Exclusions in Medical Plans*

Rod Garratt and John M. Marshall
Department of Economics
University of California
Santa Barbara, CA 93106

June 4, 1996

Abstract

The paper explains maximums and exclusions in medical plans as features of the consumer's first-best optimum. Discussion of such optimizing models is justified because competition among health plans might tend to favor those that are, from the consumer's viewpoint, more efficient. Copayment-style insurance is well known to promote excessive treatment, but the analysis here shows that it can also, through a mechanism not previously modeled, encourage inadequate treatment of serious illness. The analysis gives optimum criteria for adopting new medical techniques, and it clarifies aspects of "quality adjusted life year" as a means of evaluating treatments.

JEL codes: I11, I18.
Key words: Medical plan, copayment, insurance, exclusions, efficacy.

*This paper is a revised version of UCSB Working paper #12-94. Send correspondence to Rod Garratt, Department of Economics, University of California, Santa Barbara, CA 93106. Office: (805) 893-2849. Fax: (805) 893-8830. email: garratt@econ.ucsb.edu.

1 Introduction

In every medical system there are instances in which patients do not receive treatment. Treatment may be withheld by mistake, but in the interesting cases it is withheld intentionally, systematically, and with the conscious consent of at least some of the participants. The reasons that justify withholding treatment are diverse. They include: lack of medical necessity (cosmetic surgery), great expense (experimental treatments), high risk (experimental treatments again), questionable effectiveness of treatments (acupuncture, chiropractic), morality (abortion, euthanasia), boundary lines of medicine (physical therapy, occupational rehabilitation), inability of the patient to benefit (treatment of heart failure in the terminal stage of cancer). These examples are drawn from types of public and private health insurance available in the United States, but similar examples can be found in any health care system.

The reasons for excluding treatment are sensible when considered singly. Taken together, however, they are disturbingly ad hoc, and because they are ad hoc they raise doubts about how to draw the line between included and excluded treatments. The problem is difficult because treatments typically present themselves as incommensurate. The person has the illness, diagnosis has been made, the appropriate treatment is known, and the question is whether to give it. From this viewpoint the treatments of different illnesses are not in competition and there is no reasonable way to compare, say, the value of reducing facial scars for the victim of an auto crash and the value of resuscitation for the patient hospitalized with terminal cancer. The costs of treatments are also blurred because the person is covered by some form of health insurance or national health plan and hence the financial commitments to pay for treatments have already been made.

The difficulties are avoided by taking an ex ante viewpoint on the diverse treatments. In that perspective, the costs and values of the treatments are comparable. Thus consider an person who is not sick but is considering a future in which she might be. She knows the probabilities of contracting the various illnesses and the attributes of the treatments for them, but she does not know the specific illness she will have. The costs of treatments are the costs of contingent claims to treatment that deliver only in the event of the corresponding illness. Facing a menu of possible maladies and the costs of treating them, the consumer buys some treatments and not others. Those she buys are the ones she views ex ante as having sufficient benefit relative to cost. Other treatments are specifically excluded because she is unwilling to pay for them. In that way she selects the exclusions optimally as part of a contingent contract.[1]

Technical aspects of the theory are foreshadowed in several sources. Friedman and Savage (1948) proposed a type of utility that reappears here. The decision to be treated or not is similar to the choice of occupation in Marshall (1984) and Bergstrom (1986). The optimum contract for treatments is related to the optimum program of admissions to college in Garratt and Marshall (1994, 1995).

2 The Economics of Being Sick

In order for the consumer to choose treatments, she must foresee what will happen in each particular state of illness. Therefore, the first objective is to study demand for care in one specific sick state. The competition among sick states takes place later, when the consumer maximizes the expectation of utility over all states. This ordering of analysis is time-honored in the economics of uncertainty and makes matters as abundantly clear as they should be. However, it causes a technical problem because the von Neumann-Morgenstern axioms for expected utility actually apply to the whole problem, taken all at once. The methods of deriving expected utility representations are covered in many sources and innumerable textbooks. For the present discussion, the consumer is assumed to satisfy the axioms and possess an expected utility representation of preferences. The analysis begins at that point and explains the structure of the preferences and its implications for the choice of treatments.

The preferences of the consumer in the sick state s are shown in Figure 1. The horizontal axis measures consumption of an ordinary, divisible good

[1] A single, typically informed consumer is not expected to choose optimally from a menu of hundreds of contingent treatments. Nevertheless, the optimum exists in principle and consumers can acquire an approximation to it in practice. Because people differ, a selection of slightly differing contracts may be needed. Clark Havighurst (1995) discusses the legal ramifications of basing treatment on contracts rather than on universal standards and argues the benefits of diversity.

which represents all non health commodities. The amount of the ordinary good consumed in state s is x_s. The context of the sick utilities is provided by the top curve, the "well" utility function. Lower in the diagram, the "sick" utility function is the one that holds in the absence of medical treatment, and is interpreted as including diagnosis, compassionate care, pain control, and so on. The "treated" utility function applies when the illness is cared for by the available treatment. These various utilities are denoted $U^h(x_h)$, $U^s(0, x_s)$, and $U^s(1, x_s)$, corresponding to well, sick with no treatment, and sick with treatment. Assume that for given health status and treatment choice, the consumer is always risk averse in consumption of the ordinary good.

Demand for treatment depends upon wealth. Consider a wealth z_s measured in units of the ordinary good. The utility of an person with untreated illness is $U^s(0, z_s)$ because all wealth is spent on the consumption good. In the case of treatment, however, a cost p_s is paid to a competitive health services sector, leaving consumption of the ordinary good equal to $z_s - p_s$. Thus, the indirect utility of wealth given the treatment is $U^s(1, z_s - p_s)$, which is obtained by shifting the curve $U^s(1, x_s)$ to the right by a distance p_s, as is done in Figure 1.

In each sick state, demand for treatment is summarized in the indirect utility function, which is the upper envelope of $U^s(0, z_s)$ and $U^s(1, z_s - p_s)$. It is illustrated in Figure 1 by the heavier, non-concave curve

$$M_s(z_s) = \max[U^s(0, z_s), U^s(1, z_s - p_s)] \tag{1}$$

At low wealth the utility of being treated at cost p_s lies below the untreated curve $U^s(0, z_s)$, and the consumer prefers not to be treated. At a critical wealth the consumer becomes indifferent, and for wealth above that level the treatment is preferred in spite of its cost.

The indirect utility function in Figure 1 is of the risk-loving type identified by Friedman and Savage (1948). Thus consumers might be expected to seek gambles in wealth prior to deciding on their consumption of treatments.[2] In the present application, however, gambling turns out not to be very important. The real importance of the non-concave indirect utility functions is that they define the demand for state-contingent wealth transfers, which is in effect a demand for treatments in the states, as the wealth received is used to purchase the treatment as well as the non health commodity.

[2]The role for randomization arises in a similar fashion in Marshall (1984), Bergstrom (1986) and Garratt and Marshall (1994,1995).

Although treatments are specific to different sick states, their values are commensurate through a measure of efficacy, which is derived as follows: Define $M_s^*(z_s)$ to be the concave hull of $M_s(z_s)$, that is, the least concave function that is nowhere less than $M_s(z_s)$. This is illustrated in Figure 2. Now define the efficacy of the treatment as the slope of the linear portion of $M_s^*(z_s)$. Clearly efficacy has the units of marginal utility, and it is in fact the conditional marginal utility of the treatment, given that wealth is adjusted to z_{1s} if the illness is treated and to z_{0s} if it is not.

It is assumed that the contingent cost of treatment is the expected cost, and thus the supply of contingent wealths is risk-neutral. This assumption is justified by the law of large numbers and the availability of large financial markets. Supply of treatments is also assumed to be free of transactions costs, an assumption of convenience. Departures from risk-neutrality of the insurer would have consequences studied in numerous places including Marshall (1974). Analysis of transactions costs would follow Raviv (1979) and others (see Gollier, 1992). Fair prices are clearly the right ones from the viewpoint of social planning, but predictable modifications are needed if the theory is applied to real-world markets for medical plans or health maintenance services.

States belong to a finite set S. The probability of a state is π_s. Since contingent claims are fairly priced, π_s is also the price of a claim to a unit of wealth contingent upon state s. The consumer's endowed wealth in state s, before entering the contract, is a given quantity w_s. Finding the optimum treatments consists of trading these contingent endowed wealths for contingent insured wealths z_s. Given these notations, the problem defining the optimum medical contract is:

$$\underset{\{z_s\}}{\text{Maximize}} \ \sum_{s\in S} \pi_s M_s^*(z_s) \tag{2}$$

subject to the budget constraint

$$\sum_{s\in S} \pi_s z_s = \sum_{s\in S} \pi_s w_s \tag{3}$$

The problem in equations (2) and (3) is a concave programming problem. The solution $\{\bar{z}_s\}_{s=1}^S$ involves a Lagrangian multiplier m^* satisfying, for all states s

$$M_s^{*\prime}(\bar{z}_s) = m^* \tag{4}$$

The condition means that marginal utility is equalized across states. It is a familiar condition in the demand for fairly-priced insurance and is the result of the consumer trading wealth out of states that have low marginal utility and into those having high marginal utility until no further trades of that kind can be made.

It is possible that in some states $\bar{z}_s$ falls in the linear portion of $M^*_s(z_s)$. In any state where $\bar{z}_s$ falls in that way, the solution requires that the consumer is given a fair gamble between the wealths defining the end points of the linear portions of $M^*_s(z_s)$, the points z_{0s} and z_{1s} in Figure 2. Fortunately, the optimum can always be reached with an allocation requiring a gamble in no more than one state. The argument for this fact is constructive. Suppose that in the optimum allocation there are at least two such states. Take any two of them. Since the marginal utility of wealth in both states is the same, a small amount of contingent wealth that is moved from one state to the other using fair prices does not change expected utility. Therefore, change the optimum contract by moving contingent wealth from the first state to the second until either wealth in the first state s falls to z_{0s} or wealth in the second state s' rises to $z_{1s'}$, whichever happens first. At this point the contract remains optimum but the number of gambling states is reduced by one. If there are still two gambling states, proceed again to reduce their number by one, and repeat as often as necessary. Expected utility is maintained throughout these procedures. Eventually, at most one gambling state remains.

The optimum medical contract could be found under the restriction that gambles are not allowed. Replace $M^*_s(z_s)$ in equation (2) with $M_s(z_s)$. The problem is then a type of knapsack problem and the solution would be computed using branch and bound techniques. However, avoiding the gamble is not an improvement. The gamble arises in the optimum solution because the consumer demands it, and eliminating it reduces welfare. In any event, the issue of whether or not to allow gambles is a minor concern because a single state involving a gamble is negligible in a state space that consists of all of the diagnoses described in the library of a medical school.

3 Exclusions and Inclusions

Exclusions in the contract are contingent treatments that the consumer optimally chooses not to purchase. The basic fact about exclusions is that they exist for both grave and trivial illnesses. American health insurance plans give some clue to the situation. In them, minor cosmetic surgery is excluded and so is heroic, experimental treatment of catastrophic illness. In many cases, lifetime benefits are subject to a specific maximum, in effect excluding certain expensive treatments. Other maximums may apply to specific ailments or treatments.

Construction of the exclusions is made by arranging the sick states in order of increasing efficacy of the treatment. Efficacy is measured by the slope, denoted m_s, of the linear portion of $M_s^*(z_s)$ in each frame of Figure 3. In Figure 3, the efficacy of treatment increases from A to C. Among the states with lower slopes will be a hodgepodge of different situations. Minor cosmetic surgery (Figure 3A) has a low slope because the illness is not very damaging. Heart-lung transplant (Figure 3B) has a low slope because it is debilitating, expensive, painful and risky. Among the high-slope states is emergency treatment for an initial heart attack (Figure 3C), which is highly efficacious because the illness is grave, the treatment is beneficial, and the cost is not too high.

Exclusions in the optimum medical contract are characterized by a critical value of efficacy which is equal to the marginal utility of wealth m^*. As shown in Figure 3, treatments are always included when efficacy is greater than m^* and always excluded when efficacy is less than m^*. In the sick states of Figures 3A and 3B the tangency to $M_s^*(z_s)$ is below the linear portion – the treatments are excluded. In the sick state in Figure 3C, the tangency is above the linear portion, and that treatment is included. There may be a state in which the tangency falls on the linear portion, as discussed above, but happening upon that particular state is improbable and not illustrated.

The low efficacy states in Figure 3A and 3B are exclusions in the optimum contract. Treatment is not provided because the consumer is unwilling to purchase it contingently at a fair price. The optimum contract does provide treatment in high efficacy states like that in Figure 3C. Thus the theory explains exclusions in a wholly intuitive manner. It suggests, moreover, that intricate patterns of inclusion and exclusion observed in many medical plans are not anomalies but are instead expected.

The discussion is summarized in the following result.

Theorem 1 (Inclusion Theorem) *In an optimum contract defined by equations (2) and (3), there exists a critical efficacy level m^* such that for any state s*

a. if $m_s > m^*$ the contract includes treatment *b. if $m_s < m^*$ the contract excludes treatment*

c. if $m_s = m^*$ the contract may include treatment, exclude treatment, or supply a gamble between treatment and non treatment.

There is an optimum contract in which a gamble is provided in no more than one state.

The parallel between the critical m^* and the value of life is interesting. In one case, all life-saving projects are ranked in order of decreasing marginal cost. Given finite resources, there is an optimum cut-off level of marginal cost – the value of life – and projects having marginal cost above that level should not be undertaken. The comparability of life-saving expenditures and the existence of a cut-off level is completely intuitive. The comparability of diverse medical treatments is less intuitive. Nevertheless, the inclusion theorem shows that treatments, like safety projects, can be ordered in terms of productivity and that, again like safety projects, there is a single criterion that determines which treatments to undertake.

4 Copayment Plans

Studies of health insurance in the United States often use the copayment plan as an analytical vehicle. In a copayment plan, the person selects a treatment after she becomes ill and pays a copayment equal to a fraction of its cost. Copayment plans are clearly suboptimal, and some aspects of the suboptimality are well known. In one story, a person with a particular illness buys too much medical care because of the price subsidy (Pauly 1968). Additionally, since people cannot commit themselves to restrain their ex post demands, they pay premiums that are too high (Zeckhauser 1970).

In this paper, there is exactly one treatment for each illness and there are no feasible variations in care. The choice is to treat the illness or not. Consequently, there is no possibility of the types of waste studied by Pauly and Zeckhauser. However, a new type of inefficiency is revealed, as persons covered by copayment plans make errors in deciding whether or not to be treated.

To clarify matters, suppose that the utility benefit of the treatment in each sick state is an additive quantity B_s, *i.e.*,

$$U^s(1, w) = U^s(0, w) + B_s \tag{5}$$

Under this assumption, the efficacy of treatment has a convenient value. Recall that efficacy is the slope of the linear portion of $M_s^*(z_s)$. This slope is the increase in utility associated with the treatment divided by the difference in the optimum wealths that accompany treatment and non treatment, $z_{1s} - z_{0s}$. Since the treatment produces a parallel, vertical shift of utility and the cost produces a parallel horizontal shift, the difference $z_{1s} - z_{0s}$ is equal to the cost of treatment, p_s. That is, the optimum wealth transfer just reimburses the cost and does not affect the level of consumption of the ordinary good. Thus the benefit of treatment, after wealth adjustments, remains equal to B_s, and the efficacy of the treatment is B_s/p_s.

In the optimum medical contract there is a critical efficacy m^* and the treatment of disease s is included in the optimum plan if and only if $B_s/p_s \geq m^*$, *i.e.*,

$$p_s \leq B_s/m^* \tag{6}$$

The condition in equation (6) characterizes the optimum.[3]

The treatments selected under a copayment plan are dictated by the reservation price of the treatment. The reservation price R_s satisfies

$$U^s(1, w - R_s) = U^s(0, w) \tag{7}$$

Letting the copayment rate be θ, a treatment is chosen under the copayment plan if and only if $\theta p_s \leq R_s$, *i.e.*,

$$p_s \leq R_s/\theta \tag{8}$$

Now the treatments selected under the copayment system can be compared to those included in the optimum medical plan. Consider a group of diseases all sharing the same untreated prognosis $U^s(0, w)$. Treatments for some of the diseases are represented by points such as X, Y, or Z in Figure 4. The question is which treatments are given under the two plans?

In Figure 4, benefit is a variable B that depends on the state s, and reservation price is a function $R(B)$. The function $R(B)$ is concave, as can be seen in a short derivation from equations (5) and (7), because the utility functions are concave. The straight line of slope $1/m^*$ represents the critical relation of cost and benefit (from equation 6) in the optimum medical contract, and a treatment is included if and only if it lies on or below the line. For the

[3]If $p_s = B_s/m^*$ then the treatment is included in the sense that it is part of an optimizing gamble. As mentioned previously, there needs to be, at most, one such treatment.

copayment plan, the curve $R(B)/\theta$ represents the critical relationship of cost and benefit. From equation (8) it follows that a treatment is purchased in the copayment plan if and only if it lies on or below the curve. For instance, the point Y is included in both plans.

In looking at the group of similar illnesses, the copayment plan is sub optimum in two ways: First, consumers purchase some low-cost treatments like point X in Figure 4 that are not part of the optimum. This type of suboptimality is broadly consistent with the findings of excessive expenditure in copayment plans by Pauly and Zeckhauser. The second category of inefficiency is more interesting. Some high-cost treatments like Z in Figure 4 that should be included in the optimum are not purchased by people covered by the copayment plan. This means that people covered by copayment plans may under-spend on health care. Previous studies of copayment systems do not identify this possibility because their modelling ignores the indivisibility of medical treatment.

Actual medical plans are not true copayment systems. Most have provisions to reduce the copayment in the most expensive illnesses, either through a vanishing copayment rate or the addition of a "major medical" package that takes over the largest payments. In these real-world cases the potential for under spending on the most expensive treatments is greatly reduced, as is the usefulness of the copayment plan as a policy model.

5 New Treatments

The optimum medical contract forms a basis for evaluating new treatments. A new treatment is adopted if it has a place in the optimum contract. In this approach, superiority of a new, more expensive treatment over the old treatment is not enough to justify adoption. The extent of superiority must be sufficient to justify drawing resources away from other sick states. Conversely, a lower-cost innovation that gives worse results than the old treatment is sometimes adopted because it releases valuable resources.

In the contract method, competition of treatments across sick states is measured by the shadow-price of resources in the optimum contract. The Lagrangian multiplier m^*, which is interpreted as the marginal utility of wealth, is the needed shadow price. Its existence permits a decentralized evaluation of treatments. The decision to adopt a new treatment for one illness can be made without direct reference to all of the other illnesses and treatments but with reference only to m^*.

To be clear about how to deal with a new treatment, suppose that in a particular sick state r a new treatment "n" becomes available at the price p_r^n. The problem defining the new optimum contract is the same as the one described in equations (2) and (3) except for modifications of the $M_r(z_r)$ and $M_r^*(z_r)$ functions. In particular, $M_r(z_r)$ becomes the upper envelope of the three (instead of two) treatment options 0, 1, and n, as illustrated in Figure 5. The solution proceeds by substituting the concave hulls M^* for all of the M functions, including $M_r^*(z_r)$, as shown in Figure 5. The new optimum contract yields the new critical efficacy and shadow price m^*. In practice a new m^* would not be computed for each potential new treatment because no single sick state would have a significant effect on its value. The critical m^* would be a known constant from the viewpoint of evaluating any particular new treatment.

The decentralized method of evaluating treatments in each sick state is similar to maximizing the Hamiltonian at each point of a dynamic optimization problem. Given the right m^*, the optimum treatment in each state is chosen by a local maximization over the treatment and the wealth to be allocated to the state. The optimum treatment-wealth pair (t, z) is the one that gives the *maximum net value* of $U^r(t, z - p_r^t) - m^*z$. The sense of this rule is apparent in Figure 5. There the variable on the vertical axis is utility U and that on the horizontal axis is wealth w. The lines of slope m^* are therefore isobars of the valuation function $U - m^*z$. The tangency of one of these lines to the $M_r^*(z_r)$ function defines the treatment-wealth pair that is chosen in state r. That pair is also the one that reaches the highest value of $U - m^*z$. These considerations justify state-by-state maximization of net value.

Showing that a new treatment has the highest net value of all available treatments might be onerous. Fortunately, the comparison of net value becomes binary when the second-best treatment is already known. To see this, suppose that the incumbent treatment is option one, which was the optimum treatment before the innovation was available. Supposing further that the innovation has not affected the value of m^* materially, the net value of option one exceeds that of any other option except for, possibly, the new treatment. To evaluate the new treatment requires only a one-on-one comparison between it and treatment one, and the comparison can be made conveniently through a single expression.[4] Denoting the value maximizing

[4]In some states, the incumbent treatment might be no treatment, and that treatment

wealth for treatment one by $\bar{z}_r$ and that for the new treatment by $\bar{z}_r^n$, the value increment of substituting the new treatment for treatment one is

$$U^r(n, \bar{z}_r^n - p_r^n) - U^r(1, \bar{z}_r - p_r) - m^*(\bar{z}_r^n - \bar{z}_r) \tag{9}$$

and the rule is to adopt the new treatment when the value increment is positive.

5.1 Money Value of Treatments

Up to this point the value functions are denominated in units of utility and are therefore arbitrary up to a common, increasing, affine transformation. They become less arbitrary and more accessible when they are divided by m^*, yielding a dollar value of the improvement in treatment. In state r, where treatment one is the incumbent, the value is

$$\frac{1}{m^*}[U^r(n, \bar{z}_r^n - p_r^n) - U^r(1, \bar{z}_r - p_r)] - (\bar{z}_r^n - \bar{z}_r) \tag{10}$$

The first major additive term is the extra benefit of the new treatment. It is interpreted as the marginal rate of substitution between wealth and the probability of receiving the new treatment. Alternatively, it is the extra money the consumer standing at time zero would pay in state s to get the new treatment instead of the incumbent one. The value can be studied indirectly by making a catalog of abilities, comforts, and life spans associated with the treatments and imputing values to them. It can also be studied directly by quizzing consumers.

The second part of the expression in equation (10) is the extra cost. Clearly the new treatment is adopted if the benefit outweighs the cost. Note moreover that a new treatment can give worse outcomes in state s, and still be adopted because it saves costs. In that case, the savings are transferred into other sick states where wealth is more highly valued at the margin.

would then be the focus of the binary comparison. More generally, there may be many treatment options available in some states before the introduction of the new treatment. In such cases, the evaluation of the new treatment is not more complicated since the adoption decision is still based on a binary comparison with the incumbent treatment, whatever that may be.

5.2 Relation to QALY's

The utility functions used here are naturally interpreted as indexes of the quality and length of life, and because of that the contract method is comparable to the method of "quality adjusted life years" often discussed as a means of evaluating new or alternative treatments. The methods discussed here differ, however, in several ways. First, the contract method is broader because it covers minor ailments that do not affect the length of life at all and hardly affect its quality. Second, the contract method gives a larger role to consumer sovereignty.

The third and biggest difference is that the contract method takes full account of costs. Costs in the contract method include not only the actual costs of the treatments but also differences in optimum wealth, for instance the difference $\bar{z}_r^n - \bar{z}_r$ immediately above. The role of these differences is interesting both in itself and in its relation to the concept of quality adjusted life years. In order to clarify the role of costs in the contract method, it is convenient to reduce the range of possible behavior through some assumptions. The assumptions produce a neutral case in which the sick person has the same optimum annual consumption of ordinary goods regardless of the chosen medical treatment. The case is interesting in itself and also serves as an intellectual base-line for considering more general situations.

In the neutral case, annual utility under the incumbent treatment is $u(c)$. Lifetime utility is the undiscounted sum of annual utilities, and the market discount rate is zero. The years of survival under this treatment are J. Again consider state r, where treatment one is the incumbent. Then utility, given wealth z_r and treatment one at cost p_r, is

$$U^r(1, z_r - p_r) = \underset{c_t}{\text{Maximum}} \ \sum_{t=1}^{J} u(c_t) \tag{11}$$

$$\text{subject to } z_r - p_r = \sum_{t=1}^{J} c_t.$$

The wealth constraint does not show earnings explicitly, and one interpretation is that the person is retired. However, if the treatment enables the person to return to work, the increased earnings must be accounted for by deducting them from the cost of treatment – that is, by making them implicit in the p_r. The immediate consequence of the problem in equation (11) is that the utility associated with treatment one is

$$U^r(1, z_r - p_r) = Ju\left(\frac{z_r - p_r}{J}\right) \tag{12}$$

Now consider the new treatment n which assures survival of N years and confers a utility $u(c_t) + B$ in each year, where B may be positive, zero, or negative. Reasoning as immediately above, the overall utility from the new treatment is

$$U^r(n, z_r - p_r) = Nu\left(\frac{z_r - p_r}{N}\right) + NB \tag{13}$$

Suppose that the consumer is in an optimum contract and the critical efficacy is m^*. The optimum wealth in state r under treatment one is $\bar{z}_r$ satisfying

$$U_2^r(1, \bar{z}_r - p_r) = m^* \tag{14}$$

where the subscript "2" denotes differentiation with respect to the second argument. The annual rate of consumption in this situation is $\bar{c}$ where

$$\bar{c} = \frac{\bar{z}_r - p_r}{J} \tag{15}$$

It is easy to see that

$$u'(\bar{c}) = m^* \tag{16}$$

The neutrality of this case comes about because annual consumption with the new treatment will also be at the rate $\bar{c}$. In fact, it is readily verified that

$$U_2^r(n, \bar{z}_r^n - p_r^n) = u'(\bar{c}) = m^* \tag{17}$$

and hence the optimum wealth under the new treatment is

$$\bar{z}_r^n = N\bar{c} + p_r^n \tag{18}$$

In the neutral case the new treatment causes no adjustment of annual living standards, but it extends life (or perhaps shortens it). The marginal cost of the new treatment is $\bar{z}_r^n - \bar{z}_r$ which is, after some manipulations

$$\bar{z}_r^n - \bar{z}_r = (N - J)\bar{c} + p_r^n - p_r \tag{19}$$

The marginal cost of the new treatment includes the cost of consumption for the person in the extra years of life. The interpretation is obvious when the person is, in fact, retired, but it would be less obvious if the treatment leads to more years of working, which would be reflected in reduced values of p_r^n.

The marginal benefit of the new treatment can be written, using equation (10) and the derivations of this section as

$$\frac{1}{m^*}[(N-J)u(\bar{c}) + NB] \tag{20}$$

The first term in the bracket is the utility of additional years of life at the same level of satisfaction, and the second term is the utility increment for all remaining years. In the literature on QALY's, a major concern is with treatments that extend life by making it less pleasant. That case corresponds to a negative benefit – a negative value of B in equation (20) – which in the contract method also weighs against the advantages of longer life. To that extent the evaluation of treatments by quality-adjusted life years is parallel in effect to the contract method. However, in the contract method the disadvantages of longer life include the extra cost of maintaining it.

The contract method seems mean-spirited when it requires that the cost of longer life should be counted against a new treatment, but that semblance is deceiving. Resources saved by not prolonging life are not properly thought of as reverting to an anonymous pension fund or the social security system. Resources saved in one state are used by the consumer to prolong or enhance life in another state. Moreover, the critical marginal utility m^* can mediate between utility and money in such a way as to justify treatments that are very expensive. Finally, a new treatment that extends working years has a price that is diminished by the amount of increased earnings, an advantage for life-prolonging treatments even if the side effects are unpleasant.

5.3 Beyond the Neutral Case

A little further insight into the contract method can be gained by considering variations in the neutral case. Some treatments extend abilities as well as lengthen life, and such treatments should *enhance* consumption in the sense that, at the same consumption level, the marginal utility of wealth is higher for the person having the treatment. The condition of enhancement is stronger than normality, but it seems reasonable for many treatments. The main effect of an enhancing treatment is to increase the optimum wealth level associated with treatment. This happens because optimum annual consumption is higher for the person receiving the new treatment. So there are more years and more consumption per year. The optimum contract supplies not only the treatment but also sufficient wealth in the state to enjoy consumption that the treatment enhances.

6 Concluding Remarks

One way of lowering the cost of a health care system is to restrict the treatments that are given. The state of Oregon tried this type of cost-cutting in revising its Medicaid program. That experience proved that a medical plan can be managed by manipulating its included and excluded treatments, and it proved that exclusions are a politically sensitive issue. The criteria used to justify exclusions differed from the measure of efficacy introduced above, but in counting both the effects and the costs of treatments, the Oregon criteria used the right kinds of information.

Experimentation with public medical plans is infrequent, but the private sector is constantly experimenting. Health maintenance organizations and the many forms of managed care continually adjust their treatment plans. Such experimentation has a chance to improve the market for medical care, as successful innovations drive out their less successful competitors. This is especially true, according to the theory, if the decisions to give or withhold treatment are based on the costs and on medical studies of the benefits.

The process of innovation in the private sector is far from perfect. Providers are secretive about the many subtle exclusions that are in practice part of managed care, and the lack of openness diminishes the value to consumers of having a choice among alternative treatment plans. At least that is true if consumers are really deceived. However, the cost-savings of managed care are widely advertised, and the plans have progressively attracted more and more subscribers. Unless consumers are wholly irrational, they must recognize at some level that they are giving up treatment to get premium reductions.

References

[1] Bergstrom, Theodore. "Soldiers of Fortune?" in Walter P. Heller, Ross M. Starr and David A. Starrett, eds., *Essays in Honor of Kenneth J. Arrow, Volume II: Equilibrium Analysis*. New York: Cambridge University Press, 1986, pp. 57-80.

[2] Friedman, Milton and Savage, Leonard J. "The Utility Analysis of Choices Involving Risk." *Journal of Political Economy*, August 1948, 56, pp. 279-304.

[3] Garratt, Rod and Marshall, John M. "Public Finance of Private Goods: The Case of College Education." *Journal of Political Economy*, June 1994, 102(3), pp. 566-82.

[4] Garratt, Rod and Marshall, John M. "Optimum Tuitions, Taxes and Fees when Completion is Uncertain." *Education Economics*, December 1995, 3(3), pp. 219-234.

[5] Gollier, Christian. "Economic Theory of Risk Exchanges: A Review." in Georges Dionne, ed., *Contributions to Insurance Economics*, Boston: Kluwer Academic Publishers, 1992, pp. 3-23.

[6] Havighurst, Clark C. *Health Care Choices: Private Contracts as Instruments of Health Reform*, The AEI Press, Washington, D.C. 1995.

[7] Marshall, John M. "Gambles and the Shadow Price of Death." *American Economic Review*, March 1984, 74(1), pp. 73-86.

[8] ____________, "Insurance Theory: Reserves versus Mutuality." *Economic Inquiry*, December 1974, 12(4), pp. 476-92.

[9] Pauly, Mark V. "The Economics of Moral Hazard: Comment." *American Economic Review*, June 1968, 58(3) part 1, pp. 531-537.

[10] Raviv, Artur. "The Design of an Optimal Insurance Policy." *American Economic Review*, March 1979, 69(1), pp. 84-96.

[11] Thaler, Richard and Rosen, Sherwin, "The Value of Saving a Life," in Nestor E. Terleckyj, ed., *Household Production and Consumption*, New York, National Bureau of Economic Research, 1976, 265-302.

[12] Zeckhauser, Richard J. "Medical Insurance: A Case Study of the Tradeoff Between Risk Spreading and Appropriate Incentives," *Journal of Economic Theory*, March 1970, 2, pp. 10-26.