Abstract

In this paper, we view health insurance as a combined hedge against the two consequences of falling ill: treatment expenditures and loss in income due to permanent impairment. We discuss how an individual’s ability when healthy affects her decision on whether to buy health insurance with treatment to full recovery if ill or with partial treatment combined with cash compensation for the resulting loss in income. We find that a high-ability individual demands full recovery and equalise utility across states, while a low-ability individual demands partial treatment and cash compensation and suffers a loss in utility if ill.

JEL Nos.: I11, G22, D81

1 Introduction

Who are the individuals choosing to be partially disabled and to live from disability payments if ill, and who are the individuals choosing to fully regain health if ill? This is the focus of our paper. In particular, we study how different types of individuals prefer to be compensated if illness occurs.

---

*We are grateful to two anonymous referees of this journal for their valuable comments. We would also like to thank Agnar Sandmo, Sören Blomquist, and participants at the 1999 EEA Conference and the 1999 Oslo workshop on health economics, for helpful comments on earlier versions. Part of Asheim’s and Nilssen’s research is financed by the Research Council of Norway through HERO - The Health Economics Research Programme at the University of Oslo. Emblem gratefully acknowledges financial support from the SiS Programme at Agder University College.

†Correspondence: Anne Wenche Emblem, School of Management, Agder University College, Servicebox 422, N-4604 Kristiansand, Norway. Fax: +47 38 14 10 27.
The risk of falling ill is an inevitable fact of life. Illness entails both monetary and non-monetary consequences. First, individuals may purchase medical treatment to alleviate the adverse health effects of the illness, i.e., they incur medical expenditures. Second, illness alters individuals’ productivity in the labour market, and may in severe cases lead to a disabling impairment. Illness consequently restricts individuals’ earning potential. Last, illness implies non-monetary losses to the individuals. To be in ‘good’ health is of value in itself and influences the utility derived from other activities, e.g., from consumption.

Adequate medical treatment is available for a number of illnesses. Some diseases are indeed curable so that functionality can be fully restored to its pre-illness level, provided that the appropriate quantity and quality of treatment is undertaken. Individuals may thus avoid potentially disabling consequences of an illness by undertaking the appropriate level of medical treatment. If less than complete medical treatment is undertaken, then less than complete recovery is attained. In this case, individuals will not fully recuperate and their earning potential is reduced accordingly. Consequently, individuals’ choice with respect to what level of treatment to undertake is not only a choice regarding which extent to retrieve health, but also a choice regarding which extent to restore earning potential.

Our paper is motivated by the empirical fact that health insurance and disability insurance are integrated in a number of real-life health care systems. This is particularly prevalent in European countries where health insurance with in-kind compensation and disability insurance with cash compensation typically form parts of a public tax-financed (social) insurance system. Also in the United States, public programs include medical insurance and disability payments: low-income (poor) individuals are insured through a public program against both medical expenditures (e.g., through Medicare or Medicaid) and loss in income due to disability (e.g., through public disability programs or income support programs toward disabled such as Social Security Disability Insurance (SSDI), Supplementary Security Income (SSI), or state-based Workers’ Compensation systems). High-income individuals in the US are, on the other hand, typically (privately) insured against medical expenditures but only few have private disability insurance. Moreover, whereas low-income individuals falling into the Medicare or Medicaid program are usually considered to be restrained in the level (and quality) of health care provided, high-income individuals with private health insurance have access to high-quality care. Thus, we observe that different income groups are compensated differently if ill.1 It is an empirical fact, both in the US and in Europe, that individuals with less schooling are more likely to be disabled than those with more schooling, see for instance Haveman and Wolfe (2000). Naturally, there are many explanations to why this is so: different proneness to illness and thus different educational ‘carriers’, different types of work and thus different exposure to health risks, etc. The causal effects are also not readily apparent. We do not aspire to provide a definite answer to this challenging question, but hope to shed some light on the question by

---

1This is generally not the case in European countries with social insurance.
studying how individuals’ inherent earnings capabilities influence their decision regarding how to be compensated if ill. As will be shown, how individuals choose to be compensated if ill, determines whether they will fully recuperate from an illness or suffer a (partial) disabling impairment.

In this paper, we study a situation in which individuals face a risk of suffering an illness that, if left untreated, will fully impair their functionality in the labour market, i.e., they will be disabled. Health may however be fully restored to its pre-illness level if the appropriate level of medical treatment is undertaken. Medical treatment here includes all health-promoting medical activities leading to full recuperation of health. Moreover, these activities all take place at the hospital. We thus do not consider issues related to the need for rehabilitation (and sickness payments) after being discharged from hospital. To simplify, we assume that poor health improves instantaneously, and with certainty, when medical treatment is provided. Thus, sick individuals will immediately recuperate, in part or in full, if they receive partial or complete medical treatment. Our model is consequently atemporal and we do not consider temporary losses in income.

Naturally, individuals’ decisions regarding how much treatment to undertake if ill determine the economic consequences of falling ill: medical expenditures, loss in earnings, or both. Individuals would like to hedge against these consequences. In this paper, we allow for insurance contracts offering compensation of not only medical expenditures but also income losses due to illness-induced impairment. Hence, we expand the concept of health insurance. Traditionally, individuals are thought to insure against potential medical expenditures by holding a medical insurance and to insure against potential loss in income by holding a disability insurance. We argue that the two types of insurance provide coverage against different consequences of the same risk, namely that of suffering a loss in health. It is, hence, appropriate to consider insurance contracts that offer individuals compensation in the form of (i) medical treatment, (ii) cash compensation of income loss, or (iii) both. To our knowledge, the health-insurance literature does not discuss compensation of (permanent) income loss due to illness, i.e., disability payment. Little is thus known about individuals’ choice between different types of compensation in the case of illness. Indeed, Pauly (1986) recognizes the absence of studies of the relationship between individuals’ demand for medical insurance and their demand for insurance that pays cash if illness occurs. We show that by integrating what is traditionally considered a health insurance and disability insurance and solving for the optimal design of insurance, then individuals may indeed trade off health for consumption. In particular, we show that individuals who initially have low productivity will choose type (iii) compensation whereas individuals who initially have high productivity choose type (i).

In the literature on health insurance, focus is mainly placed on insurance against medical expenditures, assuming that illness entails only monetary losses (e.g., medical expenditures and temporary loss in earnings). The desire to re-

\footnote{As examples of the latter, Pauly (1986) mentions, among others, salary continuation insurance, disability insurance and life insurance (see his note 4).}
store health is, by and large, taken for granted. When non-monetary consequences of illness are taken into account, it is assumed that utility is state dependent and that health is either not restorable or irreplaceable. In this paper, we allow for both monetary and non-monetary consequences of illness without imposing the assumption that health is irreplaceable. Rather, we assume that health if ill is endogenous: health if ill is improved if individuals receive medical treatment. We assume that treatment is divisible, and that individuals can choose to which extent their health is to be restored (with certainty) if ill. Health is thus insurable and the non-monetary consequences of illness endogenous. Our model may thus provide a bridge between models taking only monetary consequences of illness into account, and models postulating that health is irreplaceable. A recent paper by Flochel and Rey (2002) supplements our analysis in that they study individuals’ ex post demand for health care and, given the level of health care demanded, derive their optimal level of insurance against medical expenditures. Like us, they assume that utility is a function of both consumption and health. However, they do not allow labour earnings to depend on health state, and so their analysis does not encompass disability insurance.

Our main analysis takes place in a world of symmetric information about health risks and health states. However, we show that our findings are robust, even in a situation where state of health is not verifiable, i.e., when insurers face problems of ex-post moral hazard. This is so since the integration of in-kind provision of medical treatment and cash compensation reduces an insured individual’s incentive to falsely claim to be ill when in good health. Hence, integrating medical treatment, i.e., medical insurance, and cash compensation, i.e., disability insurance, is not intrinsic to a public health-care system, but may also grow out of a totally unregulated system. Indeed, this possibility may be viewed in light of the increasing importance in the US during the last decades of Health Maintenance Organizations (HMOs) where financing and supply of medical treatment are integrated (mainly) in order to reduce moral-hazard problems. Our findings suggest that also financing of income loss due to illness could form part of such schemes. In fact, if ex-post moral hazard is a problem, then integrating medical insurance (with in kind provision) and disability insurance (with cash compensation) may induce self-selection. Integrating the two may indeed incite a larger private provision of disability insurance.

---

3 The only exceptions we know of are Byrne and Thompson (2000) and Graboyes (2000), who argue that, when the probability of successful treatment is small, the insured may be better off with cash compensation if ill, rather than going through the treatment.

4 Analyses based on health being non-restorable include Zeckhauser (1970), Arrow (1974), Viscusi and Evans (1990), Evans and Viscusi (1991), and Frech (1994). Health is irreplaceable if individuals value restored health lower than pre-illness health; see Cook and Graham (1977) and Schlesinger (1984).

5 More than 70 percent of the insured in US were enrolled in some form of managed care plan in 1993 (Glied, 2000). The term managed care organizations comprises organizations that mediate between the insured and the providers of care, e.g., by regulating the services available, as in HMOs, or that restrict the insured’s choice of providers, e.g., as in independent practice associations (IPAs) and preferred provider organizations (PPOs) (Glied, 2000).
The outline of this paper is as follows. The model is presented in Section 2 and a preliminary analysis is provided in Section 3. Our main findings are derived in Section 4. In Section 5, we discuss the case where health states are not verifiable, i.e., ex-post moral hazard. Our results are discussed in a concluding Section 6.

2 The model

Consider an individual who has preferences over consumption, \( c \), and health, \( h \). The individual faces exogenous uncertainty with respect to her state of health. She may either be healthy, which corresponds to state 1, or she may fall ill, which corresponds to state 2. The two states are mutually exclusive, jointly exhaustive, and verifiable. In state 1, the level of health is normalized to 1: \( h_1 = 1 \). In state 2, the individual is ill and suffers a complete loss in health: \( h_2 = 0 \). Health if ill may, however, be partially or fully restored (with certainty) if the individual receives medical treatment: \( t \in [0, 1] \), i.e., treatment is assumed to be a continuous variable. Medical treatment leading to full recovery is available at cost \( C \), while treatment at cost \( tC \) leads to partial recovery. It follows that if an individual receives treatment at a level leading to complete recovery, i.e., \( t = 1 \), then her level of health if ill is equal to 1: \( h_2 = 1 \). If no treatment is received, then \( t = 0 \) and health equals zero: \( h_2 = 0 \). Health in state 2 is henceforth measured by the fraction of \( C \) spent on treatment, that is, \( h_2 = t \).

Consumption in the two states are denoted \( c_1 \) and \( c_2 \), respectively.

The objective probability of falling ill is known to the individual and given by \( \pi \), where \( 0 < \pi < 1 \). The individual seeks to maximize the von Neumann-Morgenstern expected utility:

\[
(1 - \pi)u(c_1, 1) + \pi u(c_2, t),
\]

where \( u(c, h) \) is a Bernoulli utility function. We assume that \( u : \mathbb{R}_+^2 \to \mathbb{R} \) is twice continuously differentiable, strictly concave, and satisfies: \( \forall (c, h) \in \mathbb{R}_+^2, u_c > 0 \) and \( u_h > 0 \), where partial derivatives are denoted by subscripts. In particular, a strictly concave \( u \) implies that the individual is risk averse. Moreover, health and consumption are assumed to be complements in utility: \( u_{ch} > 0 \). This implies that individuals take more pleasure in consumption when health is good than when health is poor. We also assume that \( u_c(c, h) \to \infty \) as \( c \downarrow 0 \) whenever \( h > 0 \), \( u_h(c, h) \to \infty \) as \( h \downarrow 0 \) whenever \( c > 0 \), and \( u_c(c, h) \to \infty \) or \( u_h(c, h) \to \infty \) as \( c \downarrow 0 \) and \( h \downarrow 0 \). Note that our assumptions on \( u \) imply normality. As specified in the above expected utility function, utility from treatment is positive and decreasing (\( u_h > 0 \) and \( u_{hh} < 0 \)). This implies that marginal utility from treatment is higher at a low treatment ratio than at a high such ratio. (Hence, if we would measure treatment in terms of utility, then cost of treatment would be strictly convex.) Also, it follows from the properties of \( u \) that the individual

---

6 The cost of curing an illness is assumed to depend on the characteristics of the illness, rather than the characteristics of the individual suffering from it. Since all individuals face the same health risk, the cost of treatment is constant across individuals.
prefers to receive an intermediate level of treatment to an uncertain prospect of receiving either complete or zero treatment with the same expected cost.

There exists a competitive insurance market in which profit maximizing insurers offer insurance at an actuarially fair premium. Information about the individual’s probability of falling ill ($\pi$), which disease she is suffering from, and consequently, the associated costs of treatment, is symmetrically distributed among the market participants. Moreover, health state is verifiable, thus insurance policies can be made contingent on it. We hence for now rule out problems of ex-ante and ex-post moral hazard. (A situation where health state is non-verifiable is discussed in Section 5). The market for health insurance will, therefore, be efficient.

The individual’s inherent capacity to generate earnings is henceforth referred to as ‘ability’, and is given by $A$. If well, the individual enjoys her full ability, $A$. We choose our monetary unit so that total earnings are equal to $A$ when well. Hence, labour earnings in state 1 are given by $A$. Since, by assumption, leisure is not included in the utility function, we implicitly assume that leisure is constant (and thus, labour supply is fixed) across states. We assume that the individual, by spending $tC$ on medical treatment, will generate earnings equal to $tA$ when ill. Hence, labour earnings in state 2 are proportional to the rate of cost spent on treatment (recalling that $t$ gives the rate of total treatment cost ($C$) spent on recovery). Note that the following analysis does not require insurance companies to know the individual’s ability when healthy; hence $A$ may be private information.

In this model, there are two types of consequences of an illness. Firstly, the individual suffers a loss in earnings due to reduced ability (productivity) and entails medical expenditures due to illness, i.e., financial losses. Secondly, the individual suffers a direct loss in utility because of lower health (since $u_h > 0$), i.e., a non-financial loss. Since health if ill is endogenous, the size of the non-financial loss is also endogenous. Indeed, if $t = 1$, then she suffers a financial loss only, viz., the costs of treatment, while if $t < 1$, then she suffers both a financial and a non-financial loss. Consequently, a positive level of treatment reduces the non-financial loss from illness relative to a situation without treatment. In fact, if complete treatment is received, the non-financial loss is eliminated.

The risk-averse individual wishes to insure against the consequences of falling ill. Her insurance decision takes place prior to her knowing which state has occurred. Her budget constraints in states 1 and 2 are respectively given by:

$$c_1 + \pi I = A,$$

and

$$c_2 + \pi I + tC = tA + I,$$

where $\pi I$ denotes the insurance premium to be paid in both states of the world in order to receive compensation equal to $I$ if ill. As before, $A$ and $tA$ respectively give labour earnings in state 1 and 2. From this, it follows that

$$I = tC + c_2 + (c_1 - A) - tA,$$
that is, the compensation consists of a compensation of medical expenditures equal to: $tC$, and a cash compensation equal to: $(c_2 + \pi I - tA)$.\(^7\) Insurance is the only way to transfer income across the two states of the world. Combining the two budget constraints, it follows that the individual is constrained by:

$$A - c_1 = \pi [tC + (c_2 - tA + A - c_1)]$$

when \textit{ex ante} making her choice of $c_1$, $c_2$, and $t$.

We make the additional assumption about the utility function $u$ that the marginal rate of substitution of consumption for health, $u_h/u_c$, is higher at full than at partial recovery from illness, given that the expected cost of treatment is subtracted:

$$\frac{u_h(tA - \pi tC, t)}{u_c(tA - \pi tC, t)} \leq \frac{u_h(A - \pi tC, 1)}{u_c(A - \pi tC, 1)}.$$

This assumption means that, for a fixed relative price of health in terms of consumption across states, the individual wants to shift the expected cost of treatment towards the healthy state if treatment does not lead to complete recovery (i.e., if $t < 1$). A homothetic utility function satisfies this for any non-negative expected cost of treatment, but the assumption is also satisfied by other demand systems.

In the following, we characterize the individual’s demand for insurance, both with respect to level, and type of, coverage. In particular, we analyze how the individual’s level of inherent ability, $A$, influences her choice of compensation: whether to be compensated in the form of health restoration, i.e., medical treatment, and/or in the form of cash, i.e., compensation of loss in income due to incomplete recovery.

### 3 Preliminary analysis

As explained earlier, we assume that treatment leading to a health level $t$ is available at a cost $tC$ when ill. For the purpose of our analysis, however, let us be more general and ask what is the maximum utility achievable if the individual has to pay $P (\geq 0)$ for treatment $t$:

$$U(t, P, A) := \max_{(c_1, c_2)} \left\{ (1 - \pi)u(c_1, 1) + \pi u(c_2, t) \right\}$$

s.t. $$(1 - \pi)c_1 + \pi(c_2 + P) = (1 - \pi)A + \pi tA,$$

where $U : \mathbb{R}^+ \times [0, (1/\pi - (1 - t))A] \times \mathbb{R}^+ \to \mathbb{R}$. The individual is offered a positive level of treatment that may, for the purpose of defining and analyzing the $U$ function, exceed one. As specified, the maximum price she is able to pay for this level of $t$ is given by $(1/\pi - (1 - t))A$, hence the price of treatment, $P$, satisfies: $0 \leq P < (1/\pi - (1 - t))A$. Naturally, the higher the level of inherent

\(^7\)Since the premium $A - c_1$ must be paid in both states, disposable income net of the premium equals $tA - (A - c_1)$ if no cash compensation is received. Hence, the cash compensation equals $c_2 - [tA - (A - c_1)]$. 
ability \((A)\), the higher the price she can pay for treatment. Also, the higher the probability of falling ill, the less she is able to pay for treatment.

To investigate the optimization problem, form the corresponding Lagrangian:

\[
L(c_1, c_2, \lambda; t, P, A) = (1 - \pi)u(c_1, 1) + \pi u(c_2, t) + \lambda[(1 - (1 - t)\pi)A - (1 - \pi)c_1 - \pi(c_2 + P)].
\]

Given our assumptions on \(u\), the first-order necessary conditions (FOCs) give the consumption demand function in each of the two states of the world:

\[
(c_1(t, P, A), c_2(t, P, A)) \in \mathbb{R}^2_{++},
\]

satisfying

\[
u_c(c_1(t, P, A), 1) = u_c(c_2(t, P, A), t) = \lambda
\]

and the constraint (cf. equation (1)). As shown above, optimal consumption in each of the two states of the world is a function of treatment \((i.e., \text{the degree of recovery in state 2})\), price of treatment \((P)\), and income \((A)\). Equation (3) implies that, in optimum, the individual’s marginal utility of consumption is equal in the two states.

The utility function \(U\) can now be written:

\[
U(t, P, A) = (1 - \pi)u(c_1(t, P, A), 1) + \pi u(c_2(t, P, A), t).
\]

We have that \(U\) is strictly increasing in \(t\), strictly decreasing in \(P\), and strictly increasing in \(A\). Hence, we can define an indifference curve in \((t, P)\)-space, call it \(\mathcal{P}(t, A; \bar{t}, \bar{P})\), going through \((\bar{t}, \bar{P})\) and showing combinations of \(t\) and \(P\) yielding a constant level of utility. Hence, \(U(t, P, A)\) is equal to \(U(\bar{t}, \bar{P}, A)\) if and only if \(P = \mathcal{P}(t, A; \bar{t}, \bar{P})\). The slope of the indifference curve is given by:

\[
\frac{\partial \mathcal{P}(t, A; \bar{t}, \bar{P})}{\partial t} = -\frac{\partial U}{\partial t} = -\frac{\partial c}{\partial t} = \frac{\pi(u_c(c_2, t) + \lambda A)}{\pi \lambda} = \frac{u_h(c_2, t)}{u_c(c_2, t)} + A, \quad (4)
\]

where the second equality follows from the envelope theorem, and the fourth equality is implied by equation (3). This means that the marginal willingness to pay for treatment is equal to the marginal rate of substitution of consumption for health plus the additional earnings capacity generated by treatment. Since, by construction, \(\mathcal{P}(t, A; \bar{t}, \bar{P})\) is the indifference curve going through \((\bar{t}, \bar{P})\), it follows that:

\[
\frac{\partial \mathcal{P}(\bar{t}, A; \bar{t}, \bar{P})}{\partial A} = 0.
\]

\(I.e.,\) even though the indifference curve through \((\bar{t}, \bar{P})\) will shift when ability increases, it will still go by \((\bar{t}, \bar{P})\). Since \(u_{c_1} < 0, u_{c_2} > 0\), and \(\partial c_2/\partial A > 0\), it follows that the shift will be an anti-clockwise rotation, with \((\bar{t}, \bar{P})\) as fixed point, so that the slope at \((\bar{t}, \bar{P})\), \(\partial \mathcal{P}(t, A; \bar{t}, \bar{P})/\partial t\), increases:

\[
\frac{\partial \mathcal{P}(t, A; \bar{t}, \bar{P})}{\partial A} = \frac{\partial}{\partial A} \left[ \frac{u_h(c_2(t, \bar{P}, A), t)}{u_c(c_2(t, P, A), t)} + A \right] > 1.
\]
Hence, the slope of an indifference curve through any point \((\bar{t}, \bar{P})\) is increasing in ability \(A\). We will refer to this as the single-crossing property. The single-crossing property is illustrated in Figure 1 for two different values of ability, \(A_l < A_h\), where \(l\) and \(h\) denote low and high ability, respectively. As illustrated, the slope of the indifference curve going through \((\bar{t}, \bar{P})\) in \((t, P)\)-space is steeper if ability is high than if it is low.

It remains to show that \(P(t, A; \bar{t}, \bar{P})\) is a strictly concave function of \(t\), so that an individual being faced with the possibility of purchasing treatment at cost \(P = tC\) constrained by \(t \leq 1\) will have a unique level of treatment maximizing \(U(t, tC, A)\). This will be shown by demonstrating that, if \((t', P')\) and \((t'', P'')\) are different combinations yielding the same utility level given \(A\), then any interior convex combination

\[
(t, P) = (\alpha t' + (1 - \alpha)t'', \alpha P' + (1 - \alpha)P'') , \quad 0 < \alpha < 1,
\]

will yield a strictly higher utility level. Accordingly, assume \(U(t', P', A) = U(t'', P'', A) = U(\bar{t}, \bar{P}, A)\), and introduce some notation:

\[
\begin{align*}
    c'_1 &= c_1(t', P', A) & c''_1 &= c_1(t'', P'', A) \\
    c'_2 &= c_2(t', P', A) & c''_2 &= c_2(t'', P'', A).
\end{align*}
\]

Also, let \((c_1, c_2) = (\alpha c'_1 + (1 - \alpha)c''_1, \alpha c'_2 + (1 - \alpha)c''_2)\). Since \((c'_1, c_2)\) satisfies the ex-ante budget constraint (1) given \((t', P', A)\) and \((c''_1, c''_2)\) satisfies constraint (1) given \((t'', P'', A)\), it follows that also \((c_1, c_2)\) satisfies constraint (1) given \((t, P, A)\), implying that \((c_1, c_2)\) is feasible. Hence,

\[
\begin{align*}
    U(t, P, A) &\geq (1 - \pi)u(c_1, 1) + \pi u(c_2, t) \\
    &> (1 - \pi)[\alpha u(c'_1, 1) + (1 - \alpha)u(c''_1, 1)] + \pi[\alpha u(c'_2, t') + (1 - \alpha)u(c''_2, t'')]
\end{align*}
\]

\[
= \alpha U(t', P', A) + (1 - \alpha)U(t'', P'', A) = U(\bar{t}, \bar{P}, A)
\]

where the first inequality follows since \((c_1, c_2)\) is feasible, and the second inequality follows since \(u\) is strictly increasing. This means that \(P(t, A; \bar{t}, \bar{P})\) is a strictly concave function of \(t\). We will henceforth refer to this property as diminishing willingness to pay for treatment.

### 4 Main results

Due to the diminishing willingness to pay for treatment, an individual being faced with the possibility of purchasing treatment at cost \(P = tC\), constrained by \(t \leq 1\), will have a unique level of treatment \(t(A)\) maximizing \(U(t, tC, A)\). Furthermore, due to the single-crossing property, \(t(A)\) will (weakly) increase with \(A\). In fact, whenever \(0 < t(A) < 1\), \(t(A)\) is determined by

\[
\frac{\partial P(t(A), A; t(A), t(A)C)}{\partial t} = C. \tag{5}
\]

9
I.e., marginal willingness to pay for treatment equals marginal cost of treatment (note that marginal cost equals total cost). It follows that \( t(A) \) is a strictly increasing function of \( A \) when \( 0 < t(A) < 1 \).

We have that \( t(A) = 1 \) for all \( A \geq A^* \), where \( A^* \) satisfies that the indifference curve through \((1, C)\) has slope \( C \), so that unconstrained maximization of \( U(t; tC, A^*) \) leads to \( t = 1 \). By the single-crossing property, \( A^* \) is unique. Hence, we can define \( A^* \) by

\[
\frac{\partial P(1, A^*; 1, C)}{\partial t} = C. \tag{6}
\]

Since \( \frac{\partial P(t; A; \bar{t}, \bar{P})}{\partial t} > A \) for all values of \( t, \bar{t}, \) and \( \bar{P} \), we have that \( A^* < C \). Moreover, it follows from equation (3) and constraint (1) that \( c_1 = c_2 = A - \pi C \) when \( t = 1 \) and \( P = C \), implying that \( t = 1 \) is not feasible when \( A < \pi C \). Finally, since \( u_c(c, h) \to \infty \) as \( c \downarrow 0 \) whenever \( h > 0 \), it follows that \( \frac{\partial P(1, A; 1, C)}{\partial t} \to 0 \) as \( A \downarrow \pi C \). This means that \( A^* > \pi C \). Note that the individual may choose a level of treatment that enables her to fully recover \( (i.e., h_2 = h_1) \) even if \( A < C \), provided that \( A \) is greater than, or equal to, the insurance premium.

The individual’s optimal level of treatment is illustrated in Figure 2 for two different values of ability: \( A_1 < A^* \) and \( A_6 = A^* \), where \( l \) and \( h \) denotes low and high ability, respectively. As illustrated, a high-ability individual’s indifference curve in \((t, P)\)-space is tangent to the marginal-cost line for \( t = 1 \), while that of a low-ability individual is tangent to the marginal-cost line for some \( t \in (0, 1) \).

These observations partially prove the proposition below. The proposition states that the individual’s utility is constant across states if she chooses full treatment: \( u(c_1, h_1) = u(c_2, h_2) \). Her utility if ill is lower than that if well if she chooses less than full treatment \( (i.e., t < 1) \), even though she receives an insurance indemnity: \( u(c_1, h_1) > u(c_2, h_2) \). Moreover, with full treatment, she will not receive any cash payment in addition to what is required to pay for treatment, while in the case of partial treatment, her compensation will exceed the amount spent on medical treatment.

**Proposition 1** There exists a level of inherent ability if healthy, \( A^* \), where \( \pi C < A^* < C \), such that:

1. If the individual’s ability when healthy, \( A \), is high, in particular, if \( A \geq A^* \), then her optimal level of treatment is equal to one and does not vary with \( A \): \( t(A) = 1 \). Moreover, her level of consumption is identical in the two states: \( c_1(1, C, A) = c_2(1, C, A) = A - \pi C \), as is her utility: \( u(c_1, h_1) = u(c_2, h_2) = u(A - \pi C, 1) \). Her insurance coverage is in the form of medical treatment only.

2. If, however, the individual’s ability when healthy, \( A \), is low, in particular, if \( 0 < A < A^* \), then her optimal level of treatment is positive but less than one, \( 0 < t(A) < 1 \), and increasing in \( A \): \( \frac{\partial t(A)}{\partial A} > 0 \). Moreover, her level of consumption if ill is lower than if healthy: \( c_2(t(A), t(A)C, A) < c_1(1, C, A) \).
c_1(t(A)), t(A)C, A, and her utility if ill is lower than if healthy: u(c_2, h_2) < u(c_1, h_1). Her insurance coverage is partly in the form of medical treatment and partly in the form of cash.

Proof. Part (1). Given the observations prior to the Proposition, it remains to show that the individual’s utility is constant across states, and that she has insurance coverage in the form of medical treatment only. Constant utility across states follows since c_1 = c_2 = A - \pi C and h_1 = h_2 = 1, implying that u(c_1, h_1) = u(c_2, h_2) = u(A - \pi C, 1). Since cash payment equals c_2 - tA + A - c_1 (cf. footnote 6), it follows that cash payment is zero.

Part (2). By the definition of \( A^* \), \( 0 \leq t(A) < 1 \) whenever \( 0 < A < A^* \). Moreover, since \( u_h(c, h) \to \infty \) as \( h \downarrow 0 \) whenever \( c > 0 \), and \( u_c(c, h) \to \infty \) as \( c \downarrow 0 \) and \( h \downarrow 0 \), it follows from \( A > 0 \) and equation (4) that \( \partial P(t, A; tC)/\partial t > C \) if \( t \) is sufficiently small; hence, \( t(A) > 0 \). Now, the single-crossing property implies that \( dt(A)/dA > 0 \). From equation (3) and the properties of \( u \), it follows that \( c_1 > c_2 \), since \( h_1 = 1 \), and \( h_2 = t(A) < 1 \). This in turn implies that \( u(c_1, h_1) > u(c_2, h_2) \). To show that cash payment is positive, i.e., that \( c_2 - tA + A - c_1 > 0 \), we start out with the observation that \( t(A) \) is determined by \( \partial P(t(A), A; t(A), t(A)C)/\partial t = C \) whenever \( 0 < t(A) < 1 \). Thus, marginal willingness to pay for treatment equals cost of treatment: \( u_h(c_2, t)/u_c(c_2, t) + A = C \). In the hypothetical case where treatment were available also if healthy, or inversely, where health could be sold at price \( C - A \), the access to actuarially fair insurance would imply the same level of health in both states. Since this is not the case, it is a binding constraint that health if healthy cannot be sold at price \( C - A \), implying that marginal willingness to pay for health if healthy is less than \( C - A \):

\[
u_h(c_1, 1)/u_c(c_1, 1) < C - A = u_h(c_2, t)/u_c(c_2, t)\]

Hence, effectively, the relative price of health in terms of consumption is lower if healthy than if ill. Combining this finding with constraint (1) and condition (2), and recalling that \( u_{cc} < 0 \) and \( u_{ch} > 0 \), we have that \( c_1 < A - \pi tC \), and \( c_2 > tA - \pi tC \). This in turn means that \( c_1 - A < c_2 - tA \), or \( c_2 - tA + A - c_1 > 0 \). \( \square \)

Proposition 1 can be illustrated by considering the following Bernoulli utility function:

\[
u(c, h) = c^rh^s, \text{ with } r > 0, \ s > 0 \text{ and } r + s < 1,
\]

which satisfies all assumptions listed in Section 2. It is now straightforward, by using equations (1), (4), (5), and (6), to show that

\[
A^* = \frac{r + s}{r + s} C
\]

and, if \( A < A^* \), then the follow expression obtains for the cash compensation:

\[
c_2 - t(A)A + A - c_1 = \frac{r + s}{(1 - \pi)s} (A^* - A) t(A).
\]

This means that the critical level \( A^* \) increases with both a greater probability of falling ill, \( \pi \), and a greater cost of treatment, \( C \). Furthermore, for given values
of \( \pi \) and \( C \) (and hence, for a given value of \( A^* \)), the individual’s level of ability, \( A \), has an ambiguous effect on the size of the cash compensation. On the one hand, provided that \( A < A^* \), a smaller \( A \) leads to a greater relative importance of the cash compensation through the term \( A^* - A \). On the other hand, an individual with a smaller ability, \( A \), has a smaller initial endowment and will choose an insurance contract with a lower level of total compensation. This moderating effect on the size of the cash compensation is reflected by the term \( t(A) \), which decreases with a smaller \( A \) when \( A < A^* \).

5 Ex-post moral hazard

We have, in Proposition 1, shown that an ill individual with ability \( A \) lower than the critical level \( A^* \) receives partial treatment, \( t(A) < 1 \), and, in addition, a positive cash compensation:

\[
C_2(A) - t(A)A + A - C_1(A)
\]

where we from now on write \( C_1(A) = C_1(t(A), t(A)C, A) \) and \( C_2(A) = C_2(t(A), t(A)C, A) \). If, contrary to what we have assumed in the formal analysis, the two states (healthy/ill) are not observable, the availability of such a disability insurance will tempt the individual to claim that she has fallen ill, although she is in fact in good health. In this section, we show that our analysis goes through even if we allow for such ex-post moral hazard,8 provided that:

- the cash compensation is paid only in combination with treatment, and
- the disutility of receiving treatment while healthy is sufficiently great.

Consequently, in order to prevent the individual from falsely claiming to be ill, she should suffer a loss in expected utility from undergoing redundant medical treatment. Moreover, the disutility should at least balance the gains in expected utility from masquerading as ill.

Denote by \( v(c, t) \) the (direct) disutility of receiving treatment \( t \) while healthy and consuming \( c \). Assume that \( v \) satisfies, \( \forall (c, t) \in \mathbb{R}_{+}^2, v(c, t) > 0 \) and \( v_t \geq 0 \).9

To ensure that a healthy individual with ability \( A \) will not falsely claim to be ill, we must consider the possibility she may not only misrepresent her health state but also her ability, in order to receive the higher cash compensation designed to be paid to an individual with a different ability \( A' \). Hence, to prevent

---

8 Ex-post moral hazard refers to the effect of insurance on insured individuals’ incentives to reveal their true health state (i.e., the insured individual knows the state of the world, while the insurer does not, or verification of health state is too costly for the insurer). The analysis of ex-post moral hazard was pioneered by Spence and Zeckhauser (1971). For a textbook exposition, see Zweifel and Breyer (1997, Chapter 6.5).

9 The disutility \( v(c, t) \) is the difference between the utility derived from consumption \( c \) when healthy and not receiving unnecessary treatment and the utility derived from consumption \( c \) when healthy and receiving unnecessary treatment \( t \). Hence, this disutility is a function of \( c \) and \( t \).
that \textit{ex-post} moral hazard is a problem, the following inequality must hold, for any true ability \(A\) and any claimed ability \(A'\):

\[
(1 - \pi)u(c_1(A), 1) + \pi u(c_2(A), t(A)) \geq (1 - \pi)(u(c'_2, 1) - v(c'_2, t(A'))) + \pi u(c_2(A') + t(A')(A - A'), t(A')).
\]

In condition (8), \(c'_2 = c_2(A') + (A - t(A'))A'\) represents the consumption that a healthy individual with ability \(A\) will receive having purchased the optimal insurance contract of an individual with ability \(A'\) and masquerading as ill, while \(c_2(A') + t(A')(A - A')\) is the consumption that an ill individual with ability \(A\) will receive having purchased the optimal insurance contract of ability \(A'\) and truly claiming to be ill. An individual with ability \(A\) higher (lower) than \(A'\) generates higher (lower) earnings and can, therefore, sustain a higher (lower) level of consumption than can an individual with ability \(A'\).

Notice that, in condition (8), we allow the claimed ability \(A'\) to take any value, including the real ability \(A\). Hence, we do not require that an individual also lies about her ability when she misrepresents her health state, but we allow for this possibility. Of course, a misinterpretation of ability must, if it occurs, take place before the individual knows whether she has fallen ill or not; this is reflected by the last term on the r.h.s. of condition (8).

To find a sufficient condition for (8) to hold for any true ability \(A\) and any claimed ability \(A'\), the following observation is useful: We know that \((c_1(A), c_2(A), t(A))\) maximizes expected utility over all triples \((c_1, c_2, t)\) satisfying the constraint in equation (1); in particular,

\[
(1 - \pi)u(c_1(A), 1) + \pi u(c_2(A), t(A)) \geq (1 - \pi)u(c'_1, 1) + \pi u(c_2(A') + t(A')(A - A'), t(A')),
\]

where \(c'_1 = c_1(A') + (A - A')\) is the consumption that a healthy individual with ability \(A\) will receive having purchased the optimal insurance contract of ability \(A'\) and not masquerading as ill. This means that an individual with ability \(A\) will not lie about her ability unless she intends also to misrepresent her health state.

Notice that the increase in consumption that an individual with ability \(A\), having purchased the optimal insurance contract of ability \(A'\), attains by masquerading as ill, \(c'_2 - c'_1\), is equal to the cash compensation designed to be paid to an individual with ability \(A'\):

\[
c'_2 - c'_1 = c_2(A') - t(A')A' + A' - c_1(A').
\]

Moreover, by equation (9), it is a sufficient condition for equation (8) to be satisfied for any true ability \(A\) and any claimed ability \(A'\), that

\[
u(c'_2, 1) - v(c'_1, 1) \leq v(c'_2, t(A'))
\]

holds for any true ability \(A\) and any claimed ability \(A'\).

We have shown the following result.
Proposition 2 The ex-post moral hazard of an healthy individual masquerading as ill in order to obtain cash compensation does not constitute an incentive problem if, for any true ability \( A \) and any claimed ability \( A' \), the additional utility that \( A \) obtains from the cash compensation, \( c_2(A') - t(A')A' + A' - c_1(A') \), to be paid to \( A' \) in case of illness, does not exceed the disutility that \( A \) suffers from undergoing, when healthy, redundant treatment at the level, \( t(A') \), that \( A' \) is entitled to.

It is clear from this result that a healthy individual will falsely claim to be ill only after having purchased the optimal insurance contract of some ability \( A' \) smaller than \( A^* \). The reason is that, according to Proposition 1, for any ability \( A' \geq A^* \), the optimal insurance contract includes no cash compensation, implying that the disutility from undergoing redundant treatment will dominate. On the other hand, if \( A' < A^* \), there is a positive cash compensation. What can be said about how this cash compensation varies with \( A' \) for \( 0 < A' < A^* \)?

To investigate this question, it might be instructive to look at the special Cobb-Douglas case considered at the end of Section 4. It follows from the expression for the cash compensation, equation (7), that the level of claimed ability, \( A' \), has an ambiguous effect on the size of the cash compensation. While, with \( A' < A^* \), a smaller \( A' \) leads to a greater relative importance of the cash compensation through the term \( A^* - A' \), there is a moderating effect through the term \( t(A') \), which reflects that a smaller \( A' \) leads to an insurance contract with a lower level of total compensation. Hence, provided that \( v(c, t) \) does not decrease significantly with a smaller \( t \), and thus with a smaller \( A' \), we may have that (10) holds for any true ability \( A \) and any claimed ability \( A' \).

When health state is not verifiable, i.e., when ex-post moral hazard is a problem, the individual will have an incentive to masquerade as ill in order to acquire a cash compensation. However, when cash compensation is made conditional on medical treatment, we have shown that the individual’s incentive to masquerade is reduced since she will suffer a disutility from receiving redundant treatment. The ex-post moral hazard problem associated with cash compensation is hence solved in our model through the integration of treatment for illness and payments for disability. The lack of such integration can help explain why private markets for disability insurance are of little empirical significance. Naturally, one may argue that even if treatment and cash compensation were not integrated, then information about whether an individual is ill could be obtained if the insurer offering disability insurance could require information from the insurer offering medical insurance. In this case, information on the (contractually) adequate level of treatment as well as the level of treatment actually undertaken is required. It follows that the informational costs would be higher relative to a situation in which the two types of insurance are integrated.

In reality, the true health state is simple to verify for some medical conditions, while difficult for others; moreover, the disutility from receiving unnecessary treatment is significant for some types of conditions, while insignificant for others. Hence, the severity of the ex-post moral hazard problem discussed in this section varies according to the medical conditions.
6 Discussion

Our focus of attention has been on how an individual’s inherent ability at full functionality (i.e., when healthy) influences her \textit{ex ante} choice of insurance contract and her optimal level of coverage. Insurance allows the individual to allocate income between the two states of the world \textit{prior} to knowing which state has occurred. Moreover, it enables her to achieve her optimal distribution of income on consumption and health when ill. Since the individual is assumed to have perfect foresight, her optimal allocation \textit{ex ante} will be optimal also \textit{ex post}.

The novelty of this paper is the integration of what is usually thought to be different types of insurance, namely insurance against the risk of losing income due to (permanently) reduced health (ability) and insurance against the risk of incurring medical expenditures. We argue that a health insurance should offer a hedge against both potential loss in income due to reduced health and potential expenditures on medical treatment. Contrary to what is assumed in most of the health insurance literature, we allow the individual to choose whether or not to restore health if ill. We show that the individual’s marginal willingness to pay for treatment is increasing in ability, and we derive a critical level of ability at which an individual prefers to fully restore health if ill. If the individual’s inherent level of ability is sufficiently low, then she chooses to restore health only partly, thus suffering a loss in ability. In order to obtain the preferred level of consumption if ill, she will hold a contract that in addition entitles her to a cash transfer in the event of illness. She will, however, suffer a loss in utility if ill. If, on the other hand, the individual has a sufficiently high level of inherent ability, then she prefers a contract that provides for complete medical treatment and thus full restoration of health.

It is of no importance, in a world of symmetric information, whether the coverage for medical costs is paid in cash intended to cover medical bills, or directly in the form of medical treatment. The individual’s \textit{ex-ante} decision concerning what level of treatment to choose is unaffected by the way she is compensated; the fundamental decision concerns to what extent health is to be restored.\textsuperscript{10} However, as discussed in Section 5, if health state is not easily verifiable, then, due to the \textit{ex-post} moral hazard problem, it becomes essential whether medical expenditures are compensated in cash or in kind. When information about health state is asymmetric, integration of a cash compensation of income loss and an in-kind compensation of medical expenditures reduces the individual’s incentive to falsely claim to be ill.

Our findings are driven by the fact that the potential loss in income is larger, the higher the ability at full functionality. This implies that the prices of the two types of contracts differ depending on the individual’s ability. The higher

\textsuperscript{10} Arrow (1963) mentions three different ways in which costs of medical care can be covered in an insurance contract: payment directly in medical services, a fixed cash payment, and a cash payment that covers the actual costs involved in providing the necessary medical treatment. In a perfect market, individuals wishing to receive medical treatment would be indifferent between a payment directly in the form of medical treatment and its cash equivalent.
the potential income loss due to reduced ability, the cheaper is the contract offering indemnity in kind (i.e., treatment), compared to a contract offering cash compensation of income loss. Thus, the cost-benefit ratio on medical treatment is lower the higher the level of ability at full functionality.

The preceding analysis is based on a highly stylized model. We largely disregard any informational constraints causing the familiar problems of adverse selection. Furthermore, the individual is assumed to be fully informed ex ante about health consequences of illness as well as about treatment options. The insurers need not, however, know the individual’s ability at full functionality, since, even without such knowledge, first-best, zero-profit insurance contracts lead in an undistorted way to self-selection. Transaction costs associated with gathering of information about relevant treatment options and treatment costs for all types of diseases are ignored. Moreover, we make a somewhat strong assumption regarding the treatment technology: the individual recovers instantly and proportionally to the level of treatment received, and treatment is effective with respect to health. However, in spite of these limitations, our model still provides interesting findings that may be subject to further studies.

References


---

11 In Asheim, *et al.* (2003), we expand our analysis to a situation of asymmetric information about the probability of illness.


Appendix: Calculations for the Cobb-Douglas case at the end of Section 4, not intended for publication

The following Cobb-Douglas function is a Bernoulli utility function that satisfies all assumptions listed in Section 2:

\[ u(c, h) = c^r h^s, \] with \( r > 0, \ s > 0 \) and \( r + s < 1. \)

With this function, it is possible explicitly to calculate \( A^*. \) We have that

\[
\frac{\partial P(1, A, 1, C)}{\partial t} = \frac{u_h(c, h)}{u_c(c, h)} + A
\]

\[
= \frac{u_h(c_2, 1)}{u_c(c_2, 1)} + A = \frac{s}{r} (A - \pi C) + A,
\]

where the first equality follows from (4), the second equality follows since \( c_2 = A - \pi C \) when \( t = 1 \) and \( P = C, \) and the third equality follows since

\[
\frac{u_h(c, h)}{u_c(c, h)} = \frac{s}{r} \cdot \frac{c}{h}
\]

(A1)

when \( u \) is given by the Cobb-Douglas function above. Since \( A^* \) is defined by \( \frac{\partial P(1, A^*; 1, C)}{\partial t} = C, \) we can find \( A^* \) by solving

\[
\frac{s}{r} (A^* - \pi C) + A^* = C,
\]

which implies that

\[
A^* = \frac{r + \pi s}{r + s} C. \hspace{1cm} \text{(A2)}
\]

Since, for \( A < A^* \), \( t \) is determined by \( \frac{\partial P(t, A; t, tC)}{\partial t} = C, \) we get, by invoking equations (4) and (A1), that

\[
\frac{s}{r} \cdot \frac{c_2}{t} + A = C. \hspace{1cm} \text{(A3)}
\]

Moreover, by letting the cash compensation, \( c_2 - tA + A - c_1, \) be denoted by \( x, \) it follows from equation (1) that

\[
x - c_2 + tA = \pi [tC + x]. \hspace{1cm} \text{(A4)}
\]

We now have that

\[
(1 - \pi)x = \pi tC + c_2 - tA
\]

\[
= \pi tC + \frac{s}{r}(C - A) - tA
\]

\[
= \frac{1}{2} [(r + \pi s)C - (r + s)A] t
\]

\[
= \frac{r + s}{r} (A^* - A) t,
\]

where the first equality follows from (A4), the second equality follows from (A3), and the fourth equality follows from (A2). We have thereby shown that the cash compensation can be expressed as follows:

\[
c_2 - t(A)A + A - c_1 = \frac{r + s}{(1 - \pi)s} \cdot (A^* - A) \cdot t(A).
\]
Figure 1. The single-crossing property.

Figure 2. The optimal level of treatment.