. Any possible margin in excess of these two costs presumably can be made available for other purposes. But the theory says that there can be substantial utility gains even in the absence of any appreciable margin. Moreover, resource mobilization for medical care is a goal for the economy as a whole only to the extent that use of those resources in providing health care is more valuable than their other uses.

### **When Is Insurance Worth the Most?**

How is willingness to pay for insurance by a risk averse person related to the size and probability of a potential loss? The most useful proportion is this: for *risks of equal expected value*, the willingness to pay an administrative loading increases as the size of the loss increases (and the loss probability correspondingly decreases). Catastrophic coverage is worth more than “front-end” coverage. In the limit, as the loss probability approaches one and the premium therefore approaches (or even exceeds) the amount of the loss, insurance demand goes to zero.

This simple observation leads to an important intuition: if out of pocket payments (losses) can vary in amount, and if the loading is positive and proportional to the expected value of out of pocket payment, it will be rational to buy insurance with a “deductible”: a provision that excludes small losses from coverage. The reason is that the willingness to pay loading for insurance that covers such small, highly likely events will itself be small and, in the limit, must

be less than any positive loading. Thus the optimal extent of coverage, if losses vary inversely in amount and probability (as they do with medical care), will be less than complete coverage. I will return to this point below.

### **Moral Hazard: What if Insurance Affects the Amount of Loss?**

Another aspect of insurance, and especially health insurance, which will affect the amount and type of insurance demanded is *moral hazard*. This refers to situations in which the expected loss is affected by the presence and extent of insurance. (Sometimes moral hazard is defined as changes in excess of those due to any income effects; see below.)

Moral hazard in health insurance can take two forms. It is possible that the presence of insurance coverage of the cost of care received when a person becomes ill may affect the actions he takes that affect the probability of illness (Type-1 moral hazard). For example, a person who has full coverage of the cost of treatment of the “flu” may not be willing to pay the cost and make the effort to get flu vaccine, or may not make the effort in terms of hand washing to reduce that chance. If the insurer cannot tell whether the consumer was vaccinated or taken other costly precautions (hidden action) claims will be higher with insurance than without. It is also possible that the amount and cost of care used, once illness has occurred, may be affected by the presence of insurance (Type-2 moral hazard). Because insurance reduces the user price of medical care and because the premium a person pays is usually independent of that person’s use, the person responds to the lower out of pocket price by demanding more medical care and possibly more expensive types of medical care.<sup>13</sup> If the insurer cannot determine exactly how sick the person is

(hidden ex-post severity), it may be forced to use the level of spending as an indicator of the amount of loss. Then actual losses will be larger with insurance than without it.

In developed countries, extensive empirical analysis shows that significant moral hazard does characterize the kinds of insurance contracts or policies generally used. The great bulk of moral hazard for health insurance appears to be type-2 moral hazard. Full coverage insurance that makes all care free of any out of pocket payment has been shown to lead to increases in use compared to coverage with larger deductibles (but still with wealth-related catastrophic coverage) of nearly 50 percent, with very modest improvements in health outcomes. The extensive Rand Health Insurance Experiment obtained an implicit price elasticity on the order of 0.1 to 0.2. Other research suggests that elasticity varies to some extent across types of care, and can range as high as 0.7. Even a relatively low numerical value of elasticity can imply a high impact on spending if the change in coverage involves a large percentage change in out of pocket payment. For example, cutting a proportional coinsurance rate from 40 percent to 20 percent implies 50 percent reduction in the out of pocket price.

Research on the effect of insurance compared to no insurance is less definitive. The expansion in use is surely larger than that associated with varying the extent of coverage over the positive range, and the health effects are probably greater, but the magnitudes of those effects are not known with great precision.

The relationship of household income to moral hazard has been studied to some extent. The Rand experiment found that income exerted the expected positive effect on use, especially

outpatient use, at any level of co-payment. However, there was no evidence of a significant difference by income level of the effect of cost sharing on use when cost sharing was capped as a percentage of income; the demand elasticity did not vary by income. This is somewhat counter-intuitive; one would imagine that the effect of income on use should be greater for lower income people. Perhaps other costs (e.g., transportation), not covered by insurance, continue to constrain the use of low income people even when care itself is free, and this causes use to vary directly with income even then. I will return to this point below.

Most fundamentally, type-2 moral hazard causes people who seek protection against financial risk to face distorted incentives at the point of use of care. These distorted incentives cause them to use care that is worth less to them than its cost or price; this use of low (but positive) value care gives rise to the so-called “welfare cost” of health insurance. Other things equal, consumers would prefer insurance forms or types which limit this overuse, but usually limiting moral hazard also reduces the protection against risk for which insurance is demanded in the first place. One way to reduce moral hazard is to increase the extent of consumer cost sharing, but this increases the *financial* risk the consumer faces. Another way to reduce moral hazard is to offer incentives to suppliers to limit the use of low-value care (managed care), but this increases *care* risk—the risk that the person will not obtain care which really is worth its cost—unless supplier decisions are perfect and/or consumers are forced to pay out of pocket for insured care suppliers refuse to furnish. (If the person is willing to go “out of plan” and buy care denied by the managed care plan at its full out of pocket price, “care risk” gets retransformed back into financial risk; however, there are often substantial impediments to such supplementation.)

The general theoretical proposition with the strongest support here is that, holding everything else constant including the consumer's risk aversion, *optimal (and demanded) insurance coverage will be less generous, with either financial or managed care limits, the greater the extent of moral hazard.* Some medical services which display rather high degrees of moral hazard do indeed tend to have less voluntary insurance in developed countries. This is surely the case with dental care, and may also be true of mental health care and outpatient prescription drugs. Conversely, the generally high coverage of inpatient care may be explained by relatively low moral hazard combined with a quintessential low-probability, high loss scenario.

The more general point here is that the most attractive kind of insurance coverage in the presence of moral hazard will involve some patient cost sharing, and that ideal cost sharing is unlikely to be uniform. In reality cost sharing often takes the form of uniform "coinsurance" (percentage cost sharing), which provides incentives to consumers to pay attention both to the level of use and to the relative cost of different types of treatment. Sometimes it takes the form of "co-payment" (fixed monetary payment per unit of care), but even here the co-payment typically is greater for a unit with higher price (brand name drugs) than a unit with a lower price (generic drugs).

What is the effect of moral hazard on the emergence of voluntary insurance? One might suppose that moral hazard might prevent insurance from coming into existence. However, under typical assumptions this supposition is not correct.

Here is why. If the administrative cost of insurance is a constant proportion of expected benefits, and if the person would have bought some insurance in the absence of moral hazard, theory predicts that the person should still voluntarily buy *some* positive amount of insurance even with moral hazard—just less coverage. The intuition for this result is as follows<sup>14</sup>: begin with no insurance, and imagine buying insurance that covers one percent of total spending. If the person is risk averse, this marginal protection against risk will have positive value. But the welfare cost associated with reducing the out-of-pocket price to 99 percent of its true value will be close to zero because the additional care is worth almost (but not quite) what the care truly costs. Thus at least some coverage will be utility increasing. The optimal (utility-maximizing) coinsurance rate will be that rate at which the marginal benefit for increasing coverage by one more percentage point just equals the marginal welfare cost associated with that charge. Since the marginal benefit for risk reduction decreases as coverage increases, and since the marginal welfare cost of moral hazard increases from zero at zero coverage, we know that there will be an interior solution with positive (but less than complete) coverage. Thus *moral hazard is not a reason for the absence of an insurance market, but it can cause coverage in that market to be smaller than if there were less moral hazard.*

This result is modified if (as is almost surely the case) there is a fixed cost to purchasing any amount of insurance, along with a loading that perhaps increase with the generosity of coverage. (This is the cost structure for term life insurance,<sup>15</sup> and almost surely applies to health insurance as well.) Conversely, if there are some costs associated with the illness that are not covered by insurance (e.g., loss of income, pain and suffering), nominal coverage can be 100 percent of medical care costs or even more than 100 percent.<sup>16</sup>

How will moral hazard affect the demand for insurance? As noted it will cause the optimal insurance (from the buyer's viewpoint) to be less than full coverage. At the optimal coinsurance rate, the marginal welfare cost from increasing coverage just equals the marginal welfare gain from reducing risk. Thus if we know the total gross risk premium at this level of coverage, we can roughly calculate the net premium by subtracting from it the total welfare cost to that point. For example, suppose insurance which covers half of out of pocket costs increases use by 20 percent. Then the marginal welfare cost is 10 percent of spending. If the risk premium for this level of coverage were, say, forty percent of expected expenses without moral hazard, the risk premium with moral hazard would be  $(40-10)$  or thirty percent. One implication of this discussion is that the maximum willingness to obtain insurance occurs if coverage is set at the optimal level. If in contrast people are only allowed to obtain coverage which is much more generous than they would have chosen, they may decline that coverage even if the loading were zero.

Although the use-stimulating effect of moral hazard is a negative influence from a buyer's point of view (setting aside Nyman's point to be discussed later), it may be viewed as positive from a societal perspective. Moral hazard is, after all, equivalent to (or, more precisely, the consequence of) increased access. If this access and use is valued at more than its costs by others, even if it is not greatly valued by the consumer (for example, because of low income), society may wish to cause excessive moral hazard. Of course, sometimes moral hazard is explained as "use of unnecessary care," but if "unnecessary" means "useless," that is not economically correct; in the economic model (so far) consumers do not pay for useless care.

Rather the care rationed out by co-payment is care that is beneficial, but whose benefits to the consumer (in terms of willingness to pay) are less than its cost. While the income-constrained consumer may want to avoid such low value use in order to conserve household resources for more immediate needs, society may feel differently. This is even more likely to be the case if the care reduces illnesses that are contagious.

This point seems especially relevant to developing countries. One of the positive reasons why insurance is said to be desirable from a public policy perspective is because it promotes “access to care.” Far from being regarded as a welfare cost, the additional use that follows from insurance may be thought of as achieving social objectives, especially if the use occurs among people with moderate incomes.

But the preceding discussion indicates that there is a tradeoff that is usually ignored in such policy discussions. If the insurance is to be voluntary, the greater the extent of such increased access, the smaller the willingness of people to pay the insurance premium. Absent a subsidy, the power of insurance to stimulate socially desirable use is constrained. One could specify what set of regulations on coverage that would maximize additional access and use, but this increase will probably be distributed unevenly.

Finally, this discussion raises a question that is (or ought to be) at the heart of public policy toward insurance in any country. If additional access to care is not worth its cost to citizens who voluntarily purchase insurance, can we necessarily claim that it is socially desirable? If others

than the direct consumer value access, shouldn't those others pay? But in developing countries, who (or where) is this alternative source of value and financing?

### **Insurance Demand and Supply Side Cost Sharing**

Type-2 moral hazard is caused by the absence of insurer information on how severe a person's illness is. If such information were available, moral hazard could be perfectly controlled either on the demand side or on the supply side. Demand-side control would come about when insurance takes the form of a fixed dollar indemnity conditional on the state of illness. The insured would receive a payment equal to the cost of care that is optimal (in the sense that the marginal benefit from care just equals its marginal cost) for that level of illness severity; the person would then face the full cost of care for any additional use and would therefore choose to remain at the optimal level. Supply-side control would come about by paying the same amount to a provider of care, who (in a competitive market) would render care of the efficient amount and cost.

When information on ex-post illness severity is imperfect, there is a tradeoff between protection against risk and moral hazard as already shown under demand-side cost sharing. Under supply side cost sharing, there need be no financial risk, but the tradeoff is between moral hazard and "care risk": the risk that the amount of care supplied will not be appropriate to the actual state of illness.

More generally, while sometimes it may be desirable to use only supply side cost sharing,<sup>17</sup> in the general case the insurance for which consumer demand will be strongest in the presence of moral hazard will use a combination of demand and supply side cost sharing.<sup>18</sup> A general theory for this case has not been worked out, but one's intuition is that the mix will depend on which kind of risk—financial or absence of care—is more tolerable.

### **Adverse selection and voluntary insurance markets.**

The other phenomenon that can limit or inhibit the emergence of voluntary insurance is adverse selection: adverse selection occurs when insurers do not set premiums that reflect the information about a consumer's expected expenses that the consumer has. In unregulated markets, this will happen if insureds know more true information about their expected losses than the insurer knows (information asymmetry). It can also happen if insurance purchasers incorrectly *think* they know more than the insurer knows (information distortion). Even if all information is common knowledge, adverse selection can also arise if insurers are not permitted (by law or custom) to set premiums that reflect that information (for example, if insurers are not permitted to set sufficiently higher premiums for higher risks).

Adverse selection often leads to a smaller insurance market than if it were absent, can prevent the existence of a stable market, and might (though rarely) cause the market to disappear entirely. The behavior to be expected depends as well on whether insurers can know the total amount of insurance each person purchases, the conditions for entry into markets, and the degree of foresight insurers are assumed to have.

To see how adverse selection can prevent insurance from emerging, consider the extreme case in which public regulation requires private insurers to charge the same premium for a single “approved” policy to all regardless of risk. Suppose also that there are some consumers who are certain that they will make a large claim under this insurance.

Is there a premium that can cover insurers’ costs that at least some consumers are willing to pay? If insurers set a premium based on the average experience of risks at all levels, and if risk varies fairly substantially, the lower risks will be unwilling to buy insurance at that premium. Insurers would then anticipate that they would sell only to the higher risks, and therefore would have benefits costs higher than the initial “average” premium. But if they raise their premium to cover these higher expected costs, more relatively low risks will drop out. A so-called death spiral can ensue until only those sure or almost sure to incur large losses are left in the market. But because their losses are virtually certain, these people may be unwilling to pay anything more than their actual expenses for the insurance, and so insurers will be unable to cover their administrative costs.

For adverse selection to prevent the emergence of any market whatsoever, there would need to be a non-trivial number of such undetectable high risk users. If the highest risks still have a chance of low or less than average use, even with adverse selection there will still be a market, but it will be confined to insuring the uncertainty of actual losses for those known already to be highest risk.

Sometimes regulators are encouraged to require insurers to charge premiums that do not reflect the risk differences insurers can observe, on the apparently logical ground that the purpose of insurance is to “spread risk.” But the risk voluntary insurance can spread is the risk of poor health that has yet to occur, not the higher risk that results from a chronic condition that has already become evident. Whatever policy and ethical benefits flow from making transfers from lower risks to higher (already-realized) risks, such policies do inhibit the emergence of insurance markets.

If insurers were free to charge higher premiums to higher risks and if insurers knew as much about individuals’ risk levels as insurance purchasers did, the extent of the insurance market under voluntary insurance would be maximized as long as the loading is moderate. The reason is that low risks would be willing to buy insurance at low premiums, while higher risks, who expect higher benefits from insurance, would be willing to pay higher premiums as long as there is still some uncertainty about what their claims will be.

This last point is a subject of enormous confusion in the policy-oriented insurance literature. Sometime analysts conclude (without having actually done the analysis) that, in unregulated competitive insurance markets selling to non-poor people, “private insurers would have an incentive to select only low-risk subscribers” or that “high risk individuals are effectively excluded from the insurance markets due to prohibitively high premiums.” They apparently have a “cream-skimming” story in mind, but simple theory shows that cream skimming cannot exist in competitive insurance markets.<sup>19</sup> It can only exist if regulators require insurers to charge low premiums to high risks that insurers can identify. In the absence of such rules, the lower

risks and the higher risks will both face premiums based on the expected value of their out of pocket payments. In equilibrium, without more, at those premiums it should be equally profitable to sell to low or to high risks. Insurers would not have an “incentive” to select low risks, since the low premiums they could charge would make those risks not especially profitable. Likewise, they would be willing to sell to high risks if premiums charged to those risks would cover the expected value of their expenses.

Compared to some idealized but infeasible situation of mandated insurance with community rating one might be concerned about this market equilibrium. But compared to the relevant alternative of high out of pocket payments, such risk rated insurance permits utility gains by both low and high risks; indeed, under plausible assumptions about risk aversion the high risks might gain more from the opportunity to buy insurance at above average premiums than the low risks would gain from buying coverage at below average premiums. For the great bulk of “high risks,” those who are older or who have some medical condition, the premium will still be affordable if the former out of pocket expense it covers was affordable. High risks will not be excluded from the market by high premiums that cover expenses they would otherwise have had to pay out of pocket. Only the very highest risks whose expected expenses equal the premium because the loss probability is close to one will find insurance with loading unattractive. Other higher than average risk will definitely prefer to pay their risk rated premiums to going uninsured. The individual insurance market in the US which does display risk rated premiums has been found to be more effective in providing coverage to some higher risks (those who are older or who have chronic conditions), controlling for income, than in providing coverage to lower risks. Of course, any profit-maximizing insurer would prefer to select low risks and then

get them to pay the high-risk premium, but in a competitive insurance market where all insurers have the same information on risk this cannot happen.

Finally, if prospective insurance purchasers know more than insurers, there will be some adverse selection—probably not as severe as the extreme case under imposed uniform or “community” rating, but still with some failure to make adequate coverage attractive to lower risks. How severely this will impact the market depends on the elasticity of demand for insurance and on the technology for implementing rate regulation (and controlling the types of policies offered).

So will adverse selection pose an insurmountable problem for voluntary insurance markets in developing countries? The evidence is mixed but surely incomplete. The most rigorous recent evidence suggests problems only when regulation-required community rating compels insurers to ignore information they have. The possibilities for group insurance and for guaranteed renewability provisions in individual contracts are not known. I would be surprised if adverse selection served as an absolute barrier to the emergence of a voluntary insurance market, but it could serve to limit its scope. If regulators choose to impose rating limits, markets may disappear.

### **Cream Skimming and Demand**

*Cream skimming will not occur in competitive unregulated insurance markets.*<sup>20</sup> The reason is that in such markets premiums will adjust to reflect to adjust insurer perceptions of risk.

Premiums will be reduced to attract lower profitable risks, and they will be increased until higher risks become profitable.

If regulation or custom does not permit such premium adjustment, cream skimming can occur. However, if insurers are also required to enroll all who apply, in competitive insurance markets the primary manifestation of cream skimming will be an insurer decision to render more lavish care than is needed to low risks (to attract them) and cutbacks in care below the optimal level to higher risks (to get costs down closer to premiums). In the limit, there may be no financial transfers across risk classes, just inefficiency.

### **Insurance Reserves and Demand**

Commercial insurers promise to make large payments to people who suffer losses in return for the premiums they have already received. How can they, or their customers, be sure that this promise will be kept? As is discussed extensively in Dror and Preker, insurers are appropriately concerned that the benefits they owe may exceed the premiums they collect, and therefore choose to (and are often though not always required by regulators to) assemble “reserves” to cover the cost of aggregate claims in any time period in excess of aggregate premiums collected in that period.<sup>21</sup> How high to set these reserves is not an easy question to answer intuitively, although it does have an easy answer in economic theory (which differs in some important ways from what actuarial theory seems to advise). The dilemma has two parts. (1) In a risky world the maximum possible (if improbable) level of total claims can be quite high relative to total premiums; higher reserves will almost always reduce by a small but positive amount the chance

that there will not be enough premium revenue and assets to cover claims. But the cost of reducing this chance literally to zero would be enormous. So there will be an optimal non-zero probability of default at the optimal level of reserves as long as reserves are costly. (2) But “most of the time,” the chance that claims exceed premiums by a substantial margin is low because of the “law of large numbers.” As long as losses to one policyholder are not highly correlated with losses to other policyholders, which is generally the case in health care, the average claim can come as close as one likes to the average premium if the number of insured exposures grows large enough.

The normative economic solution to this dilemma is straightforward: calculate the marginal cost of sequestering capital so as to add to insurer reserves, calculate the change in the probability that claims will not be fully paid that such an addition will make possible, calculate the value to insureds of this reduction in the risk of non-payment, and set reserves at the level where the marginal expected benefits to risk averse consumers associated with a lower probability of non-payment equals the marginal cost of adding to reserves. (This approach differs from some actuarial models that simply assume some [low] target value for the “probability of ruin.”)

What do these considerations have to do with the demand for health insurance? The first practical point to note is that the relevance of either ruin or reserves to health insurance is generally thought to be considerably less (though by no means zero) than for some other kinds of insurance. The reason is that most health events are independent (my heart attack is unlikely to be correlated with yours) and most health insurance promises benefits only for one year in the future. Thus the need to hold reserves as a large proportion of premiums is generally small in

health insurance. The primary reasons to hold reserves are, first, the possibility of an epidemic (e.g., this winter's flu scare in the US substantially boosted drug claims) and, second, the uncertainty about prices, use, and technology in the next time period independent of the incidence of illness (e.g., the unexpected spread of laparoscopic surgery caused claims to surge). In developed countries, even these examples of non-independence are generally small relative to the value of total claims, but that may not be the case in developing countries. Finally, consumers sometimes seem willing to choose low priced insurers at risky financial status but then profess ignorance and the need for government help when the insurer who charged a premium "too good to be true" either cannot pay claims or exits the market. This is so even though the evidence suggests that more financially reliable insurance does generally claim higher prices. These political considerations may mean that governments will choose to force insurers to have enough reserves to be minimally risky rather than run the risk of having to bail out people who bought from the low premium, low reserve firms. Finally, from a purely paternalistic point of view one might believe that people should not be allowed to buy "risky" insurance—but the alternative to cheap low quality insurance purchasing may for some consumers to remain uninsured altogether. Mandatory insurance purchasing rules solve these kinds of problems (and many others), but may be politically difficult since they impose de facto head taxes if unsubsidized.

What kind of reserves would a private health insurer optionally choose to hold? Imagine that the expected value of benefits per person is  $\beta$  and that the insurer sets a per-person premium of  $\rho = \beta +$  administrative expenses. Then, after these expenses have been paid, an insurer which sells to  $N$  persons will have  $N\beta$  available to pay claims. *If*  $\beta$  was correctly calculated, these

collections should on average be enough to cover actual benefits as long as the benefit levels are independent and  $N$  is reasonably large (say, 10,000 persons). While one of those persons might have a costly expense, in health insurance as opposed to some other kinds of insurance, like liability, the maximum possible expense is relatively small. This assumes that the insurer does not pay for “million-dollar” heroic-measures treatments. The fraction of the insureds who get sick every year will vary, but that variation should be modest. Hence if reserves are costly, their levels ideally should be fairly modest.

The conclusion would change in two circumstances. First, if a large share of total claims was related to infectious disease, then such payouts would be sensitive to the presence of epidemics, and the assumption of independence would not hold. The key issues here are both the infectious nature of the disease and the possibility of response to epidemics, either because the disease is contagious among humans or because it can spread rapidly in response to sudden and unexpected changes in insect or animal hosts.

The other potential problem is if the cost of treatment is subject to unpredictable fluctuations due to changes in prices, wages, or new technology. An unexpected change in drug prices, or the introduction of costly new technology that becomes popular, affect all expenses together, so that pooling within the insured group will not work. Since both of these kinds of risks vary across plans, some reinsurance of the excess risk may be preferable to adding reserves. The choice depends, of course, on the cost of reserves relative to the cost of reinsurance.

## **Group Insurance Demand**

Often the private insurance made available in voluntary insurance markets in developed countries takes the form of group insurance. Insurance is arranged for some group of buyers, who then may have only one insurance plan or at most choose from a small set of plans, rather than having insurance purchased by each individual according to what the individual prefers. Most commonly the group is based on employment at a particular firm, but it may also be based on membership in a labor union, in some other non-governmental organization, or even on residence in a community.

Theory suggests several reasons that demand for insurance is sometimes channeled through a group. First, there are often tax advantages for insurance purchased in this way (usually by excluding some portion of the premium labeled the “employer payment” from taxation). Second, even in the absence of subsidy, group purchase can reduce insurer administrative costs, especially costs incurred for selling and billing. Finally, another reason sometimes suggested as to why group insurance is socially desirable (even if not necessarily desirable to all consumers) is that it “pools risk” across people with different levels of observable expected expense (based on age or the presence of chronic illness) to a greater extent than does individual insurance.

Traded off against these advantages (relative to individual insurance) are the need to settle for the insurance policy or small set of policies chosen by the group, and frequently considerable “lock-in” to the group especially for those who come to be high risks. For example, Madrian showed that higher risk people with employment based group insurance were much less likely than other workers to move to more attractive jobs.<sup>22</sup> In addition, the risk pooling advantage

may not be that strong, both because employers who provide insurance as part of total compensation do seem to vary worker wages inversely (other things equal) with some characteristics related to higher risk (like age) and because individual insurance in unregulated competitive markets typically provides protection against the onset of high risk through guaranteed renewability protection.<sup>23</sup>

The least well-known aspect of group insurance is the “group demand” for insurance. When groups contain people with different insurance demands, what determines what is actually chosen by or for the group? Possible designees for the role of decisionmaker are the average worker, the marginal worker, the worker with the most political influence in the group, and the uninformed employer.<sup>24</sup> Evidence does suggest that, nevertheless, group insurance purchases do often match characteristics of workers: groups with higher income workers with larger families choose more generous insurance coverage. Moreover, workers move across groups based in part on the insurance offered. Finally, more heterogeneous groups are more likely to offer multiple insurance options than those where all workers are similar in demand characteristics.<sup>25</sup> On the other hand, unionization does seem to be associated with higher insurance demand, given worker characteristics.<sup>26</sup>

### **The Effect of Insurance Subsidies on Demand**

Demand for insurance seems to be responsive to the presence of subsidies. Most of this responsiveness seems to occur at the group level, rather than at the level of the individual

worker.<sup>27</sup> However, there is a wide range of estimates of demand elasticities in developed countries, and little evidence for less developed countries.

### **The Demand for Protection against Risk Reclassification**

One of the characteristics of individual market insurance in a static world is that premiums charged to any person will reflect what the insurer knows about that person's level of risk, in the sense of expected expenses. From a policy perspective this kind of risk rating, however helpful it is to efficiency and the emergence of markets, is troublesome. One reason for concern is a normative judgment that there should be transfers from low risks to high risk. Even though the efficient vehicle for making such transfers is the use of formal public tax and transfer programs not related to insurance premiums, or through the use of risk adjustment in any public subsidies, there is a political temptation to favor uniform insurance premiums. But another reason for concern on the part of a currently low risk consumer is the desire to be protected against substantial increases in future or lifetime premiums if a chronic condition which results in high expected expenses over multiple periods were to occur. It turns out that most competitive individual insurers in developed countries offer protection against this risk in the form of "guaranteed renewability" provisions in the insurance policy.<sup>28</sup> These provisions commit the insurer to charging the same premium to someone who becomes a high risk as to others in the person's initial risk pool; that is, the insurer agrees not to charge discriminatory premiums based on the person's post-purchase health experience. There is evidence that actual market premiums are consistent with the functioning of guaranteed renewability.

More formally, it appears that consumers do demand to be protected against subsequent increases in premiums based on their own health experience. It is much more difficult for insurers to protect them against market-wide reasons for premium increases, such as increase in prices or costly new technology.

### **New Technology, Cost Containment, and Insurance Demand**

One characteristic of voluntary market insurance in developed countries, relative to many public insurance plans, is that market insurance has been more accommodating to the introduction of beneficial but costly new technology, and therefore less cost containing. To some extent, it appears that the “failure” of cost containment is not a defect as far as consumers are concerned, but rather the price they are willing if not eager to pay for less restrictive supply side rationing than occurs in public insurance plans. Nevertheless, an open question that remains is whether voluntary competitive insurers will cover costly technology in an efficient fashion.

Pauly has recently argued that, as long as competitive insurers are free to refuse to cover new technology and the market for medical services is also competitive, they will never add coverage that makes consumers worse off.<sup>29</sup> The empirical examples of what appear to be inefficient technological “arms races,” such as occurred in the US in the 1960s and 70s, appear to be associated with state regulations forbidding insurers from denying coverage or contracting selectively with providers that did not provide expensive technology. However, the role of markets and the impact of new technology on the demand for coverage need to be investigated further.

## **Other Reasons for Non-Purchase or Market Failure**

Economic theory identifies low risk aversion, moral hazard, and adverse selection as the primary reasons for no or low insurance demand in competitive markets. Are there other reasons why demand may be low, especially reasons germane to developing countries?

*Consumer information.* If consumers have incomplete or incorrect information about the distribution of expected losses, demand for insurance at premiums insurers require may fail to materialize. The most obvious possibility here is that consumers may underestimate ex ante the chance of an illness with relatively expensive treatment. Such beliefs of the “it can’t happen to me” variety do exist, and sometimes if the loss probability is in a low enough range, it may be rational for consumers to fail to obtain correct information.<sup>30</sup> More generally, culturally conditioned beliefs about the future or even high interest rates can lead to a kind of myopia in which severe losses are not anticipated by consumers and therefore not insured.

If closing the gaps in knowledge seems in principle attractive, the first question is whether there exists or could exist adequate data to develop estimates of illness probabilities (defined not just by the existence of illness but by its severity or other proxies for effectiveness of treatment).

Better data is almost always better, but we do not want to overemphasize the importance of perfect knowledge. Consumers surely must develop some subjective estimate of illness probability, which they can then update in a Bayesian way if better information becomes available. If insurance is supplied by commercial firms, they also will have some estimate of

expected losses (even if ambiguous). Theory and available empirical evidence suggest that ambiguity about probability is not *necessarily* a barrier to insurance demand (or the emergence of markets) as long as consumers' subjective estimates are above those of insurers, and as long as consumers are sufficiently risk averse (or protection-seeking) to pay enough to cover any amount insurers might add as a hedge against ambiguity.<sup>31</sup> It all depends on the facts; imperfect information does not necessarily make losses uninsurable.

This conclusion is strengthened if we allow for the possibility of mutual insurance.<sup>32</sup> Consider a simple model in which there is poor data on loss probability associated with treatment of some illness (e.g., stroke), and assume further that consumers differ in how likely they think the chance of an illness is. Insurance can still emerge if consumers agree that, whatever the loss probability is, it is the same or very similar for all households in a community and that there is not high correlation between losses. Then the solution is mutual insurance; in its simplest form, all households agree to share the cost of treatment for those who become ill. In this arrangement, the "premium" is the person's estimate of his household's share of the *ex post* cost. Setting aside transactions costs, making such a payment will always be preferred to risking high out of pocket expense. In a sense, those who think illness is unlikely will still join the pool because they expect their *ex post* share to be low, while those who think illness is likely but not certain will prefer to pay a relatively high premium to running the risk of an even higher out of pocket payment.

*Political Limits.* It is possible that the kind of insurance for which voluntary consumer demand would exist is politically unacceptable. "Acceptability" obviously depends on the nature of a

country's political system and the distribution of political power and private influence.

(Lobbying by medical professional associations is especially common here.)

Begin by assuming that money income in a country is and remains distributed unequally, but that (tautologically) there is no effective political consensus to redistribute it to any greater extent than is currently done. The willingness to pay out of pocket for medical care will also be distributed unequally, even among households at the same level of health somehow defined. Generally medical spending will vary positively with income (and often other socioeconomic factors like education); it therefore follows that transforming that spending into voluntary private insurance coverage will also lead to an uneven distribution of actual insurance and, under risk rating, variation in premiums for the same nominal coverage because the cost and value of benefits are greater to higher risks and richer people.

While variation across people in terms of spending for many goods and services may be politically acceptable (especially if the uneven distribution of income is acceptable), the similar variation in health insurance coverage may be regarded as undesirable based on different views about what constitutes "equity" or fairness. While many people have strong views on equity, and many also think it is important in health care even if the distribution of income and other types of consumption remain quite unequal, they often do not agree on what is fair. More to the point, clashing views on the importance of equity (and efficiency) can cause opposition to private insurance markets, often precisely because they make the inequality already inherent in a society much more obvious. Insurance markets offer a temptation to politicians and advocates to use health insurance pricing to redistribute income across income levels, or from the healthy to

the sickly; these efforts, however praiseworthy they might be on ethical or esthetic grounds, can impede the emergence and functioning of private insurance markets that best satisfy private demand.

More generally, health insurance and health care are both common objects of taxation and regulation. Sometimes the politics of this taxation and regulation can be counter-productive. For example, regulators sometimes require insurance to cover certain medical services, or to hold very large reserves, both of which can make the price of insurance too high for many people to be willing to buy it. It might be more desirable to have many people with incomplete insurance than to have a few people with generous insurance. Not all health insurance is government licensed or regulated, however. For example, many large employers provide insurance to their workers that the employers “self-insure.”

*Distrust of insurers.* It is alleged that consumers sometimes mistrust insurers when there has been a history of default.<sup>33</sup> We do know that insurers of all types with less financial stability are penalized in terms of lower premiums they can charge or smaller markets shares. It is clear that some kind of structure that will reassure consumers who pay premiums that they can collect claims without excessive delay and bother is important in establishing a functioning insurance market. Setting up insurers under the auspices of other trusted social institutions, such as hospitals, labor unions, or trade associations, can help. There is no well-developed economic theory of trust.

*Lack of competition.* Given some demand for insurance on the part of a population, insurance will be demanded in a large quantity the lower the price (in the sense of administrative costs and profit markup). Evidence does exist that limited competition among insurers can lead to higher prices. Even if insurers are non-profit, the absence of competition can lead to excessively high administrative costs. The possibility of market power cannot, however, be a reason for insurance to fail to emerge at all, since even a profit-maximizing monopoly must set a price low enough that it can sell some product.

### **Applying These Concepts to Demand for Health Insurance in Developing Countries**

*The Macro Sectoral Allocation Problem.* In any country, health insurance can be supplied either as voluntarily purchased insurance or as government subsidization of the cost of medical care. Public provision is generally financed via taxation; private provision is generally financed by voluntary payments, sometimes as individual insurance purchasers and sometimes as employment-based group insurance. In all of these three settings, citizen demand for risk protection is presumably a common motivation. Public insurance demand may additionally be motivated by externalities, either the technological ones associated with uninsurance coverage of care for contagious disease or the altruistic and paternalistic motivation that reflects consensus about the health and health care of fellow citizens.

The tax bases used in almost all countries also tend to redistribute income as well as publicly purchased goods. In doing so, they also distort economic behavior, and the greater the level of

this distortion, the smaller (other things equal) the demand for goods to be provided through the public sector.

Employment-based insurance chosen on a voluntary basis responds to worker valuations of such coverage as well as to the magnitude of administrative cost savings associated with group insurance. The precise connection between the insurance profit-maximizing employers will choose to provide and the insurance demand of heterogeneous workforces is unclear.

Which of the three methods will be chosen will depend on the relative costs and benefits of each. Compared to either public provision or group insurance, individual insurance can allow each person to get exactly the insurance demanded, but the administrative cost will be high. In contrast, group insurance will generally have lower administrative cost but less perfect tailoring to individual desires.

Public provision also tends to have low explicit administrative cost, but (in contrast to either form of private insurance) the use of the tax system generates economic distortion or “excess burden.” This observation may be especially relevant for developing countries with poorly administered tax systems. In such countries it will be rational to limit the amount of insurance furnished through the public sector. The reason is not that incomes are low, but that tax-collected funds are costly and therefore scarce. In this sense, emergence of private insurance, which can be less costly, is not just an undesirable but unavoidable alternative to public provision; it may be the desirable instrument to use when the public sector is too costly to be efficient.

A strategy that should be considered as an alternative to full public provision is subsidization of private insurance. This can be a way to tap private willingness to pay and yet still achieve equity and efficiency goals of the public sector. *If the public subsidy program could be appropriately designed*, in general it would dominate any program in which the government provides fully paid insurance (public or private). The intuition behind this result is straightforward: since insurance is bound to be worth something to citizens (even if not enough to cover its full cost), inducing people to make private payments to match public subsidies should always be possible. At a minimum, these private payments would lower the need for administratively costly public funds. Whether they also lower the excess burden of taxation depends on whether redistributive or other reasons for tax disincentives can be reduced. If, for example, political constraints require public funding to be redistributive in a way that deters work effort, then greater use of tax funds will cause more distortion. Of course, one could reduce some other redistributive taxes to offset any higher taxes to finance public insurance, but this may prove politically difficult.

Will moral hazard and/or adverse selection impede an effort to convert the out-of-pocket payment which exists into private insurance? This seems to be the most important possible source of concern. We can address this question either in the context of the conventional normative economic model that attributes to government a desire for economic efficiency, or in the context of a positive model of government in which political pressures and rent-seeking motivations may prompt regulation and control over private insurance.

In the first (normative) context, the use of conventional tools of deductibles and coinsurance in indemnity type insurance should be sufficient to *permit* moral hazard to be controlled well enough to permit a market to emerge. As long as spending is verifiable, potential coverage should be possible. It may even be possible and desirable to implement “true indemnity” coverage in which payment depends only on the evidence of the existence of a treatable illness (like a fracture), and need not require information on actual spending.

We begin by assuming that the entity that implements collective choice (“the government”) is under no political or economic constraints. It can reallocate the population’s total resources instantly; it can levy non-distortive taxes on precisely those households it wishes to tax; it can allocate subsidized services in a minimally constrained way; and it can impose whatever out-of-pocket payments for medical care that it chooses. (This last power is somewhat superfluous if the second one is present, but we will assume it for the present.)

A government with such power would be able to implement the allocational and distributional objectives described by Musgrave.<sup>34</sup> Its “Allocation Branch” would choose the level of medical services by comparing marginal benefit from care to each person with a given illness and other characteristics to marginal service cost. Then the “Distributional Branch” would choose how to pay for this allocation in the way that satisfies the society’s distributional objectives, imposing nondistortive taxes and transfers from private income where necessary.

If this model approximates reality, then we should not find people paying such high levels of out-of-pocket payment as to exclude them from appropriate care. This powerful government

would see that they get what they need. In the ideal scenario, it would have zero out-of-pocket payments for all risk averse people, and use its hypothesized powers of control to prevent moral hazard.

The relevant marginal benefit to be considered by this government would have two parts. One part would represent the value of medical care to the household receiving it; if households differ in their values because of tastes, these differences would be taken into account and there would be no necessary implication that use should be completely uniform. The other component of marginal benefit reflects potentially positive externalities for others in society of a given household's greater use of medical care. If the care reduces contagious disease, this external benefit should be obvious. But even if the illness that might be effectively treated is not contagious, altruistic and humanitarian motives might cause others to attach positive value to the relief of suffering.

In this model, there would be no role for the "solidarity principle" applied to medical care. That principle has no rigorously precise interpretation, but its general idea is that people should have access to medical care according to their marginal benefit ("need") and pay based on some assumed ability to pay. The first part of this principle is still carried over into the ideal outcome, but the second part becomes superfluous or harmful. If the general tax and transfer system is doing an adequate job of achieving the desired redistribution, there is no need to tailor particular taxes, even ones that may to some extent be earmarked for health care, to achieve distributional objectives. In other words, there is no need for each tax to meet distributional or equity goals; all that is required is that the package of taxes do so.

For example, if a country decided to fund health insurance for the non-poor with a uniform premium, that would not necessarily limit an ability-to-pay or rich-to-poor transfer. Compared to a country that used a proportional wage tax, this country would just have to have greater progressivity in its other taxes. If not all taxes are identical in terms of targeting or excess burden, the choice will be complex and it is even possible, some have agreed<sup>35</sup> that distributional objectives can be better achieved through the health tax and spending budget. But the main point is that, either way, there is no basis for a breakeven system operated according to the solidarity principle. And if specific health care funding is supplemented with general revenues, as is often the case,<sup>36</sup> then the overall distributional pattern is affected at the margin by the general revenue taxes: the health insurance tax is irrelevant.

Obviously it is not realistic to assume that actual governments can or will choose to do what welfare economics says. One reason for deviation has already been suggested: different practical taxes have different efficiency or excess burden implications in addition to their distributional goals. So next, I assume that efficient taxation is not possible, but continue to assume government makes the aggregate welfare maximizing allocation given the cost it faces.

If a tax causes economic distortion, it imposes two limits on the provision of social goods. Most obviously, there will be limits on the amount of funding that can be received at any tax rate; make the tax rate too high, and less money may be collected than at a lower tax rate. So if the tax on the base available to the country is set at the revenue maximizing level, but that amount is less than the cost of the ideal levels of social goods, there will be undersupply of social goods. More

generally, the theory of excess burden tells us that the economic cost of transferring resources from the private to the public sector is higher when taxes are distortive. In effect the cost of spending  $X$  raised through taxation is more than  $X$ . This excess burden raises the cost of providing social goods, and the efficient response is to provide less of them.

In such a case, we can have a situation where the government chooses a level of provision of some social goods like medical care that falls far short of the level at which the marginal private benefit for such goods equals their marginal resource costs because of the “tax on a tax” character of excess burden. If some alternative method to fund that spending exists (such as private purchase or private insurance), we might expect such purchases to emerge, and the resulting mixed system would be efficient.

However, the pattern of such additional private purchases would not replicate the pattern of the government program. The government program in principle takes account of both private benefits from health care and external or social benefits. But the supplemental private purchases take account only of private benefits, but will have a less strict budget constraint. As a result it is obvious that only those with high private benefits—those with strong demands for health care or insurance—will engage in such supplementary purchases.

Some of the discussion of actual systems imagines that there is a new tax base that can be tapped. But if a new tax were feasible, the question arises of why it was not already used to finance the basic (and chronically underfunded) pre-existing social system. “Political Feasibility” is an all-purpose but possibly true answer. It may be that, in contrast to our

assumptions to this point, politics inhibited provision of services whose marginal benefits exceeded their marginal tax cost, but that restructuring proposals for additional spending in the form of new insurance (rather than national social insurance or general public spending) will garner more political support. While this is possible, it should not be taken for granted.

### **Health Insurance, Income, and Demand**

Another possible motivation for the purchase of health insurance has recently been suggested by John Nyman, although this hypothesis has yet to win widespread approval. But it may be especially applicable in this context.<sup>37</sup>

The standard theory of insurance, already described above, envisions it as protecting financial wealth. People, in this theory, buy insurance to cushion the financial blow of the cost of care they would have purchased even in the absence of insurance. In the case of medical services, the threats to wealth are the large out-of-pocket payments previously discussed. But suppose that there is treatment that the consumer knows she might need for a life-threatening illness whose cost is greater than her financial wealth, and greater even than the present discounted value of her lifetime earnings or consumption. We will not observe spending on this treatment for this person in the absence of insurance. But if the person attaches enough value to survival, above and beyond any productivity effects, Nyman argues that she may be willing to pay the premium for health insurance which (in effect) “buys” this survival. If so, insurance will be associated with more spending than in its absence, but this increase in spending will *not* be inefficient moral

hazard. As Nyman explains, the increase in spending here is really an income effect resulting from the higher wealth insurance creates in the “very sick” state.

We do not know how important this motivation is in developing countries. It depends on the form of the person’s utility function for survival and the reason for the demand for medical care. Here is a helpful way to think about this problem in comparative terms. Consider the demand curves for treatment of a given illness of people at different income levels. Plausibly, but perhaps in contrast to the Rand results, assume that higher income people always buy more care at various user prices, that the intercept at which some care is bought rises with income, but that the effect of income on quantity demanded is larger at higher user prices than at lower user prices. The important thing to note is that, at any user price, lower income people may have more elastic demand curves than higher income people. Under the conventional view, this means that, other things equal, lower income people will prefer insurance with higher levels of coinsurance because their demand displays more moral hazard. But this should not be the case if the larger increase in demand proceeds from the basis suggested by Nyman. Indeed, a test of his hypothesis would be provided by examining empirically the relationship between income and insurance demand in unregulated unsubsidized markets. His theory might even suggest that lower income people are more likely to demand comprehensive insurance (to help them “afford” costly care) than would higher income people who could more easily pay out of pocket.

According to Nyman, the willingness to pay a premium in excess of the expected value of no-insurance spending can be quite high, much higher than what we would attribute to risk aversion alone. If this hypothesis is true, it implies a substantial demand for coverage of expensive but

highly effective (life saving) treatments even by consumers of moderate wealth. More importantly, it means that the increase in spending associated with such coverage might not represent inefficient moral hazard. At present, both the positive and normative analysis of these cases in developed countries is very incomplete. More work needs to be done on the relationship between income and wealth and the demand for both medical care and insurance.

---

<sup>1</sup> Levit K, et al. Health Spending Rebound Continues in 2002. *Health Affairs* (January/February 2004): 23(1); 147-159.

<sup>2</sup> Cutler DM, Zeckhauser RJ. The Anatomy of Health Insurance. In AJ Culyer and JP Newhouse, eds., *Handbook of Health Economics, Volume 1A* (Amsterdam: Elsevier Science B.V., 2000), 563-643; 568.

<sup>3</sup> Vate M, Dror DM. To Insure or Not to Insure? Reflections on the Limits of Insurability. In *Social Reinsurance: A New Approach to Sustainable Community Health Financing*, Dror DM, Preker AS, eds. (Washington, D.C.: The World Bank, 2002): 125-152.

<sup>4</sup> Nyman JA. The Value of Health Insurance: The Access Motive. *Journal of Health Economics* (April 1999); 18(2): 141-152.

<sup>5</sup> Vate and Dror, 2002.

<sup>6</sup> Gertler P, Sturm R. Private Health Insurance and Public Expenditures in Jamaica. *Journal of Econometrics* (Amsterdam: March, 1997); 77(1): 237.

<sup>7</sup> Arhin D. "Health Insurance in Sub-Saharan Africa: What are the Options?" Paper presented at Symposium on Health Care Financing at The European Conference on Tropical Medicine, Hamburg, 1995.

<sup>8</sup> **Wagstaff and Pradhan, 2003.**

<sup>9</sup> Arhin D. "Health Insurance Demand in Ghana: A Contingent Valuation." Paper prepared for IHEA Conference, Vancouver, 1996.

<sup>10</sup> Brown W, Churchill C. Insurance Provision in Low-Income Countries. (Calmeadow: Microenterprises Best Practice Project, Development Alternatives, Inc., 2000.)

<sup>11</sup> Pauly MV, Nichols LM. The Non-Group Health Insurance Market: Short on Facts, Long on Opinions and Policy Disputes. *Health Affairs Web Exclusive* (October 23, 2002) <http://www.healthaffairs.org/WebExclusives/2106Pauly.pdf>, October 23, 2002.

<sup>12</sup> Arhin-Tenkorang D. "Experience of Community Health in the African Region." Health Finance for Poor People: Resource Mobilization and Risk-Sharing, Alexander Preker and Guy Carrin, eds. (Washington: World Bank 2005), pp. 157-198

<sup>13</sup> Pauly MV. The Economics of Moral Hazard. *American Economic Review* (June, 1968); 58, 531-537.

<sup>14</sup> Shavell S. On Moral Hazard and Insurance. *Quarterly Journal of Economics* (November 1979); 93(4): 541-562.

<sup>15</sup> Cawley J, Philipson T. An Empirical Examination of Information Barriers to Trade in Insurance. *American Economic Review* (September 1999); 89(4): 827-846.

<sup>16</sup> Schlesinger H, Doherty N. Severity Risk and the Adverse Selection of Frequency Risk. *Journal of Risk and Insurance*, (December 1995); 62 (4): 649-665.

<sup>17</sup> Randall EP, McGuire TG. Supply-Side and Demand-Side Cost Sharing in Health Care. *Journal of Economic Perspectives* (Fall 1993); 7(4): 135-151.

<sup>18</sup> Pauly MV, Ramsey SD. Would You Like Suspenders to Go with That Belt? An Analysis of Optimal Combinations of Cost Sharing and Managed Care. *Journal of Health Economics* (August 1999); 18(4): 443-458.

- 
- <sup>19</sup> Pauly MV. Is Cream Skimming a Problem for the Competitive Medical Market? *Journal of Health Economics* (April, 1984); 3(1):88-95.
- <sup>20</sup> Pauly, *JHE*, 1984.
- <sup>21</sup> Dror DM, Preker AS. *Social Reinsurance: A New Approach to Sustainable Community Health Financing* (Washington, D.C.: The World Bank, 2002).
- <sup>22</sup> Madrian BC. Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock? *The Quarterly Journal of Economics* (February 1994); 109(1): 27-54.
- <sup>23</sup> Pauly MV, Herring B. Pooling Health Insurance Risks. (Washington: American Enterprise Institute, 1999.)
- <sup>24</sup> Pauly MV, Goldstein GS. Group Health Insurance as a Local Public Good. *The Role of Health Insurance in the Health Services Sector*, R. Rosett, ed., (New York: National Bureau of Economic Research), 1976.
- <sup>25</sup> Bundorf MK. Employee Demand for Health Insurance and Employer Health Plan Choices. *Journal of Health Economics* (January 2002); 21(1): 65-88.
- <sup>26</sup> Herring B, Pauly MV. Incentive-Compatible Guaranteed Renewable Health Insurance. NBER Working Paper 9888, July 2003. <http://www.nber.org/papers/w9888>, September 10, 2003.
- <sup>27</sup> Washington E, Gruber J. Subsidies to Employee Health Insurance Premiums and the Health Insurance Market. *National Bureau of Economic Research Working Paper 9567*, March 2003, <http://dsl.nber.org/papers/w9567.pdf>, 28 January 2004.
- <sup>28</sup> Herring and Pauly, 2003.
- <sup>29</sup> Pauly MV. Market Insurance, Public Insurance, and the Rate of Technological Change in Medical Care. *The Geneva Papers on Insurance and Risk* (April 2003); 28(2): 180-193.
- <sup>30</sup> Pauly MV, Kunreuther H. Neglecting Disaster: Why Don't People Insure Against Large Losses? *The Journal of Risk and Uncertainty* (January 2004); 28(1): 5-21.
- <sup>31</sup> Kunreuther and Pauly, 2004.
- <sup>32</sup> Pauly MV, Kunreuther H, Vaupel J. Public Protection against Misperceived Risks: Insights from Positive Political Economy. *Public Choice* (May, 1984); 43(1): 45-64.; Doherty N. The Design of Insurance Contracts When Liability Rules are Unstable. *Journal of Risk and Insurance* (June 1991); 58(2): 227-246.
- <sup>33</sup> Weber A. Insurance and Market Failure at the Microinsurance Level. In *Social Reinsurance: A New Approach to Sustainable Community Health Financing*, Dror DM, Preker AS, eds. (Washington, D.C.: The World Bank, 2002): 203-222.
- <sup>34</sup> Musgrave RA. The Theory of Public Finance. (New York: Prentice Hall, 1959.)
- <sup>35</sup> Besley T, Coate S. Public Provision of Private Goods and the Redistribution of. *The American Economic Review* (Nashville: September, 1991); 81(4): 979 (6 pages).
- <sup>36</sup> **Croatia, 2002.**
- <sup>37</sup> Nyman, 1999.