Moral Hazard and Health Insurance when Treatment is Preventive

S. Hun Seog
KAIST Business School
Korea Advanced Institute of Science and Technology
Hoegiro 87, Dongdaemun-Gu, Seoul, 130-722, KOREA
Email: seogsh@business.kaist.ac.kr
Phone: +82-2-958-3527
Fax: +82-2-958-3160

This Draft: November 2008

Abstract: We consider a two-period model under moral hazard when treatment is also preventive. In the second period, the treatment level under moral hazard is higher than that under no moral hazard. However, it may be lower than that under moral hazard, when overinsurance is not allowed. In the first period, the treatment level tends to be higher when treatment is preventive than when not. The treatment level is higher as the discount factor is higher. The results imply that allowing moral hazard may improve efficiency of the health insurance market.

Keywords: moral hazard, health insurance, treatment, prevention, two period model.
JEL Classification: I110, D820, G220, H400

Moral Hazard and Health Insurance when Treatment is Preventive

1. Introduction

In general, preventive health care is referred to the health care that is consumed before the illness occurs. Prevention is often classified into primary prevention and secondary prevention. Simply put, the primary prevention aims to lower the probability of illness, while the secondary prevention aims to lower the severity of illness. Primary preventive care includes healthy dieting and reducing smoking. On the other hand, secondary preventive care includes medical examinations and diagnostic screening.[1]

The cost of preventive care is conventionally excluded from insurance coverage in the insurance literature, since the preventive care is considered a choice by consumers, not a random event (for example, see Zweifel and Breyer, 1997). Prevention is discussed mostly in the context of moral hazard. When the insurance benefit cannot be contingent on the prevention level, the coverage for treatment will affect the selection of the prevention level of the consumer (Ehrlich and Becker, 1972; Shavell, 1979; Kenkel, 2000). As a result, an optimal insurance coverage for treatment should take into account its effect on the prevention level. Recently, however, a few studies argue that insuring preventive care costs may be optimal, if, for example, it allows more efficient risk sharing or lower treatment costs (Barigozzi, 2004; Ellis and Manning, 2007; Newhouse, 2006).

While prevention is distinguished from treatment in the existing literature, there may be no clear distinction in practice as pointed out by Ellis and Manning (2007). For example, early screening through self examination may lead to more sophisticated diagnostic examination. Such diagnostic examination can be considered preventive care because it will lower the severity of the possible illness. However, it can also be a part of treatment care, if it detects a disease. Moreover, treatment care itself may work as preventive care for future illness. For example, one of the aims of treatment for diabetes is to lower the risk of developing other complications. Treatment for a stroke includes preventive care to lower the risk of future strokes.

Based on the above observation, this paper attempts to understand the optimal treatment level and its insurance coverage when treatment care is also preventive. For this, we set up a simple two-period model where the consumer may be sick in each period. Insurance for treatment is available in each period. Departing from the existing literature, we assume that treatment in the first period may affect the probability of being sick in the second period. Insurance coverage and treatment care level in the first period will be affected by their effects on the second period outcome.

A related issue is the tradeoff between moral hazard and risk sharing, as the need for insurance for prevention is often discussed in relation with the need for more generous insurance coverage (Ellis and Manning, 2007; Newhouse, 2006). Since Pauly (1968) and Zeckhauser (1970), the moral hazard problem is one of main economic issues in health insurance. On the one hand, health insurance increases welfare level by rendering more efficient risk sharing. On the other hand, it may lower the welfare level by causing the

[1] See Kenkel (2000) for the review of prevention in the health economics.

moral hazard problem, i.e., the excessive utilization of health care. The optimal health insurance coverage is determined by balancing the welfare gains and losses. In general, neither full insurance nor no insurance is optimal.

Several health economics papers report that the moral hazard problem in the U.S. health insurance is not properly controlled (Feldstein, 1973; Feldstein and Friedman, 1977; Manning and Marquis, 1996; Pauly, 1974, 1986). The current average U.S. coinsurance rate (cost sharing by consumers) is known to be about 25%. However, for example, Manning and Marquis (1996) find that an optimal coinsurance rate is about 45%, although the efficiency loss with 25% coinsurance rate is not large. From this viewpoint, efficiency will be improved by increasing coinsurance rate.

Contrary to the above findings, however, there is an increasing sentiment that the moral hazard concern might be excessive. Given the fact that over 40 million Americans have no health insurance (U.S. Census Bureau, 2008), it is often argued that the current health insurance overemphasizes the moral hazard problem, resulting in excessive cost sharing. From this viewpoint, efficiency will be improved by lowering coinsurance rate.

A strand of literature has also sought for the possibility that the conventional approach might overemphasize the moral hazard problem. For example, Nyman (1999a, b) argues that the conventional measure of the total welfare loss may be exaggerated (for debates, see Blomqvist, 2001; Manning and Marquis, 2001). Newhouse (2006) also argues that lower cost sharing may result in increases in consumption of health services that eventually reduces the total costs.

This paper adds to the literature by investigating the tradeoff between moral hazard and risk sharing when treatment is also preventive. The optimal treatment level in our model will be higher than that when treatment is not preventive. An important implication of our results is that insurance may have to be more generous than when treatment is not preventive. Another interesting finding is that the treatment level may be lower under moral hazard than under no moral hazard if we do not allow overinsurance. This result challenges the traditional belief that moral hazard is associated with overutilization of medical care.

The remainder of the paper is composed as follows. The next section outlines the model. The third section solves the second period problem, and the fourth section solves the first period problem. The fifth section discusses the implication of the results on the health insurance debates. The last section concludes.

2. The Model

We consider a two-period model in which a consumer faces a random health loss in each period. Period is denoted by $t = 1, 2$. The consumer is an expected utility maximizer with endowment wealth of W in each period. The von Neumann-Morgenstern utility is denoted by $U(\cdot)$. The discount factor for period 2 is denoted by β. There is an uncertainty regarding health status in each period. The uncertainty is described by two states of nature, denoted by s_t, in period t: the no loss state ($s_t = 0$) and the loss state ($s_t = 1$). The loss state can occur with probability p_t in period t. The consumer suffers a fixed health loss D_t in the loss state in period t. In the loss state, the consumer may receive medical treatment of cost x_t which enhances the health level by $H_t(x_t)$. We assume that $H_t'(x_t) > 0$ and $H_t''(x_t) < 0$. We assume that health level and health loss are expressed in monetary terms. We further assume that treatment in period 1 includes (primary) preventive health care in period 2, i.e., $p_2 = p_2(x_1)$ with $p_2' < 0$ and $p_2'' > 0$.

The consumer may purchase health insurance for treatment before the realization of health status in each period. For simplicity, we focus on the short term coinsurance contract, where a_t is the insurance payment rate by the insurer. The indemnity I_t is thus given by $a_t x_t$. The insurance premium is denoted by Q_t. We assume that the premium is actuarially fair, $Q_t = p_t I_t = a_t p_t x_t$. We also assume that the payment is made right after treatment in each period.

When the consumer purchases insurance contract (Q_t, I_t), her expected utility in period t is expressed as follows.

$$EV_t = (1 - p_t)U(W - Q_t) + p_t U(W - Q_t - D_t + H_t(x_t) - x_t + I_t)$$

Let us denote W_{ts} for the wealth in the state s in period t: $W_{t0} = W - Q_t$ and $W_{t1} = W - Q_t - D_t + H_t(x_t) - x_t + I_t$. Let us also denote U_{ts} for $U(W_{ts})$.

The overall expected utility in period 1 can be denoted by

$EU = EV_1 + \beta EV_2$,

where β represents the discount factor.

The moral hazard problem occurs because the consumer selects treatment x_t in the loss state given the insurance contract. The details, however, are different between periods (see the following sections). In period 2, the consumer selects treatment x_2 to maximize the ex post utility after the health loss occurs. In period 1, however, the choice of treatment x_1 also depends on its effect on the second period utility.

In sum, the consumer's problem is to select $\{x_1, a_1; x_2, a_2\}$ to maximize her expected utility, given the moral hazard problem. Let us solve the problem by a backward induction from the second period problem.

3. Second period problem

In period 2, the consumer selects treatment x_2 to maximize the ex post utility after the health loss occurs. That is, an optimal x_2 solves

$$\text{Max}_{x2}\ U(W - Q_2 - D_2 + H_2(x_2) - x_2 + a_2 x_2), \text{ given } Q_2 \text{ and } a_2.$$

Assuming an interior solution, this condition can be expressed as

$p_2 U_{21}' \cdot (H_2' + a_2 - 1) = 0.$

Under the assumption that $U' > 0$, the condition is equivalent to

$H_2' + a_2 - 1 = 0.$

On the other hand, the consumer selects insurance coverage a_2 to maximize the expected utility. Therefore, given p_2, the second period problem can be stated as follows.

$$\text{Max}_{a2,\, x2,\, Q2}\ EV_2 = (1 - p_2)U(W - Q_2) + p_2 U(W - Q_2 - D_2 + H_2(x_2) - x_2 + a_2 x_2) \qquad (3.1)$$

s.t. $Q_2 = p_2 a_2 x_2$ $[\lambda_2]$

$H_2' + a_2 - 1 = 0$ $[\mu_2]$

To solve the problem, let us denote λ_2 and μ_2 for the Lagrange multipliers attached to constraints. For simplicity, let us suppress the time subscript 2 in this section, unless stated otherwise. Now, the Lagrangian can be expressed as follows.

$$L = (1 - p)U_0 + pU_1 + \lambda(Q - pax) + \mu(H' + a - 1) \quad (3.2)$$

First order conditions are

$$L_x = pU_1'(H' + a - 1) - \lambda pa + \mu H'' = 0. \quad (3.3)$$

$$L_a = pU_1'x - \lambda px + \mu = 0.$$

$$L_Q = -(1 - p)U_0' - pU_1' + \lambda = 0.$$

$$L_\lambda = Q - pax = 0$$

$$L_\mu = H' + a - 1 = 0$$

No moral hazard case

As a reference, let us first find the first best outcome, assuming there is no moral hazard. No moral hazard case corresponds to $\mu = 0$ in (3.3).

=>

$$L_x = pU_1'(H' + a - 1) - \lambda pa = 0. \quad (3.4)$$

$$L_a = pU_1'x - \lambda px = 0.$$

$$L_Q = -(1 - p)U_0' - pU_1' + \lambda = 0.$$

$$L_\lambda = Q - pax = 0$$

=> $\lambda = (1 - p)U_0' + pU_1' = U_1'$, $H' = 1$,

Thus, $U_0' = U_1'$, implying that $-D + H(x) - x + ax = 0$.

As a result, the solution x* and a* satisfy the following.

$$H'(x^*) = 1, \quad (3.5)$$

$$a^* = [D - H(x^*)]/x^* + 1.$$

The first condition in (3.5) implies the treatment is determined so that the marginal benefit of treatment equals the marginal cost. The second condition implies that the risk is fully removed. In sum, the first best solution achieves the cost efficiency in treatment and the complete risk hedge.

Technically, a* can be negative if H(x*) is very high. Since this case is not interesting, we will assume a* > 0 throughout this paper. On the other hand, a* can be greater than 1, exhibiting overinsurance, if D > H(x*). Given that a health loss is often not fully compensated, it may be the case that D > H(x*). Let us summarize the results as follows.

Lemma 1: [First-best outcome]

(A) Treatment is determined to maximize the treatment efficiency.

(B) Overinsurance is optimal, if the health status is not fully recovered.

[proof] See the text above. ///

On the other hand, it is commonly assumed in literature that overinsurance is not allowed: $0 \le a \le 1$. The solution in such a case will be called the *constrained first best solution* to distinguish it from the first best one. If $D \le H(x^*)$, then the constrained first best solution is the same as the first best solution. If $D > H(x^*)$, two solutions will differ from each other. From now on, we will focus on the case of $D > H(x^*)$. Let us add superscript ** for the constrained first best solution. The solution exhibits the following properties.

Lemma 2: [Constrained first-best outcome]
When the health status is not fully recovered with the first-best treatment, the constrained first-best outcome has the following properties.

(A) Insurance coverage is full.

(B) The treatment is greater than the first-best treatment.

[proof] (A) The optimal coverage, a^{**}, should be 1, since the first best coverage a^* is greater than 1, when $D > H(x^*)$, from lemma 1. (B) Suppose on the contrary that $x^{**} \le x^*$, implying $H'(x^{**}) \ge 1$. Given $a^{**} = 1$, we should have $\lambda^{**} = U_1'H'(x^{**})$ from $L_x = 0$ in (3.4). Plugging this expression for λ into $L_Q = 0$, we have $U_0 \le U_1$, which implies that $H(x^*) \ge H(x^{**}) \ge D$, a contradiction. ///

With full coverage and $H(x^{**}) < D$, we have $U_0 > U_1$.[2] From (3.4), $\lambda^{**} = (1-p)U_0' + pU_1' \equiv EV' < U_1'$. In sum, the solution satisfies the following.

$H'(x^{**}) = \lambda^{**}/U_1' < 1$. (3.6)

$a^{**} = 1$.

Although insurance coverage is full, the risk is not fully hedged, since the health status is not fully recovered. In order to compensate the health loss, the consumer selects a high level of treatment.

Moral hazard case

Now, let us go back to the moral hazard case. The solution will solve the first order conditions (3.3). From (3.3), we obtain the following.

$L_\mu = 0 \Rightarrow H' = 1 - a$. (3.7)

$L_Q = 0 \Rightarrow \lambda = EV' = (1-p)U_0' + pU_1'$

$L_a = 0 \Rightarrow \mu = -p(1-p)x(U_1' - U_0')$

$L_x = 0 \Rightarrow -a\{(1-p)U_0' + pU_1'\} - (1-p)x(U_1' - U_0')H'' = 0$

$L_\lambda = 0 \Rightarrow Q = pax$.

From (3.3), we can characterize the solutions as follows. Let us add superscript m for the solution.

[2] Given $a^{**} = 1$, $L_a = 0$ is ignored.

Proposition 1: [Second-period outcome]

(A) Insurance coverage (i) is partial, and (ii) is less than the first best coverage.

(B) The treatment is greater than the first-best treatment.

[proof] (A) (i) First, full (or over-) insurance is not optimal, because $a^m \geq 1$ implies $H' \leq 0$, an infinite treatment. Second, no insurance is not optimal, either. Suppose on the contrary that $a^m = 0$. In this case, $x^m = x^*$, since $H'(x^m) = 1$. This implies $U_1' > U_0'$, since $-D + H(x^*) – x^* = -a^*x^* < 0$. On the other hand, for no insurance to be optimal, we should have $L_a \leq 0$ at $a = 0$. From $L_x = 0$ in (3.3), $a^m = 0$ is followed by $\mu = 0$. Then, $L_a \leq 0$ implies $\lambda \geq U_1'$, which in turn implies $U_0' \geq U_1'$, since $\lambda = (1 – p)U_0' + pU_1'$ from $L_Q = 0$. A contradiction. Thus, $0 < a^m < 1$. (ii) Since $a^m > 0$, the first order conditions (3.3) fully characterize the solution. We also observe from (3.3) that $\lambda > 0$ and $\mu < 0$. If $a^* \geq 1$, the result is clear. Now assume that $a^* < 1$. Let us suppose on the contrary that $a^m \geq a^*$. For any x, $H(x) – (1-a^m)x \geq H(x) – (1-a^*)x$. Since x^m maximizes $H(x) – (1-a^m)x$, we have $H(x^m) – (1-a^m)x^m \geq \text{Max}_x [H(x) – (1-a^*)x] \geq H(x^*) – (1-a^*)x^*$. Thus, $-D + H(x^m) – (1-a^m)x^m \geq -D + H(x^*) – (1-a^*)x^* = 0$. Then we have $U_0 < U_1$, implying $\mu > 0$, a contradiction. (B) The result that $x^m > x^*$ follows from $H'(x^m) = 1 – a^m < 1$ with $a^m > 0$. ///

The results are typical in the moral hazard literature. Let us provide intuitive explanations as follows. Full coverage is not desirable under moral hazard. No insurance is not desirable either, since without insurance the consumer is exposed to too high a risk. The insurance coverage under moral hazard is less than the first best coverage, since higher risk sharing lowers moral hazard costs. Part (B) states that moral hazard leads to a higher treatment than the first best treatment. With a lower level of coverage than the first best, the consumer is exposed to a high risk of a health loss. To compensate the health loss, the consumer will select a higher level of treatment than the first best.

Let us make additional comments on this proposition. First, note that the Lagrange multiplier μ is negative. This result implies that the insurer wants the consumer to reduce the treatment level, which confirms the moral hazard problem. Second, the proposition does not say that $x^m > x^{**}$. The relative sizes between x^m and x^{**} are ambiguous. For this, let us define b by $H'(x^{**}) = 1 – b$. Given that $H'(x^{**}) < 1$, $b > 0$. Thus, $x^m > x^{**}$ if and only if $a^m > b$.

From (3.6), $b = 1 - \lambda^{**}/U^{**}_1' = 1 – EV^{**}'/ U^{**}_1'$.

Under moral hazard, from (3.7),

$a^m = x^m H''(x^m)(1 – U^m_1'/EV^{m}')$.

Let us define $R^{**} = EV^{**}'/ U^{**}_1'$, and $R^m = EV^{m}'/U^m_1'$. Note that $R^{**}, R^m < 1$. With these notations, we have the following result.

Corollary 1:

$x^m > x^{**}$ if and only if $x^m H''(x^m)(1 – 1/R^m) > 1 – R^{**}$. (3.8) [proof] see the text above. ///

Although (3.8) is not of an explicit form, the following inferences can be made. First, as R^{**} is close to 1, x^m is likely to be greater than x^{**}. To see this, note that R^{**} will be close to 1 when the risk of the consumer under no moral hazard is well hedged. In such a case, x^{**} will be close to x^*. Since x^m is greater than x^*, x^m is likely to be greater than x^{**}. On the other hand, as R^m becomes close to 1, x^m is likely to be smaller than x^{**}. To see this, note that the treatment level is likely to be high (low) and the risk hedge level is low (high) if the moral hazard problem is severe (not severe). Thus, high R^m is likely to be associated with low x^m.

The above observation may look contradictory to the conventional belief that moral hazard leads to a higher treatment level. However, recall that $x^m > x^*$. The only factor to affect the difference betweeen x^* and x^{**} is the coverage constraint. In the no moral hazard case with the coverage constraint, risk is not fully hedged due to the coverage constraint. As a result, a higher level of treatment is needed to compensate for the incomplete risk hedge. If the risk hedge benefit is high enough, then the treatment level can be higher than that under moral hazard. In the appendix, we present an example of a CARA utility to show that x^m is greater than x^{**} when the risk hedge benefit is high. The example shows that x^{**} tends to be greater than x^m when risk aversion is greater, or when the health damage is larger.

Our next concern is the behavior of treatment as coverage changes under no moral hazard. Under moral hazard, the increase in coverage is followed by the increase in treatment from (3.3). However, under no moral hazard, the treatment may increase or decrease as coverage increases as shown below. This result may have an interesting implication on the debates on moral hazard in the health insurance, since the treatment increase following coverage increase is often interpreted as moral hazard (see section 5).

For this, let us fix coverage a. From $L_x = 0$ and $L_Q = 0$ in (3.4), we have

$$M(x, a) = U_1'(H' + a - 1) - [(1 - p)U_0' + pU_1']a = 0. \qquad (3.9)$$

Totally differentiating this expression with respect to a, we can find dx/da as follows.

$dx/da = - M_{xa}/M_{xx}$, where

$$M_{xx} = U_1''(H' + (1\text{-}p)a - 1)^2 + U_1'H'' + p(1\text{-}p)a^2U_0'',$$

$$M_{xa} = U_1''(1\text{-}p)x(H' + (1\text{-}p)a - 1) + p(1\text{-}p)axU_0'' + (1\text{-}p)(U_1' - U_0').$$

Since $M_{xx} < 0$, the sign of dx/da equals the sign of M_{xa}.

In M_{xa}, the first term is negative, since from $L_x = 0$ and $L_Q = 0$, $H' + a - 1 = a\lambda/U_1'$, implying that $H' + a - 1 - pa = a(1\text{-}p)U_0'/U_1' > 0$. The second term is also negative, while the third term is positive. Without further technical assumptions, the sign of M_{xa} is ambiguous. However, we can show that dx/da is decreasing near the first best coverage and is increasing near zero coverage.

Corollary 2:

Under no moral hazard, treatment may increase or decrease in coverage, in general. Treatment decreases in coverage, where coverage is close to the first best. Treatment increases in coverage, where coverage is close to zero.

[proof] See the text above for the ambiguity of sign of dx/da. At $a = a^*$, $dx/da < 0$ since $U_1' = U_0'$. When a is close to zero, $H' + a - 1$ is close to zero from $L_x = 0$. Thus, $dx/da > 0$ for small a, since $M_{xa} \approx (1-p)(U_1' - U_0') > 0$. ///

In Figure A.1 of the appendix, we present an illustrative graph of treatment as a function of coverage of the basis case under no moral hazard. It shows that treatment increases near zero coverage and decreases near the first best coverage ($a^* = 21.22$).

Before moving to the first period problem, let us present the following comparative statics that will be useful in the next section. For clarity, let us revive the time subscript. Recall that the probability of loss p_2 is affected by x_1. Considering the solutions as functions of x_1 and using the envelope theorem, we have the following.

$dEV_2^m/dx_1 = (\partial EV_2^m/\partial x_2)\partial x_2/\partial x_1 + (\partial EV_2^m/\partial a_2)\partial a_2/\partial x_1 + (\partial EV_2^m/\partial \lambda_2)\partial \lambda_2/\partial x_1 + (\partial EV_2^m/\partial \mu_2)\partial \mu_2/\partial x_1 + (\partial EV_2^m/\partial p_2)\partial p_2/\partial x_1 = (\partial EV_2^m/\partial p_2)\partial p_2/\partial x_1$

$= -p_2'U_{20} + p_2'U_{21} - \lambda_2 p_2' a_2 x_2 > 0.$ (3.10)

That is, the increase of treatment in period 1 has a positive effect on the expected utility in period 2. This result is intuitive, since an increase in x_1 lowers the loss probability in period 2.

4. First period problem

In the first period, the consumer selects the treatment x_1 and coverage a_1. The problem is similar to the second period problem, except that the consumer now needs to take into account the effect on the second period utility. As in period 2, the consumer selects treatment x_1 after the health loss occurs. However, the objective function includes not only the ex post utility, but also the discounted expected utility of period 2. That is, an optimal x_1 solves

Max $U(W - Q_1 - D_1 + H_1(x_1) - x_1 + a_1x_1) + \beta EV_2^m$, given Q_1 and a_1,

where EV_2^m is optimally determined, given x_1, in period 2, as described in the previous section. The first order condition for an interior solution is

$p_1U_{11}'\cdot(H_1' + a_1 - 1) + \beta\, dEV_2^m/dx_1 = 0,$

where dEV_2^m/dx_1 is from (3.10).

Now, the problem can be stated as follows.

$$\text{Max}_{a1,\, x1,\, Q1}\ EU = EV_1 + \beta EV_2^m = (1 - p_1)U_{10}(W - Q_1) + p_1U_{11}(W - Q_1 - D_1 + H_1(x_1) - x_1 + a_1x_1) + \beta EV_2^m \quad (4.1)$$

s.t. $Q_1 = p_1a_1x_1$

$p_1U_{11}'(H_1' + a_1 - 1) + \beta dEV_2^m/dx_1 = 0$

Note that $U_{10} = U(W - Q_1)$, and $U_{11} = U(W - Q_1 - D_1 + H_1(x_1) - x_1 + a_1x_1)$. To solve the problem, let us denote λ_1 and μ_1 for the Lagrange multipliers attached to constraints. For simplicity, let us suppress the time subscript 1 in this section, unless stated otherwise. Now, the Lagrangian can be expressed as follows.

$$L = (1 - p)U_0 + pU_1 + \beta EV_2^m + \lambda(Q - pax) + \mu[pU_1'(H' + a - 1) + \beta dEV_2^m/dx] \qquad (4.2)$$

First order conditions are as follows.

$$L_x = pU_1'(H' + a - 1) + \beta dEV_2^m/dx - \lambda pa + \mu[pU_1''(H' + a - 1)^2 + pU_1'H'' + \beta d^2EV_2^m/dx^2] \qquad (4.3)$$

$$= -\lambda pa + \mu[pU_1''(H' + a - 1)^2 + pU_1'H'' + \beta d^2EV_2^m/dx^2] = 0$$

$$L_a = pU_1'x - \lambda px + \mu[pU_1''(H' + a - 1)x + pU_1'] = 0$$

$$L_Q = -(1 - p)U_0' - pU_1' + \lambda - \mu pU_1''(H' + a - 1) = 0$$

$$L_\lambda = Q - pax = 0$$

$$L_\mu = pU_1'(H' + a - 1) + \beta dEV_2^m/dx = 0$$

=>

$$L_Q = 0 \Rightarrow \lambda = EV_1' + \mu pU_1''(H' + a - 1)$$

$$L_x = pU_1'(H' + a - 1) + \beta dEV_2^m/dx - \lambda pa + \mu[pU_1''(H' + a - 1)^2 + pU_1'H'' + \beta d^2EV_2^m/dx^2]$$

$$\Rightarrow \mu = \frac{EV_1' pa}{pU_1''(H'+a-1)\{H'+(1-p)a-1\} + pU_1'H''+\beta EV_2^{m}{}''} - pU_1'(H'+a-1) - \beta EV_2^{m}{}'$$

$$\lambda = \frac{EV_1'\{pU_1''(H'+a-1)^2 + pU_1'H''+\beta EV_2^{m}{}''\}}{pU_1''(H'+a-1)\{H'+(1-p)a-1\} + pU_1'H''+\beta EV_2^{m}{}''} - pU_1''(H'+a-1)\{pU_1'(H'+a-1) + \beta EV_2'\}$$

Let us add superscript "1m" for the solution. The solution to (4.3) has the following properties.

Proposition 2: [First period outcome]

(A) Insurance coverage is partial.

(B) Treatment is greater under the preventive treatment case than under the no-preventive treatment case.

[proof] (A) Although there is an additional term $\beta dEV_2^m/dx_1$, the logic of the proof of proposition 1(A) can be used to show that full or zero coverage cannot be optimal for x^{1m} to be an interior solution. (B) Let us first consider the first order conditions under no preventive effect. These first order conditions can be obtained from (4.3) by noting that $dEV_2^m/dx = 0$.

$$L_x' = -\lambda pa + \mu[pU_1''(H' + a - 1)^2 + pU_1'H''] = 0 \qquad (4.4)$$

$$L_a' = pU_1'x - \lambda px + \mu[pU_1''(H' + a - 1)x + pU_1'] = 0$$

$$L_Q' = -(1 - p)U_0' - pU_1' + \lambda - \mu pU_1''(H' + a - 1)= 0$$

$$L_\lambda' = Q - pax = 0$$

$$L_\mu' = pU_1'(H' + a - 1) = 0$$

Although the problem is the same as in the previous section, (4.4) is different from (3.3) since the incentive constraint is written differently: $pU_1'(H' + a - 1) = 0$, instead of $H' + a - 1 = 0$. The reason for this difference is to make a clear comparision between the preventive case and the no preventive case.

Since $dEV_2^m/dx > 0$, we should have $H' + a - 1 < 0$ under the preventive case. As a result, given a,

the treatment level will be higher under the preventive case than under no preventive case, cetris paribus.
By comparing (4.3) with (4.4), we can show that $x^{1m} \geq x^m$. Suppose on the contrary that $x^{1m} < x^m$. In such a case, a^{1m} should be lower than a^m. Now, let us slightly increase a^{1m} to a', so that $x^{1m} < x' < x^m$ where x' is the treatment corresponding to a'. Then, we reach a contradiction to the optimality of (a^{1m}, x^{1m}), since
$EV_1(a^{1m}, x^{1m}) + \beta EV_2(x^{1m}) < EV_1(a', x') + \beta EV_2(x^{1m}) < EV_1(a', x') + \beta EV_2(x')$.
Note that the first inequality follows from the fact that (a^m, x^m) maximixes $EV_1(a, x)$ and the second inequality follows from the fact that EV_2 is increasing in x. ///

The intuition of part (A) is the same as in propositon 1. Part (B) is also intuitive, since the increase in treatment has an additional positive effect on the future expected utility under the preventive case. By the similar reasoning as in part (B), we can show the more general fact that the optimal treatment will increase as β increases. In words, as the consumer concerns more about the future, she will select a higher level of treatment. This result is also intuitive.

Corollary 3:
A shortsighted consumer selects a lower level of treatment than a farsighted consumer, ceteris paribus.

On the other hand, the relative size of insurance coverage is not clearly determined.[3] However, an intuitive explanation can be obtained from the obsrvation of the first order conditions. Fixing a, let us denote $x = x(a)$, $\lambda = \lambda(a)$, and $\mu = \mu(a)$ be the corresponding solutions.[4] From the first order conditions, we have $a^{1m} > a^m$, if
$L_a|_{am} = pU_1'x - \lambda px + \mu[pU_1''(H' + a - 1)x + pU_1'] > 0$, where variables are evaluated at $a = a^m$. The first term is the benefit of the marginal coverage from the utility increase after the loss. The second term is the costs following the premium increase. These two terms can be referred to as the non-moral hazard net benefit of the marginal coverage. The third term measures the moral hazard costs from the increased coverage. Under the preventive case, the increase in coverage changes the costs from the moral hazard, since the increase of treatment has preventive effects. The net effect should weigh between the non-moral hazard net benefit and the moral hazard costs. When the former is greater than the latter, then the optimal coverage a^{1m} is greater than a^m, cetris paribus. Intuitively, as the preventive effect is greater, allowing more treatment is preferred. As a result, insurance coverage will be greater. However, if the preventive effect is small, the moral hazard costs may be greater than the non-moral hazard net benefit. In this case, insurance coverage will be reduced.

5. Discussion

In the second period problem, we find that the treatment level is generally higher under moral

[3] Determining the relative sizes of coverage will require complicated technical assumptions.
[4] We need to ignore $L_a = 0$ condition.

hazard than under no moral hazard. However, when overinsurance is not allowed, the treatment level can be lower under moral hazard. This finding may contradict the conventional belief that moral hazard leads to a higher utilization of resources. The treatment level is high under no moral hazard, since it helps hedge the risk. Without coverage constraint, risk is hedged by overinsurance. With the coverage constraint, a high coverage may be needed to lower the risk exposure. A main cause for this result is that the health damage is so high that the damage cannot be fully compensated in an affordable range of treatment. This result shares the similar spirit with Ellis and Manning (2007) arguing that uncompensated health costs lead to lower cost sharing.

It is important to note that although the treatment under moral hazard is lower than that under no moral hazard, it does not mean that an increase in the treatment level improves efficiency. This is because the treatment level is a second best one.

In the first period problem, we investigate the optimal treatment level, when treatment not only increases the health level, but also lowers the future possibility of health damage. The treatment level tends to be higher when treatment is preventive than when not. We also find that the treatment level is higher as the discount factor is higher.

The results have some implications on the current debates on the health insurance coverage in the U.S.A. (Gruber, 2008; Newhouse, 2006; Nyman, 1999b)

(i) Suppose that the current coverage level is set without taking into account the preventive characteristic of treatment. That is, the coverage is set at a^m. If treatment were not preventive, the treatment would be x^m. Given that treatment is preventive, the consumer selects treatment $x(a^m)$ which is higher than x^m. It would appear that excessive treatment or overinsurance exists. The insurer then may want to lower the coverage level to control the moral hazard problem. Lowering coverage will reduce the treatment level, as anticipated. It clearly reduces the health care costs spent in that period. However, it does not mean that efficiency is improved.

The effect of lowering coverage on efficiency depends on the relative sizes of a^{1m} and a^m. If $a^{1m} \geq a^m$, then reducing coverage will lower efficiency. However, if $a^{1m} < a^m$, then lowering coverage toward a^{1m} will increase efficiency. That is, controlling moral hazard does not necessarily improve efficiency, even though it lowers the short-term health care costs. This result emphasizes the importance of the prevention characteristic of treatment. When treatment is preventive, the relevant health care costs should be the long-term costs, incorporating the effects of treatment on the future health.

Corollary 4:
Suppose that the preventive characteristic is ignored in the health insurance design. Then, allowing moral hazard may improve efficiency.

(ii) Another case for moral hazard may be found if the consumer is shortsighted. Suppose that the consumer thinks of β as zero, while the true β is positive. For coverage given, the consumer selects a lower level of treatment under $\beta = 0$ than under $\beta > 0$. For example, with a^{1m}, the consumer selects x' that is less than x^{1m}. Now, there is less treatment. If a benevolent social planner knows the true β of the consumer, then it is possible to improve efficiency by increasing coverage, which leads to higher treatment. That is, allowing moral hazard may improve efficiency.

Corollary 5:
Suppose that consumers are more shortsighted than they should be. Then, allowing moral hazard may improve efficiency.

(iii) Finally, the discussion in the previous section (corollaries 1, 2) has an interesting implication on the debates on the moral hazard and health insurance coverage. For a highlighting example, suppose that there is no moral hazard problem. However, suppose further that health insurance is designed falsely assuming that a moral hazard problem exists. For example, let us assume that coverage is set at a^m. With a^m, the consumer, under no moral hazard, will select, say $x^*(a^m)$, not x^m. Now, the increase in coverage from a^m may be followed by the increase in treatment. Note, however, that it does not imply the existence of moral hazard (corollary 2). Moreover, if $x^{**} > x^*(a^m)$, then the increase in coverage leading to the increase in treatment may improve efficiency.

This case is illustrated in the appendix. In table A.1, we report $x^*(a^m)$ in a CARA utility example, showing that $x^{**} > x^*(a^m)$ in many cases. As an example, let us consider the base case. From table A.1, we know that $a^m = 0.71$, $a^* = 21.22$ and $a^{**} = 1$ in the base case. Figure A.1 depicts the treatment as a function of coverage under no moral hazard. In the range of low coverage, the increase in coverage will lead to the increase in treatment. Both treatment and coverage increase until (and after) full coverage. Since the (constrained) first best outcome is obtained with full coverage, efficiency is improved by increasing coverage toward one.

However, if the increase in treatment is interpreted as moral hazard, the insurer may not want to increase coverage resulting in a failure in improving efficiency.[5] This observation points out the possibility that the moral hazard problem may be overemphasized.

6. Conclusion

We consider a two-period model under moral hazard when treatment is also preventive. In the second period, we find the standard result that the treatment level under moral hazard is higher than that under no moral hazard. However, it may be lower than that under moral hazard, when overinsurance is not allowed. In the first period, the treatment level tends to be higher when treatment is preventive than when not. The treatment level is also higher as the discount factor is higher. These results imply that allowing moral hazard may improve efficiency of the health insurance market if the original insurance contract is designed ignoring the preventive characteristics of treatment or if the consumers are myopic. We also show that the increase of treatment following the coverage increase does not necessarily imply moral hazard.

[5] See, for example, Gruber (2008) and Manning and Marquis (1996) for the discussion of the deadweight loss due to the negative elasticity of the health care demand with respect to its price.

These findings imply that the moral hazard problem is possibly overemphasized in literature.

Appendix

We provide a numerical example for optimal treatments and insurance coverage under moral hazard and no moral hazard.

Assumptions and notations

$U(W) = -\exp(-AW)$, where A is a constant.

$U_0 = U(W - Q) = -\exp\{-A(W\text{-}Q)\}$

$U_1 = U(W – Q – D + H(x) – (1 – a)x) = -\exp\{-A(W – Q – D + H(x) – (1 – a)x)\}$

$H(x) = h\sqrt{x}$, where h is a positive constant.

$H'(x) = (1/2)hx^{-1/2} = h/(2\sqrt{x})$.

$H''(x) = -(1/4)\, hx^{-3/2} = -h/(4x\sqrt{x})$.

$Q = pax$

$EV = (1\text{-}p)U_0 + pU_1$.

Let us first find the outcomes under moral hazard. The program under moral hazard can be stated as follows.

Max EV (A.1)

s.t. $Q = pax$

$H'(x) + a – 1 = 0$.

From the second constraint, we have

$a = 1 – H'(x) = 1 - h/(2\sqrt{x})$. (A.2)

Thus, $Q = pax = p[1 - h/(2\sqrt{x})]x = p[x - h\sqrt{x}/2]$. (A.3)

We also have

$Q' = dQ/dx = p[1 - h/(4\sqrt{x})]$.

$H(x) - (1\text{-}a)x = h\sqrt{x} - h\sqrt{x}/2 = h\sqrt{x}/2$.

$d\{H(x) - (1\text{-}a)x\}/dx = h/(4\sqrt{x})$.

Now, let us replace Q and a in EV with the above results.

$EV = -[(1\text{-}p)e^{-A(W\text{-}Q)} + pe^{-A(W – Q – D + H(x) – (1 – a)x)}]$ (A.4)

$EV' = dEV/dx = -[(1\text{-}p)e^{-A(W\text{-}Q)}(aQ') + pe^{-A(W – Q – D + H(x) – (1 – a)x)}(AQ' - A\, d\{H(x) – (1\text{-}a)x\}/dx)] = 0$.

$\Rightarrow -[(1\text{-}p)e^{-A(W\text{-}Q)}(Q') + pe^{-A(W – Q – D + H(x) – (1 – a)x)}(Q' - d\{H(x) – (1\text{-}a)x\}/dx)] = 0$.

$\Rightarrow -e^{-A(W\text{-}Q)}\ [(1\text{-}p)\ (Q') + pe^{-A(– D + H(x) – (1 – a)x)}(Q' - d\{H(x) – (1\text{-}a)x\}/dx)] = 0$.

$\Rightarrow (1\text{-}p)p[1 - h/(4\sqrt{x})] + pe^{-A(– D + h\sqrt{x}/2)}(\ p[1 - h/(4\sqrt{x})] - h/(4\sqrt{x})) = 0$

$\Rightarrow (1\text{-}p)[1 - h/(4\sqrt{x})] + e^{-A(– D + h\sqrt{x}/2)}(\ p[1 - h/(4\sqrt{x})] - h/(4\sqrt{x})) = 0$

$$=> (1-p)(4\sqrt{x} – h) + e^{-A(– D + h\sqrt{x}/2)}(p[4\sqrt{x} – h] - h) = 0$$

$$=> (1-p)(4\sqrt{x} – h)[1-p + pe^{-A(– D + h\sqrt{x}/2)}] - he^{-A(– D + h\sqrt{x}/2)} = 0 \qquad (A.5)$$

In sum, from (A.2) and (A.5), the solution under moral hazard is determined as follows:

x^m solves (A.5) and $a^m = 1 - h/(2\sqrt{x^m})$.

Now, let us consider a case of no moral hazard. The first best outcome will solve the following program.

Max EV (A.6)

s.t. Q = pax

Let us replace Q with pax in EV and solve the first order conditions.

$\partial EV/\partial x = 0$

$$=> -e^{-A(W-pax)} [(1-p)pa + pe^{-A(– D + h\sqrt{x} – (1 – a)x)}(pa - \{h/(2\sqrt{x}) – (1-a)\})] = 0. \qquad (A.7)$$

$\partial EV/\partial a = 0$

$$=> -[(1-p)e^{-A(W- pax)} + pe^{-A(W – pax – D + h\sqrt{x} – (1 – a)x)}]$$

$$-[(1-p)e^{-A(W- pax)}(apx) + pe^{-A(W – pax – D + h\sqrt{x} – (1 – a)x)}(A)(p-1)x)] = 0$$

$$=> -[(1-p)e^{-A(W- pax)}p - pe^{-A(W – pax – D + h\sqrt{x} – (1 – a)x)}(1-p)] = 0$$

$$=> -[e^{-A(W- pax)} - e^{-A(W – pax – D + h\sqrt{x} – (1 – a)x)}] = 0$$

$$=> – D + h\sqrt{x} – (1 – a)x = 0 \qquad (A.8)$$

From (A.8) to (A.7),

$$(A.7) => [(1-p)pa + p(pa - \{h/(2\sqrt{x}) – (1-a)\})] = 0$$

$$=> a - h/(2\sqrt{x}) + (1-a) = 0$$

$$=> -h/(2\sqrt{x}) + 1 = 0 \qquad (A.9)$$

From (A.8) and (A.9) the solution under no moral hazard is determined as follow:

$x^* = h^2/4,$ (A.10)

$a^* = (4/h^2)D – 1.$ (A.11)

Note that $a^* > 1$ if $D > h^2/2$.

Assumging that $D> h^2/2$, the coverage a* exhibits overinsurance. If we do not allow overinsurance, we have the following (constrained first best) outcome.

$a^{**} = 1.$ (A.12)

x** solves EV' = 0

=>

$-e^{-A(W-pax)}$ [(1-p)pa + $pe^{-A(-D + H(x))}$(pa - H'(x)] = 0.

=> $-e^{-aA(W-pax)}$ [(1-p)p + $pe^{-A(-D + h\sqrt{x})}$(p - h/(2√x))] = 0.

=> (1-p) + $e^{-A(-D + h\sqrt{x})}$(p - h/(2√x)) = 0. (A.13)

Table A.1 provides some numerical results for some parameter values. The table shows that x^m becomes lower than x** as the risk aversion (A) is higher or as the health damage (D) is higher.

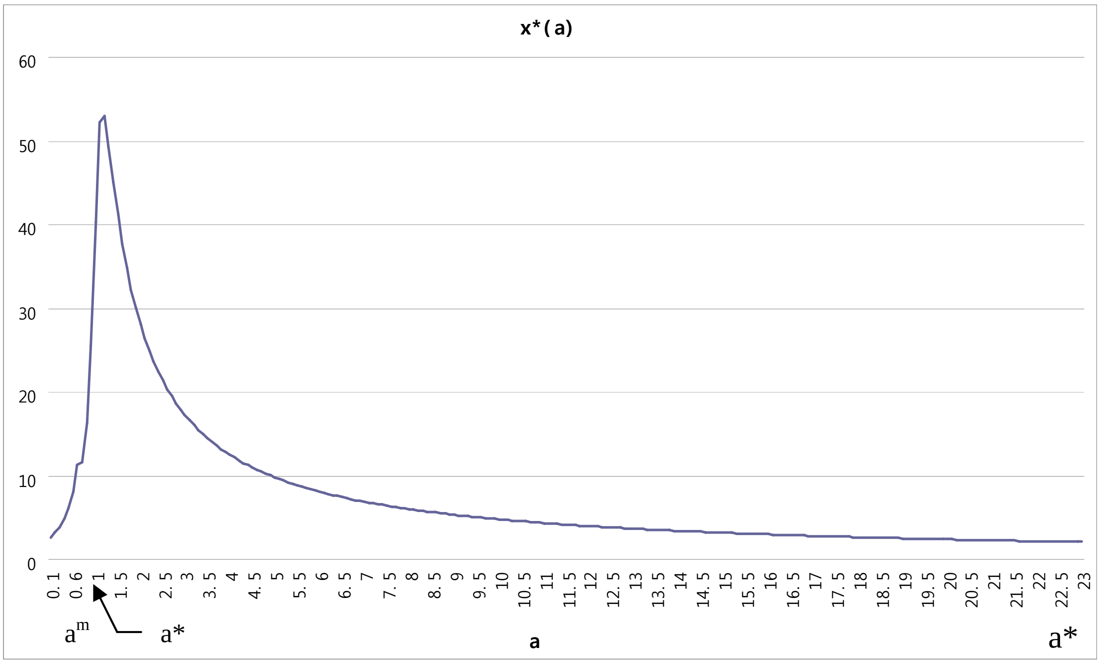
Table A.1

A numerical example

Para-meter	Base Case	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2	D3
p	0.200	0.200	0.200	0.200	0.100	0.300	0.400	0.200	0.200	0.200	0.200	0.200	0.200
A	0.100	0.200	0.050	0.005	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100	0.100
h	3.000	3.000	3.000	3.000	3.000	3.000	3.000	1.000	5.000	10.000	3.000	3.000	3.000
D	50.000	50000	50.000	50.000	50.000	50.000	50.000	50.000	50.000	50.000	10.000	30.000	70.000
Out-come													
x^m	26.812	29.519	15.985	3.523	59.795	17.798	14.592	3.125	61.066	68.692	5.183	17.359	29.153
x^{**}	40.323	54.570	20.675	3.189	71.790	22.068	13.252	5.842	46.858	25.000	4.104	18.778	52.789
x^*	2.250	2.250	2.250	2.250	2.250	2.250	2.250	0.250	6.250	25.000	2.250	2.250	2.250
$x^*(a^m)$	11.643	12.700	7.339	2.410	27.058	7.319	5.497	1.348	27.478	40.094	3.002	7.932	12.547
a^m	0.710	0.724	0.625	0.201	0.806	0.644	0.607	0.717	0.680	0.397	0.341	0.640	0.722
a^{**}	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
a^*	21.222	21.222	21.222	21.222	21.222	21.222	21.222	199.00	7.000	1.000	3.444	12.333	30.111

Figure A.1

Treatment as a function of coverage in the base case under no moral hazard: x*(a)



References

Barigozzi, F. (2004), Reimbursing Preventive Care, *Geneva Papers on Risk and Insurance Theory* 29: 165-186.

Blomqvist, A. (2001), Does the Economics of Moral Hazard Need to Be Revisited? A Comment on the Paper by John Nyman, *Journal of Health Economics* 20: 283-288.

Ehrlich, I. and G. S. Becker (1972), Market Insurance, Self-insurance, and Self-protection, *Journal of Political Economy* 80: 164-189.

Ellis, R. P. and W. G. Manning (2007), Optimal Health Insurance for Prevention and Treatment, *Journal of Health Economics* 26: 1128-1150.

Feldstein, M. (1973), The Welfare Loss of Excess Health Insurance, *Journal of Political Economy* 81: 251-280.

Feldstein, M. and B. Friedman (1977), Tax Subsidies, the Rational Demand for Health Insurance, and the Health Care Crisis, *Journal of Public Economics* 7: 155-178.

Gruber, J. (2008), Covering the Uninsured in the United States, *Journal of Economic Literature* 46: 571-606.

Kenkel, D. S. (2000), Prevention, in Culyer, A. J. and J. P. Newhouse (Eds.), *Handbook of Health Economics* (Amsterdam: North Holland): 1675-1719.

Manning, W. G. and M. S. Marquis (1996), Health Insurance: the Tradeoff between Risk Pooling and Moral Hazard, *Journal of Health Economics* 15: 609-639.

Manning, W. G. and M. S. Marquis (2001), Health Insurance: Tradeoffs Revisited, *Journal of Health Economics* 20: 289-293.

Newhouse, J. P. (2006), Reconsidering the Moral Hazard-Risk Avoidance Tradeoff, *Journal of Health Economics* 25: 1005-1014.

Nyman, J. A. (1999a), The Value of Health Insurance: The Access Motive, *Journal of Health Economics* 18: 141-152.

Nyman, J. A. (1999b), The Economics of Moral Hazard Revisited, *Journal of Health Economics* 18: 811-824.

Pauly, M. V. (1968), The Economics of Moral Hazard: Comment, *American Economic Review* 58: 531-537.

Pauly, M. V. (1974), Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection, *Quarterly Journal of Economics* 88: 44-62.

Pauly, M. V. (1986), Taxation, Health Insurance, and Market Failure, *Journal of Economic Literature* 24: 629-675.

Shavell, S. (1979), On Moral Hazard and Insurance, *Quarterly Journal of Economics* 93: 541-562.

U.S. Census Bureau (2008), *Income, Poverty, and Health Insurance Coverage in the United States: 2007*, http://www.census.gov/prod/2008pubs/p60-235.pdf.

Zeckhauser, R. (1970), Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives, *Journal of Economic Theory* 2: 10-26.

Zweifel, P. and F. Breyer (1997), *Health Economics* (New York, NY: Oxford University Press).