JOURNAL OF ECONOMIC THEORY 2, 10–26 (1970)

Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives*

RICHARD ZECKHAUSER

Department of Economics, Harvard University, Cambridge, Massachusetts 02138

Received September 9, 1969

The primary purpose of medical insurance is to spread a risk, the risk of incurring substantial medical expenses.[1] With risk spreading, individuals will not pay the full amounts of such expenses. Insurance provision will thus introduce a perverse incentive toward overexpenditure if, as usually is the case,

(1) the insureds have substantial influence over the amount that is spent on their own behalf in any particular medical circumstance, and
(2) the level of reimbursement by the insurance plan is a positively associated function of the expenses incurred by its insureds.

This is one example of a frequently occurring situation in which no practicable market structure will simultaneously produce optimal risk-spreading and appropriate incentives for individual action. The best that can be done, as we would suspect, is to find a happy compromise with some risk-spreading and some incentive.[2]

* I am greatly indebted to Mr. Milton Weinstein for his most capable research assistance. Prof. Kenneth Arrow gave me perceptive comments. This research was sponsored in part by a National Science Foundation grant, GS1537, to the Harvard Institute of Economic Research.

[1] Medical insurance might well be employed as an integral part of a redistribution scheme. The argument that transfers in kind are less efficient than lump sum transfers may be invalidated if externalities enter the arena. In our society at least, it seems that we are more concerned about the medical services received by the poor than we are about many other elements of their consumption. This gets into sticky questions involving economic values for lives and deaths, and the use of the market mechanism for regulating decisions in this area. These questions are avoided in this paper because it is assumed that all citizens are identical and have identical levels of assets.

[2] See the articles by Arrow [1, 2] and Pauly [3]. It is not only because individuals' utility functions for money display risk aversion that we are concerned with insurance. Without adequate insurance protection most citizens would be unable to afford needed medical care in some extreme situations. Medical insurance may mean the difference between life and death.

Part of this conflict in objectives can be overcome by having an insurance plan which specifies particular levels of reimbursement for particular expenses associated with particular conditions. But it seems unlikely that any prespecification procedure will remove from the hands of the insured all discretion with respect to all expenditures which affect costs to the insurance plan.

The Model

For simplicity, assume that all individuals are identical with respect to preferences, levels of assets, and susceptibilities to different medical ailments. The object is to find the insurance plan that maximizes individuals' expected utilities. The possible outcomes for each individual are the n mutually exclusive, possible medical conditions. For each particular condition, i, an individual will have a utility function, $u_i(x, w)$, that has two arguments: the level of medical expenditure on his behalf, x, and his wealth, w.[3] It is assumed that $u_i(x, w)$ is a positively valued function of both its arguments. If the values of these utility functions are to be combined in the usual fashion to give expected utilities, it is necessary that the utility function for each condition be a cardinal function of the von Neumann–Morgenstern variety and that the utilities be calibrated across as well as within conditions.[4]

The insurance plan will be represented by the sharing function $g(x)$ which gives the portion of his medical expenditure paid for by the individual.[5] The insurance plan pays $x - g(x)$. For insurance coverage each individual pays a premium, P. His wealth after medical expenditure will equal his initial wealth, w_0, less the sum of the premium he pays for medical insurance and the share of his medical expenditure he must pay himself; that is, $w = w_0 - P - g(x)$.

Given the sharing function and insurance premium, the individual in

[3] The underlying utility function may be one for wealth and health. The derivative function using medical expenditure rather than health as an argument would be linked through the production function for health in the particular condition. Uncertainty as to the effectiveness of medical care would introduce a stochastic element into the production relationship.

In this paper money magnitudes are measured in dollar units. For convenience dollar signs are usually omitted.

[4] There must be consistency in the unit intervals for the utility functions associated with the different conditions, but absolute utility levels need not be comparable. What will be required, for example, will be comparisons between utility increases from medical expenditures on cancer and from those on appendicitis.

[5] The $g(x)$ function need not be constrained to be nonnegative. A plan may give cash payments over and above medical expenses should particular conditions occur. No such plan is optimal in the examples considered here.

condition i will choose the x, call it x_i, which maximizes his utility function for that condition. This relationship is indicated

$$x_i \text{ maximizes } u_i(x, w_0 - P - g(x)). \tag{1}$$

Let p_i represent the probability that the individual meets condition i, that he has the utility function u_i. His expected utility, U, for a particular insurance plan as represented by its sharing function and premium is

$$U = \sum_i p_i u_i(x_i, w_0 - P - g(x_i)). \tag{2}$$

The collected premiums must equal the cost of the plan. This means that for each of the identical individuals, the expected cost he imposes upon the plan must be equal to the premium he pays, assuming of course that there are no administrative costs. This constraint can be written

$$P = \sum_i p_i(x_i - g(x_i)). \tag{3}$$

The optimal insurance plan is represented by the function $g(x)$ that maximizes (2) subject to (3), given that the w_i's are selected as specified in (1).

The Question of Numbers

There might be a bit of confusion on the method of calculation to select x_i given the constraint on P that is presented in (3). In choosing his x_i no individual will consider its effect on P. [P will be taken as a constant when he maximizes expression (1)]. The reason is that there are very large numbers of people in the model, and the influence of the plan's expenditures on his behalf on the size of P will be negligible.

The large population supports the assumption that the percentage of people in a particular condition will equal the expected percentage. The probability that the actual percentage will get arbitrarily close to the expected one goes to 1 as the size of the population goes to infinity, assuming that the random devices that pick the conditions for the different individuals are independent of each other. This independence assumption may not be valid if, for example, epidemics are a possibility. To allow for such nonindependent cases we must consider the constraint on P be that total premiums equal expected cost, but not necessarily actual cost.

Possible Insurance Plans

Insurance plans may differ on particulars. The simplest plan would pay no attention to medical conditions; it would offer the same sharing function regardless of the insured's condition. At the opposite extreme would be an insurance plan which offered a different sharing function for each possible condition. Such a plan would require that the administering authority be able to distinguish perfectly among all conditions. Such exceptional powers of discrimination might not be available. In between these two extremes would be a plan which allowed the insuring authority to discriminate to some extent amongst different medical conditions. The conditions would be placed in aggregative categories. There would be a different sharing function for each category. We will call plans which offer no discrimination Case I plans—those with perfect discrimination Case II plans—those with limited discrimination Case III plans.

An Example with Three Medical Conditions

To illustrate the procedure for selecting an optimal insurance plan, consider a simple case with three possible conditions: perfect health, appendicitis, cancer. The respective probabilities of occurrence for the three conditions are .94, .05, .01. When the individual is in perfect health, medical expenditure in his behalf will do nothing to raise his utility. When he has appendicities, his utility will be vastly improved by medical expenditures that are sufficient to pay for a routine appendectomy. Beyond that point, extra expenditure gives him only a few low priority frills, things like a private room.

With cancer the incentive problem looms large once again, but the risk-spreading aspects of insurance are perhaps most important. If one is paying for one's own cancer treatment, income effects can substantially reduce the amount of desirable expenditure.[6]

The Utility Functions

The three utility functions were chosen to approximate the properties described above. The conditions are indexed: perfect health, $i = 1$; appendicitis, $i = 2$; cancer, $i = 3$. The utility functions are given by the general formula

$$u_i(x, w) = -c_i e^{-d_i x} - e^{-tw}. \tag{4}$$

The specific values of the parameters are $c_1 = 0$, $c_2 = .08$, $c_3 = .25$, $d_1 = 1$, $d_2 = .03$, $d_3 = .008$, and $t = .002$.

[6] Ask a man who has one chance in 100 of contracting cancer next period how much of a premium he would like to pay now in return for 100 times that amount to spend on treatment should he contract cancer next period. His answer will likely be well in excess of 1% of the total he would spend of his own funds had he contracted cancer and had no insurance. Indeed, the amount of his preferred premium might well be a few per cent of his total wealth in which case it would be quite impossible for him to expend 100 times that amount on his own.

The separable and additive form of the utility functions has no particular significance except that it facilitates computation. The indifference curves relating the two arguments will have the usual shape. There will be no Giffen goods.

The qualitative aspects of these utility functions perhaps can best be understood by seeing how much an individual would spend given various insurance sharing functions. For ease of computation in this paper, we consider only linear $g(x)$ functions, that is functions of the form $g(x) = a + bx$. Table 1 gives the amounts that an individual with a particular condition will spend on medical treatment depending upon the particular cost-sharing insurance scheme. In drawing up this table a has been omitted. It only affects total expenditure through relatively minor income effects.[7]

TABLE I
Optimal Expenditures for Different b's
Wealth after Premium, $w_0 - P$, Equals 10,000

	$b = 1$	$b = 1/2$	$b = 1/6$	$b = 1/31$
Appendicitis, $i = 2$				
Own Expenditure	630.70	336.70	120.74	25.34
Total Expenditure	630.70	673.40	724.42	785.52
Cancer, $i = 3$				
Own Expenditure	6,320.51	5,108.49	2,866.06	788.49
Total Expenditure	6,320.51	10,216.98	17,196.33	24,443.04

The exponential form of the additive portions of the utility functions leads to reasonable behavior. As b decreases, as the percentage of his own costs an individual must pay goes down, the amount that he will spend on his own behalf may increase or decrease. However, beyond a point it is sure to decrease. The total expenditure on health that the individual will choose will always increase with decreases in b.[8]

[7] By way of illustration, with $a = 100$ the total expenditures on appendicitis and cancer respectively would be: $b = 1$, 636.95 and 6391.93; $b = 1/2$, 676.63 and 10,272.53; $b = 1/6$, 725.52 and 17,225.74; $b = 1/31$, 785.74 and 24,450.50.

[8] If a log function is used as the money portion of the utility function, with an exponential function for the medical care part, it is easier to get utility functions that lead to more spectacular and diverse changes in the insured's expenditure of own resources in response to changes in b. If log functions are used for both portions of the utility function, it turns out that the individuals' expenditure of own resources is directly proportional to b, and that total expenditure is independent of b. This does not seem to be realistic behavior to expect in general.

CASE I: Reimbursement from Insurance Solely a Function of Expenses to the Insured

Case I, we remember, offers the same sharing function regardless of medical condition, including the condition "health." The cost when a member incurs costs in any medical circumstance is solely a function of the total expenses of the insured. The individual is reimbursed the same amount when he spends \$100 on cosmetic surgery as when he spends \$100 to have a tumor removed.

Consider our example with initial wealth, w_0, equal to 10,000. The optimal sharing scheme turns out to be

$$g(x) = .03367x.$$

The expenditures under this scheme will be,

$$x_2 = 766.33, \qquad g(x_2) = 25.80, \qquad \text{and}$$
$$x_3 = 23{,}700.57, \qquad g(x_3) = 798.00.$$

The premium will be 266.05.

It is easily verified that each sharing scheme is consistent with one and only one level of premium. The relationship between the two is somewhat complex. The x_i's, which in effect determine the premium, depend not only upon the sharing scheme, but also upon the premium itself. The optimal sharing scheme thus defines a set of x_i's, and a consistent premium.

It will always be possible to have $a = 0$ in an optimal Case I plan. This is intuitively obvious upon reflection. A negative a is just a lump sum payment from the insurance plan to the individual irrespective of his condition. A positive a, conversely, is a required fixed payment by the individual to the plan. All other variables would take on the same value if these negative or positive a's were replaced by an a equal to 0 and if the original value of a were added to the premium.

Marginal Conditions at the Optimum

To determine the marginal conditions for the optimal insurance plan, we must first consider the maximizing response by the individual as set out in (1). Remember $g(x)$ is of the form $a + bx$. Following the traditional maximization methods of calculus, the individual in condition i will choose x so that

$$u_{i1} = bu_{i2}, \tag{5}$$

where u_{ik} represents the partial of u_i with respect to its kth argument.

For the optimal insurance plan, the partials of expected utility, U, defined in (2), with respect to a and b will be 0. The marginal condition with respect to a is

$$\frac{\partial U}{\partial a} = \sum_i p_i u_{i1} \frac{\partial x_i}{\partial a} + \sum_i p_i u_{i2} \left(\frac{-\partial P}{\partial a} - 1 - b \frac{\partial x_i}{\partial a} \right) = 0. \tag{6}$$

Given (5), we know that the first term on the right cancels the third term within the parentheses. This yields that, at the maximum,

$$\frac{\partial P}{\partial a} = -1. \tag{7}$$

This condition will always be satisfied, for, as we argued above, reducing a and increasing the premium accordingly will leave medical expenditures, the x_i's, and everything else just the same. Differentiation of (3) reveals $\partial P/\partial a$ will be equal to -1 if, as has just been established, $\partial x_i/\partial a = 0$. Most simply, an optimal plan can always be found with $a = 0$.

The marginal condition with respect to b is that

$$\frac{\partial U}{\partial b} = \sum_i p_i u_{i1} \frac{\partial x_i}{\partial b} + \sum_i p_i u_{i2} \left(\frac{-\partial P}{\partial b} - x_i - b \frac{\partial x_i}{\partial b} \right) = 0. \tag{8}$$

Given (5), two terms cancel once again. Simplify to get

$$\frac{\partial P}{\partial b} = \frac{-\sum_i p_i u_{i2} x_i}{\sum_i p_i u_{i2}}. \tag{9}$$

Differentiation of (3) yields

$$\frac{\partial P}{\partial b} = \sum_i p_i \left[(1 - b) \frac{\partial x_i}{\partial b} - x_i \right]. \tag{10}$$

If the medical expenditures, the x_i's, are the same for all conditions, the u_{i2}'s will also be equal. This means that (9) will be satisfied if the first term within the parentheses in (10) is zero. Thus, in this case the optimum plan will have $b = 1$; it will have the individual pay all marginal costs. This accords with intuition. With identical expenditures for all conditions, risk spreading is not of consequence, but appropriate incentives will still be valued. This same argument holds if the utility functions are risk neutral with respect to gambles on money.

When, as is usual, the x_i's differ and the utility functions display risk aversion, risk spreading must be considered. The u_{i2}'s will vary together with the x_i's so long as $g(x)$ is an increasing function of x. This implies that at the optimum $\partial P/\partial b$ will be less than $-\sum_i p_i x_i$. Therefore, $(1 - b)(\partial x_i/\partial b)$

must be less than 0. As $\partial x_i/\partial b$ is known to be negative, b must be less than 1. This is a general property whenever risk spreading is important. The greater its import, as indicated by the variance in the values of u_{i2}, the smaller will be the optimal b.

CASE I-A: THE OPTIMAL PLAN WITH A DEDUCTIBILITY OR LIMIT PROVISION

Our Case I insurance scheme employed a single linear function for $g(x)$ for all values of x. In many circumstances it may be desirable to have an initial range of expenditure all of which is covered by the insurance plan; beyond this initial range a linear sharing rule would apply. With such a limit plan the individual in perfect health will have no dealing with the insurance plan other than paying his normal premium. Such a plan distinguishes in effect between perfect health and all other conditions. This enables it to place some additional burden of medical expenditure on those fortunate individuals who suffer no medical ailment. A limit plan would be of the form

$$\begin{aligned} g(x) &= 0 && \text{for} \quad x \leqslant k, \qquad \text{and} \\ g(x) &= a + bx && \text{for} \quad x > k, \end{aligned}$$

where k would be expected to be $-a/b$.

Alternatively, the insurance plan might incorporate a deductibility feature, the insured paying all or some fixed amount of expense. A plan with a deductible amount would be of the form

$$\begin{aligned} g(x) &= x && \text{for} \quad x \leqslant k, \qquad \text{and} \\ g(x) &= a + bx && \text{for} \quad x > k, \end{aligned}$$

where k would be expected to be $a/(1 - b)$.

Among plans with a limit or deductible feature, the optimal one has a limit feature. It is

$$\begin{aligned} &g(x) = 0 \qquad \text{for} \quad x \leqslant 767.10, \qquad \text{and} \\ &g(x) = .03448(x - 767.10) = -26.45 + .03448x \qquad \text{for} \quad x > 767.10. \end{aligned}$$

The premium associated with this scheme is 267.32. The expenditures are

$$\begin{aligned} x_2 &= 767.10, & g(x_2) &= 0, \\ x_3 &= 23{,}686.75, & g(x_3) &= 790.33. \end{aligned}$$

This Case I-A plan[9] gives a somewhat higher expected utility than the optimal Case I plan. Its advantage accrues because it can better exploit the condition perfect health. In effect, the optimal Case I-A plan gives a lump sum payment to the individual if his condition is not perfect health.

CASE II: Reimbursement from Insurance a Different Function of Expenses to the Insured for Each Condition

There is a tremendous loss in efficiency if no distinctions are made amongst the reimbursement schedules for different medical conditions. Ideally, an insurance plan would work like a perfect contingent claims market. Each possible medical condition would be indexed as a state of the world. The actual transfer that would take place, the amount of reimbursement from the insurance plan (the net after premium would be negative in cases in which the individual is relatively healthy), would depend only upon the particular state of the world.

Other Grounds for Drawing Distinctions

In real life, situations may be distinguishable on many more grounds than the particular disease or illness. Distinctions made on other criteria may be part of an optimal plan. An insurance plan may pay for some services but not others. The first X days in a hospital may be covered, but not those that follow. Charges for round-the-clock nurses services may be reimbursed on a different plan than are doctors' fees. In short, the type of distinguishability mentioned here should be thought of as representative of many other types of distinguishability.

Simulating a Contingent Claims Market

A perfect contingent claims market can be simulated by an insurance plan of the type described in Case I if the cost sharing function can be allowed to differ for each different medical condition. This would allow us to employ a set of $g_i(x)$ functions, a different function for each condition i. The objective, as before, would be to maximize (2) subject to (3)

[9] It is interesting to observe that the insured spends nothing above the fixed stipend when he has appendicitis in this Case I-A plan. He would not wish to reduce medical expenditure below this fixed amount even if he could receive b times all savings. This result is not unusual; it occurs again with the example with the Case III plan. However, it is not a universal happening with these plans. Thus far I have found no satisfactory explanation of this curiosity.

given the responses defined by (1). Now however, the $g(x)$ functions would be indexed over i as well.

The optimal set of $g_i(x)$'s has each function merely a constant term. Upon reflection, this is not surprising for it is the general way in which contingent claims markets operate. The simplest way to prove this intuitive result is to solve the equivalent problem for an actuarially fair contingent claims market. With such a contingent claims market, the individual's objective would be to find the m_i's and w_i's that maximize expected utility,

$$U = \sum_i p_i u_i(m_i, w_i), \tag{11}$$

subject to

$$\sum_i p_i(m_i + w_i) = w_0 . \tag{12}$$

The constraint equation insures that the expected value of the sum of health care expenditure and after-expenditure wealth equals initial wealth, as it would with an actuarially fair market. Case II essentially enables an individual to set up any insurance plan he wishes so long as he pays its actuarial cost.

The maximization carried out for the three conditions and utility functions from the previous example yields the optimal values of the m_i's and the w_i's:[10]

m_1^*	w_1^*	m_2^*	w_2^*	m_3^*	w_3^*
0	9752.18	656	9752.18	21,502	9752.18

From these values we can derive the specific form of the optimal Case II plan. It is

$$g_2(x) = -656 + x \qquad x > 656, \qquad \text{otherwise } 0;$$
$$g_3(x) = -21{,}502 + x \qquad x > 21{,}502, \qquad \text{otherwise } 0.$$

The expenditures with the plan will be

$$x_2 = 656, \qquad g_2(x_2) = 0$$

and

$$x_3 = 21{,}502, \qquad g_3(x_3) = 0.$$

The premium is 247.82.

[10] The w_i*'s are equal to each other, but this result should not be generalized. It occurs only because wealth enters the utility function for each condition in the same separable and additive way. Individuals' cardinal preferences for wealth might be very different for different medical conditions.

Case II works like a contingent claims market. The individual receives lump sum amounts from the insurance plan that depend solely on his medical condition. For each condition just this lump sum, the 100% coverage limit for that condition, is spent. Less is spent on medical care than with Case I or Case I-A. There is no overexpenditure because there is no sharing of costs at the margin; in other words, there is no distortion of incentives.

This absence of distortion gives the optimal Case II plan an expected utility advantage over the optimal Case I-A plan which in turn is superior to the optimal Case I plan. Case II plans are, by their very structure, the best that can be done with any actuarially fair plan.[11]

The working of a Case II plan requires that the insurance plan be able to distinguish unambiguously among all medical conditions. The distinctions must be drawn not only among diseases, but also among different levels of severity of the same disease. Quite obviously, any attempt to achieve the complete discrimination required by a Case II plan is doomed to failure. It would entail a prohibitively high investigative expense by the administering authority. It would encourage its insureds to evade through misclassification. It would be likely to create the unattractive atmosphere that accompanies most any plan in any area that requires an exact specification of some condition or characteristic of its individual members. Furthermore, it would likely lead to errors in classification, errors that would create injustices as well as inefficiencies.

[11] Professor Arrow has argued that moral suasion may play a role in preventing individuals from over-expending when they are under cost-sharing insurance plans. Consider a Case I situation in which each individual is guided by moral considerations and incurs additional expense as if there were no reimbursement from the insurance plan, as if all expenses were his expenses. It is not difficult to see that given such a moral constraint a Case I plan can be developed which will yield nearly equivalent results to those we got with Case II. A Case I plan which has a very small b (the individual pays only a small fraction of total expense) can be worked out this way. Large amounts can be spent at little cost to the beneficiary. Every beneficiary, following the categorical imperative of Immanuel Kant, will spend that amount he would have contracted to spend had he agreed in advance of knowing his condition to an actuarially fair insurance plan. Income effects are thus circumvented and the risk of incurring large medical expenses is appropriately spread among the entire group of insureds.

Difficulties arise if the w_i^*'s are not equal. A Case I plan with moral suasion will then be unable to achieve Case II results. To achieve an optimum in such a situation, the insureds would need to be empowered to divert some of the reimbursement from the plan from medical expenditures to personal wealth. (Otherwise, the insurance plan would operate like a restricted contingent claims market in which only certain types of claims were sold. On this matter, see Zeckhauser [4], particularly p. 157.)

CASE III: Reimbursement from Insurance a Different Function of Expenses to the Insured for Different Sets of Conditions

To develop our Case III plan we bore the limits and constraints on discrimination possibilities in mind. Case III is supposed to represent the types of plans that might be practical to institute in the real world. It recognizes that although it might be impossible to administer an insurance plan that had different reimbursement schedules for each possible medical condition, it still in general will be feasible to make some distinctions. Blue Cross–Blue Shield plans, for example, generally allow different amounts for the treatment of particular classes of conditions. The schedules of reimbursement for "extras" may be somewhat less particularized by condition.

Case III insurance plans have different reimbursement schedules for m different categories of conditions. Each category contains conditions among which it is not practical for the plan to distinguish. Within each category the level of reimbursement from the plan is allowed to vary as a function of the expenditures by the insureds, the people best able to determine the amount that should be spent on their own medical care.

Case III is a mixture of Cases I and II. As with Case I, a single reimbursement schedule may thus cover many conditions. The Case II element is introduced because there are many such categories, each with its own schedule.

Within the jth category there are n_j conditions, the utility functions for which will be indexed u_{ji}, where i gives the particular condition in the category. The j, i pair thus fully identifies the condition. Let x_{ji} represent the expenditure on his health selected by the individual in condition ji to maximize

$$u_{ji}(x, w_0 - P - g_j(x)). \tag{13}$$

The object is to find the reimbursement functions, the $g_j(x)$'s that maximize expected utility. Expected utility can be expressed

$$U = \sum_{j=1}^{m} \sum_{i=1}^{n_j} p_{ij} u_{ji}(x_{ji}, w_0 - P - g_j(x_{ji})). \tag{14}$$

The constraint on the premium is a global one; the same premium is charged regardless of the category of condition. This means, unfortunately for ease of computation, that Case III situations cannot be treated as a number of independent Case I situations. The constraint equation is

$$P = \sum_{j=1}^{m} \sum_{i=1}^{n_j} p_{ji}(x_{ji} - g_j(x_{ji})). \tag{15}$$

Let us restrict our attention as before to linear reimbursement schedules. The object then is to find the $g_j(x)$'s of the form $a_j + b_j x$ that maximize (14) subject to (15), recognizing that in each instance the x_{ji}'s will be selected to maximize (13).

Finding an Optimal Case III *Insurance Plan*

This could all be set up as a mammoth optimization problem which could be solved in time. Fortunately, economic intuition enables us to streamline the procedure by suboptimizing it. Assume for the moment that a manager is selected for each category of conditions. He is responsible for choosing the a_j and b_j for his category. As the manager for his category he is interested in choosing an a_j, b_j pair that maximizes the expected utility for individuals whose conditions are in his category. But he also has a responsibility to the central insurance plan. He must contribute as much as possible to the support of other categories of conditions. This support is measured as the amount of premiums paid by people with conditions in his category less the amount the insurance plan pays to these people. In categories representing serious ailments the net contribution will, of course, be negative.

Category j will contribute the

$$\sum_{i=1}^{n_j} p_{ji}(P - x_{ji} + g_j(x_{ji})), \tag{16}$$

for each member of the insurance plan. The expected utility it contributes for each member is

$$\sum_{i=1}^{n_j} p_{ji} u_{ji}(x_{ji}, w_0 - P - g_j(x_{ji})). \tag{17}$$

(When this expression is multiplied by the constant $1/\sum_{i=1}^{n_j} p_{ji}$, the reciprocal of the probability of having a condition in category j, it gives the expected utility of individuals with conditions in the category.)

Unfortunately, (16) and (17) cannot be maximized simultaneously. They are respectively negatively and positively associated functions of both a_j and b_j, thus with respect to maximization they conflict with one another. It is possible to maximize a weighted sum of the two expressions. Assign the weight 1 to (16) and λ to (17). If we maximize the weighted sum, λ represents the tradeoff rate at the margin between financial contribution and expected utility.

There is one last complication, the premium. From (15) we know that the sum over all values of j of the amounts indicated in (16) must equal zero. We also know that the size of the premium will affect the maximization procedures for (16) and (17). That is, a manager will need to know P as well as λ before he can carry out his assigned maximization, before he can pick out his a_j and b_j.

The problem can be handled as one of decentralized response to announced prices. A central authority announces a P and a λ. The managers maximize in response and send in the amounts of their net financial contributions. The value of λ must be varied until these net contributions sum to zero. If the sum is less than zero, λ must be decreased, the financial contributions to the plan must receive more emphasis. Conversely, λ is increased if the sum of the contributions is greater than zero.

When the zero yielding λ is found, the central authorities ask each of the category managers to inform them how much of a per capita contribution they are going to make to expected utility. These are summed to give the expected utility for the plan.

This process is now repeated systematically for other values of P. The central authorities can then discover which value of P together with its appropriate λ gives the highest expected utility for the entire plan. The fact that each manager maximizes the same weighted sum insures that the tradeoff rate between the two valued objectives is the same within each category. This is a condition dear to the heart and familiar to the mind of the economist.

Fortunately, the optimization process is not difficult. Expected utility is a unimodal function of P (each taken with its appropriate λ). This is not surprising if we realize that increasing P is a way of increasing the "exploitation" of the healthy by the sick. By varying P to find the maximum expected utility for the whole plan, the central authorities are finding the optimal extent of this "exploitation."

An Example of Case III *with Three Categories*

To illustrate the maximization procedure consider an example with three distinguishable categories of diseases. They are perfect health, appendicitis, and cancer. There is only one condition in the perfect health category. For appendicitis there are two conditions: ordinary, (which has the utility function used above for appendicitis), and complicated. For cancer there are two conditions: serious (which has the utility function used above for cancer), and extremely serious. Table 2 gives the details of the situation. For any values of the parameters of the example, the individual will spend more when he has complicated appendicitis or extremely serious cancer than he would with the less severe form of the disease.[12]

[12] With after premium wealth of 10,000 and no insurance plan, the individual will

TABLE II

An Example to Illustrate a Case III Plan

	Perfect Health $j = 1$	Appendicitis $j = 2$		Cancer $j = 3$	
		Ordinary $i = 1$	Complicated $i = 2$	Serious $i = 1$	Extremely Serious $i = 2$
Prob. of Occurrence	.94	.045	.005	.008	.002
Utility Function		$c = .08$	$c = 10{,}000$	$c = .25$	$c = 100$
Parameters		$d = .03$	$d = .03$	$d = .0008$	$d = .0008$
$t = .002$ all conditions					

The optimal Case III plan is

$$g_2(x) = .5376(x - 675) = -362.88 + .5376x \qquad x > 675, \qquad \text{otherwise } 0;$$

and

$$g_3(x) = .1493(x - 22{,}250) = -3321.93 + .1493x \qquad x > 22{,}250, \qquad \text{otherwise } 0.$$

The premium for the optimal Case III plan is 276.06. The optimal Case III plan is not easily compared with the other plans because it includes new medical conditions and somewhat reduced probabilities for the old conditions. To get some feeling for the relationship between the plans we calculated the optimal Case I plan for the medical conditions and probabilities in this example. The optimal Case I plan had its b equal to .03086, and a consistent premium of 284.34. The plans are compared in Table 3.

As we would expect, more is spent on every condition with the Case I plan. With the Case I plan the individual is paying only 3.1 % of any marginal costs; incentives for appropriate expenditure are sorely insufficient. The Case III plan manages to provide much more appropriate incentives. The individual pays 53.8 % of the marginal expenditure on

spend 997.45 on complicated appendicitis as opposed to 630.70 for the ordinary variety. He will spend 8460.31 on extremely serious cancer as opposed to 6320.51 for serious cancer. With $b = 1/6$ these comparisons on total expenditure are 1111.32 as opposed to 724.42 and 22,482.92 as opposed to 17,196.33.

TABLE III
Expenditures with Optimal Case I and Case III Plans

		Appendicitis		Cancer	
		Ordinary	Complicated	Serious	Extremely Serious
Case I	x	768.15	1158.55	23,913.44	30,866.29
	$g(x)$	23.71	35.76	738.07	952.66
Case III	x	675	1052.68	23,385.53	28,839.72
	$g(x)$	0	203.05	169.48	983.54

appendicitis and 14.9% of the marginal expenditure on cancer. The Case III plan gives the individual adequate medical coverage, (the total expenditures are in excess of what they would be with the optimal Case II plan, the best that can be done with any actuarially fair scheme) with much less distortion of incentives.

Summary and Conclusions

The model illustrates the importance in any insurance plan of the tradeoff between risk spreading and incentives. The flatter the $g(x)$ function, the smaller the value of $g'(x)$, the greater is the emphasis on risk spreading. To have appropriate incentives for expenditure, on the other hand, $g'(x)$ must equal 1 at the margin. One way to overcome this conflict in objectives is to have different $g(x)$ functions for different classes of conditions. The varying structures of the functions will carry out much of the risk spreading objective. To have different $g(x)$ functions it is necessary to be able to discriminate among different classes of conditions. This is a specific instance of a general problem with contingent claims markets—that is the problem of distinguishing among the different states of the world.

This medical insurance model explicitly recognizes an imperfection that is perhaps somewhat different that the usual imperfections that throw us into a second-best world. Human beings act to maximize their own welfares. This is no difficulty whatsoever in the competitive model under certainty. With externalities and public goods appropriate charges can be levied to overcome what would otherwise be inefficiencies. But in a world of uncertainty, methods of charging which bring about efficient levels of expenditure will prevent effective risk-spreading. In such a world you will be damned if you do not introduce a risk-spreading procedure, but you will damned in another way if you do. In the optimal plan, we saw that a mixed damnation hurts the least.

REFERENCES

1. K. J. ARROW, The economics of moral hazard: Further Comment, *Amer. Econ. Rev.* **58** (1968), 537–38.
2. K. J. ARROW, Uncertainty and the welfare economics of medical care, *Amer. Econ. Rev.* **53** (1963), 941–73.
3. M. V. PAULY, The economics of moral hazard, *Amer. Econ. Rev.* **58** (1968), 531–37.
4. R. J. ZECKHAUSER, Uncertainty and the need for collective action, *in* The Analysis and Evaluation of Public Expenditure: The PPB System, Joint Economic Committee, U. S. Congress, 1969.