The Purpose and Limits of Social Health Insurance

by Peter Zweifel (University of Zurich)*
pzweifel@soi.unizh.ch

1. Introduction and motivation

This contribution seeks to answer two related questions. First, what is the purpose of social health insurance? Or put in slightly different terms, what are the reasons for social (or public) health insurance to exist, even to dominate private health insurance in most developed countries? And second, what are the limits of social health insurance? Can one say that there is “too much” social health insurance in the following two senses: Should the balance be shifted towards the private alternative? And is the degree of coverage excessive?

Accordingly, the plan of this paper is as follows. Section 2 is devoted to the efficiency, public choice, and equity reasons that may explain the existence (but not necessarily the prominence) of social health insurance. After having examined the demand for social insurance, its supply is analyzed in section 3. A simple model expounding the trade-off between innovation and risk selection effort of a private and a social health insurer will be developed. The innovation dimension importantly comprises efforts at containing moral hazard; therefore, the model may provide a preliminary answer also to the issues of ’excessive coverage provided by social health insurance’. This is followed by the

economic theory of the optimal health insurance contract in section 4, which provides the benchmark for a discussion of actual outcomes of both a private and a regulated social health insurance market (section 5). It would be preferable to be able to describe and evaluate ‘equilibria’; however, there is not only a price and a single quality dimension to be determined and evaluated. Therefore, the discussion will be somewhat descriptive at this points. Section 6 provides a survey of the future challenges confronting health insurance and an assessment of social and private insurance to cope with them. The final section offers a few conclusions and suggestions for future research.

2. Why social health insurance?
Most developed countries have some kind of collective financing for health services, either through tax (e.g. the National Health Service on the United Kingdom) or through their contributions to social health insurance (henceforth: SHI). This type of insurance is usually characterized by mandatory membership, at least for the vast majority of the population, open enrolment, and community rating, i.e. a prohibition to charge premiums related to individual risk. From a normative point of view, the institution of SHI can be defended on both efficiency and equity grounds, whereas positive economics can explain its existence in democracies on the basis of public choice models.

2.1 Efficiency reasons: characteristics of private health insurance markets
SHI may be efficiency-enhancing if it mitigates or even eliminates possible market failures.

(a) Asymmetric information. Ever since the seminal contribution by Rothschild and Stiglitz (1976), private competitive insurance markets are suspected to exhibit adverse selection due to asymmetric information. If the insured have more precise information on their individual risk distribution than does the insurer, the only sustainable Rothschild-Stiglitz equilibrium is a separating one in which only the highest risk types are offered complete coverage at actuarially fair premiums. Lower risks
obtain more favorable terms but are rationed in terms of coverage. They would prefer to have more coverage, but this would make their contract attractive to unfavorable risks. Compared to such a separating equilibrium, SHI which forces all individuals into a pooling contract with partial coverage can achieve a Pareto improvement. High risks are made better off because they pay lower premiums for the mandated part of their coverage, whereas low risks benefit from improved total (social plus private) coverage (Newhouse 1996). However, it is unclear to what extent asymmetric information on health risks is really a problem since medical exams are used to determine the risk of an insured.

(b) Altruism and free riding. Altruistic rich members of society may be willing to subsidize the provision of health care to the poor if they are more interested in the health than in the subjective well-being of the poor (Pauly 1970). Private charity is not suitable to reach an efficient allocation since donations to the poor, whether in cash or in kind, have a public-good characteristic, increasing the utility not only of the donor but also of other altruistic members of society. Either a tax-financed National Health Service or SHI with compulsory membership and contributions according to ability to pay solve this free-rider problem of potential donors. Note that this argument implies that social health insurance is an institution serving the interests of the rich rather than the poor, as argued by Feldstein (2005).

(c) Optimal taxation when health and income are correlated. A related justification of SHI is derived from the theory of optimal taxation (Cremer and Pestieau, 1996). If abilities cannot be observed by tax authorities, the extent to which income taxation can be used for redistribution from the high-skilled to the low-skilled is limited because the high-skilled can always “pretend” to be low-skilled by reducing their labor supply. However, if there is a negative correlation between ability and the risk of illness, a mandatory SHI with community rating implicitly redistributes between the ability groups in the desired fashion and thus improves social welfare. It must be emphasized, however, that this justification departs from Paretian welfare economics by postulating a specific redistributive goal.
2.2 Equity reasons

A further justification, also known as the ‘principle of solidarity’, relates to the achievement of equality of opportunity. People differ in their health risk already at birth, and some indicators of risk are readily observable. Moreover, with the rapid progress of genetic diagnostics and the spread of tests during pregnancy, the ability to measure individual health risks of newborns will become more and more pronounced. In private health insurance (PHI), these differences in risk immediately translate into differences in premiums. This causes persons who are endowed by nature with a lower stock of ‘health capital’ and are thus already disadvantaged to have to pay a higher price for the same coverage on top of their handicap. Behind the veil of ignorance, risk-averse individuals would prefer to see at least the financial cost of illness spread equally between member of society.

Table 1: Alternatives for achieving solidarity

<table>
<thead>
<tr>
<th>Alternatives</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current transfer</td>
<td>Permits full competition in PHI (or SHI) both in premiums and products and full information on risk</td>
<td>Means testing; definition of benchmark contract</td>
<td>Pauly et al. 1992</td>
</tr>
<tr>
<td>Lump sum transfer for lifetime</td>
<td></td>
<td>Means testing; longevity risk shifted to beneficiaries</td>
<td>Knappe 2005</td>
</tr>
<tr>
<td>Regulation: Community rating</td>
<td>Relieves public budget</td>
<td>Induces cream skimming and necessitates a RAS as secondary regulation</td>
<td>Van de Ven et al. 2000</td>
</tr>
</tbody>
</table>

Note: RAS = risk adjustment scheme

There are in principle two main ways to achieve solidarity in health insurance (see Table 1). First, PHI premiums can be subsidized for those who would have to pay excessive contributions in two variants. The transfer could be on a current basis. Or it could be a lump sum, equal to the estimated present value of future excess premiums over the whole expected life span of beneficiaries. Both have the important advantage of permitting full competition in PHI (or SHI), including insurers acquiring information
about true risk. Besides means testing and the need to define a benchmark contract to determine the amount of the subsidy, the second variant has the disadvantage of shifting the risk of longevity to beneficiaries. The second main alternative is a monopolistic SHI scheme with open enrolment and community rating that prevents differences in health risk from being translated into differences in contributions but induces cream skimming and risk adjustment schemes (RAS, see below) as a secondary neutralizing regulation.

2.3 Public choice reasons

In PHI, redistribution occurs purely by chance, from consumers who do not suffer a loss during the life of the contract to those who do. By way of contrast, social insurance mixes in elements of systematic redistribution. The fact that contributions are not (or not fully) graded according to risk (OECD 2004) alone serves to systematically redistribute wealth from high risks to low risks. In SHI, this redistribution not only affects wealth through its financing side but the benefit side as well, viz. medical services and health. This makes social health insurance an ideal means for a politician who seeks office (or re-election) by catering to the interests of groups who are sufficiently organized to have an effect on the outcome of an election (Gouveia 1997; Hindriks and De Donder 2003; Tullock 2003). The redistributive effects of SHI can be described as follows.

(a) Redistribution of wealth. Using SHI as a vehicle for systematic redistribution has the important advantage that net payers have considerable difficulty determining the systematic component of redistribution. For example, when the contribution to SHI amounts to a payroll tax (as e.g. in Germany), high wage earners pay more for their health insurance. However, they are uncertain about the systematic redistribution component of their contribution because the expected value of their benefits may also be higher than average (Henke, 2005). This may have two reasons: Preventive effort may be affected negatively by a higher wage, resulting in higher health costs, and demand for medical services may increase because short-term disability benefits usually increase with wages, creating a spill-over moral hazard effect (Zweifel and Manning 2000). Therefore, their higher contribution appears “justified”, masking a tax component which if collected as a tax would likely be opposed. However, as shown by
Feldstein (2005) for Medicare in the United States, SHI in fact benefits the middle and upper classes, who live long enough to use the health care services covered by Medicare.

(b) Redistribution of medical care. There are two effects here. First, there is an income effect because some individuals who would have demanded less or no medical care without insurance coverage now demand a positive amount of it (Nyman 2003). Indeed there is (macro) evidence suggesting that medical care is a normal good (Gerdtham et al. 1992; Miller and Frech 2004; Zweifel et al. 2005). Insurance coverage then amounts to an in-kind redistribution from the rich to the poor if the supply of medical services is not infinitely elastic and if the price elasticity of demand for medical care is not lower for the rich than the poor (which is doubtful, see Newhouse et al. 1993, ch. 11). However, there is also a price effect because health insurance boosts the ‘true’ (out-of-pocket) willingness-to-pay (WTP) for medical care depending on the rate of coinsurance (Zweifel and Breyer 1997, ch. 10). For example, if ‘true’ WTP is 100 and the rate of coinsurance is 25 percent, observed WTP is 400.

(c) Redistribution of health. Altruism with regard to health is probably more marked than with regard to income, although comparative evidence seems to be lacking (the methodology for measuring distributive preferences for health is still in its infancy, see Olsen 2000). Therefore, politicians can claim to have a mission when seeking to guarantee “health for all” (the famous slogan of the World Health Organization). Equal access to health insurance then may be seen as an important factor for securing equal access to medical care, and to the extent that medical care is effective at the margin [for which there is some evidence, see e.g. again Miller and Frech (2004) and Lichtenberg (2004) for securing equal health status Culyer and Wagstaff (1993)].

If SHI indeed contributes to winning votes and to increasing the chance of (re-)election of a democratic government, one would expect public expenditure for it to increase around election time. One piece of available evidence relates to two types of public expenditure by the Dutch government, expressed as shares in GDP, between ca. 1956 and 1993, viz. health (such as subsidies to hospitals) and tax contributions to
social insurance in general. Van Dalen and Swank (1996), cited in Zweifel (2000a), find that while public expenditure on health does not vary around election time, transfers in favor of social insurance are systematically higher during the years prior, concurrent with, and after an election. The estimated effect is 13 percent, e.g. an increase from 15 to 17 percent of GDP. In addition, the share of the pensioners in the population is significantly related to both types of public expenditure. Nowadays pensioners are not poor, but they do go to the polls. The evidence thus is compatible with governments proposing social health insurance schemes to benefit pivotal voter groups.

**Conclusion 1:** While the efficiency reasons for social (health) insurance have received much attention in the literature, they are found not fully convincing. As to the equity reasons, targeted premium subsidies emerge as an alternative to community rating. On the whole, the available empirical evidence suggests that public choice reasons (i.e. winning votes) may well be the crucial reason for the existence and even more for the growth of social (health) insurance.

### 3. The supply of social health insurance

According to the argument advanced toward the end of the previous paragraph, governments (and public administration) could be seen as the suppliers of social health insurance. This is true in systems of the National Health Service type, where in fact the government itself provides the insurance function while also acting as the organizer of medical care. This type will be called „public health insurer“ in the following. However, in a majority of industrialized countries, health insurers are not incorporated in the government’s budget; they will be called „(competitive) social health insurers“.

#### 3.1 Supply by a competitive social health insurer

Supply of health insurance can be characterized by several dimensions (Zweifel, 2005), such as comprehensiveness of the benefit package, the amount of loading included in the premium, the degree of vertical integration of healthcare providers, the amount of
effort devoted to risk selection, and the degree of seller concentration. For simplicity, only two decision variables will be analyzed below which however are of particular interest for policy. On the one hand, the insurer can devote effort to innovation \((i)\), resulting in new benefits covered but also – and even more importantly – in a better control of ex-post moral hazard (i.e. moral hazard given that illness has occurred). Developing managed care alternatives or contracts with bonus options for no claims are examples of such costly efforts.

On the other hand, the insurer can invest in risk selection effort \((s)\), trying to ‘skim the cream’, an activity that has no social value (assuming that the threat of being found to be a high risk does not induce preventive effort on the part of consumers). Let \(\mu(i,s)\) denote the share of high risks in the insurer’s population at risk; this share not only depends on \(s\) but also on \(i\) because innovation typically appeals to younger consumers (which are lower risks at least on average). The premium (and hence the present value of their flow \(P\)) is regulated to be uniform and constant for simplicity. However, high and low risks differ in their probability \((\pi^h, \pi^l)\) of claiming benefits during the planning period of the insurer. As noted above, innovation effort also has the effect of lowering the present value of losses \(L\), which are assumed not to depend on the type of insured, again for simplicity. Finally, both innovation and selection effort have a price of one.

Although social health insurers may not per se pursue the maximization of expected discounted profit \(E\Pi\), they still want to ensure their economic survival in the face of competition. To this end, accumulating reserves is of some importance. However, this ultimately implies behavior no different from maximizing expected discounted profit (of course under regulatory constraint, which also may result in a planning horizon that differs from a for-profit insurer). Therefore, the objective function of such a social health insurer may written as (see Zweifel and Eisen, 2002, ch. 5.5.2 for a similar model),

\[
\max_{i,s} E\Pi = \mu(i,s) (P - \pi^h L(i)) + (1 - \mu(i,s)) (P - \pi^l L(i)) - i - s,
\]

with \(\frac{\partial \mu}{\partial i} < 0\), \(\frac{\partial \mu}{\partial s} < 0\), \(\frac{\partial L}{\partial i} < 0\).
Neglecting boundary solutions, the first-order conditions for an optimum are given by

\[
\frac{\partial E\Pi}{\partial i} = \frac{\partial \mu}{\partial i} (P - \pi^h L) + \mu \left( -\pi^h \frac{\partial L}{\partial i} \right) - \frac{\partial \mu}{\partial i} (P - \pi^i L) + (1 - \mu) \left( -\pi^i \frac{\partial L}{\partial i} \right) - 1 = 0
\]

(2)

\[
\frac{\partial E\Pi}{\partial s} = \frac{\partial \mu}{\partial s} (P - \pi^h L) - \frac{\partial \mu}{\partial s} (P - \pi^i L) - 1 = 0.
\]

(3)

Focusing on equation (2) first, and multiplying it by \((i/\mu)\), one obtains

\[
\frac{\partial \mu}{\partial i} \left( P - \pi^h L \right) - \frac{\partial \mu}{\partial i} \left( P - \pi^i L \right) - \frac{\partial \mu}{\partial s} \left( P - \pi^i L \right) / \mu = \frac{i^*}{\mu^*}
\]

(4)

The first term on the left-hand side (LHS) is an elasticity \([e(\mu, i)]\) indicating how much a one percent increase in innovative effort serves to decrease (in percent) the share of high risks in the insurer’s population. It is treated as a constant in the following, although its value in general will depend on the levels of both \(i\) and \(s\). The term in brackets is also negative. With a common present value of premiums \(P\), the high risks cause a negative contribution to expected profit and the low risks a positive one. Together, these two terms define a first component of the marginal benefits of innovative effort.

The second component again contains an elasticity \([e(L, i)]\), which indicates the effectiveness of innovation in terms of lowering the amount of future losses \(L\). The term in brackets multiplied by \(L\) is nothing but the overall expected value of discounted future losses. This makes sense: Efforts at controlling ex-post moral hazard have a particularly high marginal benefit if the initial amount of expected losses is high; accordingly, the optimal value of innovation \(s^*\) is higher ceteris paribus (see the right-hand side (RHS) of equation (4)). However, the last factor \((1/\mu)\) shows that this benefit is dissipated across the high risks. The higher their share, the smaller this second component of benefits of innovation. Finally, the RHS of equation (4) is nothing but the marginal cost of innovation, again distributed across the high risks.

Turning to equation (3) now and multiplying it through by \((s/\mu)\), one obtains
The first term on the LHS is again an elasticity \[ e(\mu, s) \], indicating the effectiveness of selection effort. Not surprisingly, the term in brackets shows the negative overall contribution to expected discounted profits. Therefore, the greater the difference between the two types of risk in the face of the uniform premium, the higher the optimal amount of selection effort \( s^* \), ceteris paribus [see the RHS of eq. (5)]. However, its marginal cost can again be distributed over the number of high-risk insured \( \mu \).

Of course, a thorough discussion of the interplay of the two decision variables and their responses to exogenous changes would necessitate a full-blown comparative static analysis. This is relegated to the Appendix. The result with regard to risk adjustment can be stated as follows. Risk adjustment is designed to neutralize a competitive insurer’s incentive to ‘skim the cream’ by making an insurer who has an above-average share of favorable risks pay a contribution into the risk adjustment scheme (RAS) which can be used to ‘subsidize’ a competitor who has an above-average share (for an extended discussion, see Van de Ven and Ellis, 2000). In the present context, the net discounted premium \( P \) would be raised by the contribution to be paid to the RAS for a high risk, ideally causing the difference \( P - \pi^h L \) in equation (5) to go to zero, while the RAS payment received would increase \( P \) to make \( P - \pi^l L \) zero as well. In this case, equation (5) implies \( s^* = 0 \) as an optimum: the incentive for risk selection is neutralized.

However, the RAS has a side effect on innovative effort. For indeed, the first term on the LHS of equation (4) goes to zero as well. After all, one component of the marginal benefit of innovation was to attract the mobile, younger, usually favorable risks. With this term equal to zero, \( i^* \) has a lower value, ceteris paribus. Now the ceteris paribus clause does not really hold because selection effort also is lowered. The second benefit component on the LHS is strongly lowered (its derivative w.r.t. \( \mu \) being \(-\pi^l L/\mu^2\)), which is true of the RHS as well, its derivative being \(-i^*/\mu^2\). On balance, it appears that this indirect effect of the RAS on innovative effort is unlikely to dominate the first-order effect, which points toward a decrease of \( i^* \). Indeed, the experience of those
countries that have introduced competition between social health insurers combined with a RAS (such as Germany, the Netherlands, and Switzerland) has been somewhat disappointing. Health insurers have been rather slow launching new products; of course, they may have been hampered by regulation in addition to the RAS (Beck et al., 2003).

3.2 Supply by a public health insurer

Since the manager of a public insurance scheme is a public official, the full set of interactions between a politician, a bureaucrat, and a voter should be specified in principle [as e.g. in Alesina and Tabellini (2002); see also Boldrin and Rustichini (2000) and Hammond and Knott (1996); for a comparative description of regulation of social health insurers, see Maarse et al. (2004)]. Here, a much simpler alternative will be presented that has the advantage of facilitating comparisons with section 2.1 above.

From the outset, there are at least two institutional differences that need to be noted. First, a public insurance scheme typically is not allowed to build major reserves. Reserves are also unnecessary because economic survival of the scheme is assured by the government. This means that a public official pursuing his or her mission prefers to have a balanced budget. However, if there is a deviation $D$ from a balanced budget, the likelihood $\rho(D)$ that the envisaged utility level can be in fact attained decreases. The public official faces a certain probability of losing his or her position (the utility associated with the possible alternative employment is normalized to zero for simplicity). The official’s objective function can then be written

$$EU = \rho(D) U(D), \text{ with } \frac{\partial U}{\partial D} D < 0.$$  

If there is a deficit ($D < 0$), then the official has an increase in utility if $D$ increases towards zero ($\frac{\partial U}{\partial D} > 0$); if the scheme has a surplus ($D > 0$) however, a further increase in $D$ causes a decrease in utility [for a similar formulation in the case of a public hospital, see Zweifel and Breyer (1997), ch. 9.3.2]. The respective marginal utilities are normalized to $\frac{\partial U}{\partial D} = \pm 1$ below.
The second difference concerns the decision variables. The public scheme being a monopoly enrolling the entire population, it has no reason to exert any risk selection effort; moreover, the share of low risks $\mu$, being exogenous, does not respond to innovative effort $i$. The only decision variable remaining therefore is $i$, innovative effort.

In view of these considerations, and focusing on the case of a deficit ($D < 0$, $\frac{\partial U}{\partial D} = 1$), one can write the public health insurer’s objective function as

$$
(7) \quad \max_i EU = \min \rho(D)D = -\rho(D)\left\{\mu (P - \pi^h L(i)) + (1 - \mu)(P - \pi^l L(i))\right\} - i.
$$

Noting that $D$ depends on $L(i)$, the first-order condition for an interior solution reads

$$
(8) \quad \frac{\partial EU}{\partial i} = \frac{\partial \rho}{\partial D} \left\{ -\mu \pi^h \frac{\partial L}{\partial i} - (1 - \mu) \pi^l \frac{\partial L}{\partial i} \right\}D + \rho \left\{ -\mu \pi^h \frac{\partial L}{\partial i} - (1 - \mu) \pi^l \frac{\partial L}{\partial i} \right\} - 1 = 0.
$$

This can be multiplied by $(i/\rho)$ to become

$$
(9) \quad \left\{ \frac{\partial \rho}{\partial D} \frac{D}{\rho} \left( \frac{\partial L}{\partial i} \frac{L}{\rho} \right) \left[ \mu \pi^h L + (1 - \mu) \pi^l L \right] - \left( \frac{\partial L}{\partial i} \frac{L}{\rho} \right) \left[ \mu \pi^h L + (1 - \mu) \pi^l L \right] \right\} = \frac{i^*}{\rho}.
$$

On the LHS, there is first the elasticity $e(\rho, D)$ of the probability of regulatory sanction w.r.t. to a deviation from budget balance, reflecting the responsiveness of the government to budgetary problems with public health insurance. This elasticity is negative in the case of a deficit when budget balance is approached ($dD > 0$). The second elasticity shows the effectiveness of innovative effort in terms of the reduction of expenditure to be covered, i.e. of loss $L$, precisely as in section 2.1 above. The term in brackets is the expected value of the future losses to be paid by the scheme, possibly discounted by a different rate of discount than in the case of a competitive social insurer. The RHS of the equation shows that the marginal cost of innovation is “distributed” across the probability mass of the possible sanction $\rho$. 
3.3 Comparison between the two types of health insurance

Equations (4) and (9) may now be compared. Since on average, the $\mu^*$ of the competitive insurers must be equal to the fixed $\mu$ of the public scheme, the two values are set equal to simplify discussion. With regard to the first component of marginal benefit, the one in equation (4), involving only differences in contributions to expected profit, is typically smaller than the one of equation (9), involving the full expected value of losses, unless of course the government hardly reacts to any deviation from budget balance $\left[ e(\rho, D) = \frac{\partial \rho}{\partial D} D = 0 \right]$. On the other hand, the second benefit term of equation (4), being targeted on the high risks ($\mu$), is much larger than the corresponding term of equation (9). On the whole, the marginal benefit of innovative effort likely is larger for a competitive social health insurer than for its public counterpart.

Turning to the RHS of equations (4) and (9), the per-unit marginal cost $i^*/\mu^* = i^*/\mu$ of innovation needs to be compared to the public insurer’s value, $i^{**}/\rho$. If the government pursues a policy of accommodating deficits incurred by the public scheme, $\rho$ is low, calling for a low value of $i^{**}$ to balance the low marginal benefit of innovation. If however the probability of sanction is high (compared to $\mu$, the probability of a member of the population being an unfavorable risk), equation (9) may solve for a value of $i^{**}$ that exceeds $i^*$. On the whole, the impression is that competitive insurers might be somewhat more innovative.

Of course, this analysis is still partial in that the premium is fixed in both cases. If it is made to depend on $i$, the issue becomes of whether $e(P, i) > 0$ or $e(P, i) < 0$. In the first case, innovation involves a change in the benefit package that increases consumers’ WTP for insurance coverage. Usually, the new benefit package is more costly, causing the cost elasticity $e(L, i)$ to turn positive. In the second case, innovation means the launching of policies that impose restrictions on consumers, calling for a reduction in the contribution to health insurance. The question then becomes of whether $e(P, i)$ or
\( e(L,i) \) is greater in absolute value. Evidence from market experiments performed in Switzerland (Zweifel et al., 2005) suggests that while \( e(P,i) \) may be substantial on average, the decrease in WTP caused by managed care restrictions for some subpopulations may be less than \( e(L,i) \), i.e. the savings achieved by these options (Lehmann and Zweifel, 2004).

**Conclusion 2:** A simple model of behavior of a competitive health insurer subject to premium regulation predicts a weakening of interest in product innovation when regulation is by complemented by risk adjustment. Assuming realistic parameter values, a monopolistic public health insurer has a still weaker incentive to innovate.

**4. The design of an optimal health insurance contract**

The efficiency reasons given above for the existence of SHI with compulsory membership are convincing only if the design of the SHI contract is in some sense ‘optimal’ from the point of view of the representative consumer (see e.g. Zweifel and Breyer 1997, ch. 6). In view of the tendency towards full coverage of most SHI schemes, an issue of particular importance are the circumstances justifying copayments, i.e. deviations from full coverage of health care expenditures.

(a) **Administrative costs.** Copayment provisions can be called for to save administrative costs such as costs of handling claims. For this reason, and assuming expected utility maximization on the part of consumers, it is optimal to exclude partially or entirely expenditures on health care items that occur frequently but in limited amounts such as minor medications (Mossin 1968). More specifically, if administrative costs are proportional to the expected volume of health expenditures, a feature of the optimal insurance contract is a fixed deductible (Arrow 1963).
(b) Non-insurable loss. Illness typically involves not only financial costs but also non-financial losses such as pain and suffering. Optimal health insurance equalizes marginal utility of wealth in all states of nature but this is not equivalent to full coverage if there are complementarities between non-financial and financial losses. In particular, if marginal utility of wealth is lower in case of illness than in good health (due e.g. to reduced ability to enjoy expensive types of consumption), optimal health insurance does not fully reimburse the financial loss (Cook and Graham 1977).

(c) Ex-ante moral hazard. If the insurer cannot observe preventive effort on the part of the insured, a high degree of coverage reduces the incentive for prevention. Hence there is a trade-off between risk spreading through insurance and maintaining incentives to keep the risk of illness low. This trade-off leads to a premium function which is convex in the degree of coverage, such that full coverage should be particularly expensive (Ehrlich and Becker 1972). In SHI such a premium function is nowhere observed, although it could be easily administered because consumers cannot circumvent the progressive premium schedule by purchasing many insurance contracts with limited coverage and low premiums each.

(d) Ex-post moral hazard. If the insurer could observe the health status of the insured, the optimal type of health insurance would provide indemnity payments, i.e. the insurance payment would not depend on the insured’s health care expenditure. With asymmetric information, however, linking reimbursement to expenditure is inevitable. Still, copayment provisions are needed to fend against overconsumption of medical care. The optimal copayment rate is higher the more price elastic the demand for the particular type of medical services (Spence and Zeckhauser 1971, see also Zweifel and Breyer 1997, ch. 6). Empirical evidence e.g. from the RAND Health Insurance Study (Manning et al. 1987) shows that there is a small albeit significant price elasticity of demand for most medical services (for a survey of the evidence, see Zweifel and Manning 2000).

Conclusion 3: The optimal health insurance contract suggests several reasons for stopping short of full coverage. While administrative costs should be
recovered by a deductible, the presence of non-insurable losses may and
that of moral hazard definitely does commend a measure of proportional
cost sharing, reflecting the price elasticity of demand for medical care in
the last-mentioned instance.

5. The limits of social health insurance

There appear to be at least four types of limits to social health insurance. First, the
incentive structure of social insurers discussed in section 3 hampers product innovation.
Second, the features of the optimal contract as described in section 4 imply that
coverage provided by social health insurers needs to be limited in view of moral hazard
effects. A third limit of a more institutional character emanates from the fact that health
risks, while important, are only one type among several that need to be considered in
The fourth and ultimate limit of course is nothing but the willingness of citizens to pay
still higher contributions for higher-quality but more expensive health care, about which
some evidence will be presented at the end of this section.

5.1 Limits created by regulation

Traditionally, social insurance is associated with contributions that are not graded to
risk and even uniform across the population. This uniformity would be undermined if
competitive social health insurers were to launch new products that fetch a higher
premium. The only way to permit innovation is to let competitors with little potential to
increase their market share run experiments, which if successful are imposed on the
entire market. Of course, this is a far shot from the innovation process in actual markets,
where most innovations fail, adoption occurs with considerable lags (Henke, 2005), and
some competitors never adopt but already search for a still more promising alternative.
In the case of a monopolistic scheme, innovation is only possible through a majority
decision in parliament (or even a majority decision of the voters in a direct democracy).
These impediments cause the current provision of health insurance to lag behind
developments of preferences and restrictions prevalent in the population by years if not
decades (Zweifel and Breyer, 1997, ch. 11.1).
5.2 Limits imposed by the behavior of insurers
The analysis of section 3 shows that the capability of innovation, i.e. of adapting the insurance product to changing demands, is limited in SHI as traditionally understood. When risk adjustment is imposed in order to ‘marry’ uniform premiums in the face of different expected cost with competition, insurers considering innovation fear the financial sanction that goes along with attracting young clients. Finally, in the case of social health insurance not provided by a multitude of suppliers but a monopolistic scheme, the incentive for innovation is stifled even more.

5.3 Limits imposed by institutional design
Undoubtedly, health risks loom large in the lives of citizens. However, there are other risks confronting people over their life cycles. Adopting the categories of social insurance, one would want to distinguish the risks of accident, disability, old age, unemployment, increase in family size, and death of main breadwinner (Breyer, 2005; Zweifel, 2000b). While the relationships between these risks have not been fully researched yet, the available evidence points to positive correlations. This implies that the three assets to be managed over the life cycle, viz. health, wealth, and wisdom (Williams 1998) are likely to be positively correlated, an unfavorable situation for risk-averse individuals. Therefore, insurance as a system should at least mitigate these positive correlations, e.g. by paying higher than expected benefits in one branch if there is a shortfall below expected benefits in some other. However, preliminary research at the macroeconomic level suggests that in several important countries, trend deviations of payments are positively rather than negatively correlated across categories. For example, trend deviations in payments of German social unemployment insurance are positively correlated not only with those of old age and pensioners insurance but also of social, health and accident insurance (see table 2). Not one out of a total of six correlation coefficients is significantly negative.
Table 2: Correlations of trend deviations in the benefits of German social insurance, 1975-1993

<table>
<thead>
<tr>
<th></th>
<th>SOAS (1)</th>
<th>SEB (2)</th>
<th>SHI (3)</th>
<th>SAI (4)</th>
<th>SUI (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOAS (1)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEB (2)</td>
<td>-0.86*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHI (3)</td>
<td>0.45</td>
<td>-0.33</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAI (4)</td>
<td>0.91*</td>
<td>-0.81*</td>
<td>0.67*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>SUI (5)</td>
<td>0.76*</td>
<td>-0.75*</td>
<td>0.65*</td>
<td>0.83*</td>
<td>1</td>
</tr>
</tbody>
</table>

* Correlation coefficient significantly different from zero (5% significance level or better)

SOAS: Old age and survivors insurance
SEB: Employee benefits
SHI: Social health insurance
SAI: Social accidents insurance
SUI: Unemployment insurance


Also, social health insurance fails to make up for shortfalls in the benefits of private insurance. While employee benefits for old age, which are counted as social insurance, are negatively correlated with private health insurance (as they should be for portfolio variance reduction), trend deviations of all the other branches of social insurance correlate positively with at least one of the lines of private insurance (see table 3). Conversely, again not a single out of 20 coefficients of correlation is significantly negative.

Table 3: Correlations of trend deviations in the benefits of German private and social insurance, 1975-1993

<table>
<thead>
<tr>
<th></th>
<th>SOAS (5)</th>
<th>SEB (6)</th>
<th>SHI (7)</th>
<th>SAI (8)</th>
<th>SUI (9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLI (1)</td>
<td>0.27</td>
<td>-0.08</td>
<td>0.26</td>
<td>0.39</td>
<td>0.25</td>
</tr>
<tr>
<td>PHI (2)</td>
<td>0.79*</td>
<td>-0.72*</td>
<td>0.56*</td>
<td>0.92*</td>
<td>0.63*</td>
</tr>
<tr>
<td>PAI (3)</td>
<td>-0.41</td>
<td>0.28</td>
<td>-0.15</td>
<td>-0.41</td>
<td>-0.26</td>
</tr>
<tr>
<td>PGI (4)</td>
<td>0.43</td>
<td>-0.25</td>
<td>0.16</td>
<td>0.54*</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Correlation coefficient significantly different from zero (5% significance level or better)

PLI: Private life insurance
PHI: Private health insurance
PGI: Private general liability insurance
PAI: Private accident insurance
OAS: Old age and survivors insurance
SEB: Employee benefits
SHI: Social health insurance
SAI: Social accident insurance
SUI: Unemployment insurance

Of course, this argument could be turned around to read that German private insurers do not meet their task of making up for the shortfalls in benefits occurring within social insurance. However, as shown in table 4, as many as 3 out of 10 correlation coefficients are significantly negative this time, pointing to at least some diversification effects achieved within private insurance. On the whole, then, social health insurance in Germany (but in Austria, Switzerland, and the United States as well, see Zweifel and Lehmann, 2001) might be largely responsible for present insurance systems keeping the volatility of individuals’ assets larger than necessary.

Table 4: Correlations of trend deviations in the benefits of German social insurance, 1975-1993

<table>
<thead>
<tr>
<th></th>
<th>SOAS (1)</th>
<th>SEB (2)</th>
<th>SHI (3)</th>
<th>SAI (4)</th>
<th>SUI (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOAS 1</td>
<td>1</td>
<td>-0.86*</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SEB 2</td>
<td>-0.86*</td>
<td>1</td>
<td>-0.33</td>
<td>-0.81*</td>
<td>-0.75*</td>
</tr>
<tr>
<td>SHI 3</td>
<td>0.45</td>
<td>-0.33</td>
<td>1</td>
<td>0.67*</td>
<td>0.65*</td>
</tr>
<tr>
<td>SAI 4</td>
<td>0.91*</td>
<td>-0.81*</td>
<td>0.67*</td>
<td>1</td>
<td>0.83*</td>
</tr>
<tr>
<td>SUI 5</td>
<td>0.76*</td>
<td>-0.75*</td>
<td>0.65*</td>
<td>0.83*</td>
<td>1</td>
</tr>
</tbody>
</table>

* Correlation coefficient significantly different from zero (5% significance level or better)

SOAS    Old age and survivors insurance
SEB     Employee benefits
SHI     Social health insurance
SAI     Social accident insurance
SUI     Unemployment insurance

This (admittedly preliminary) evidence suggests that the same amount of resource could produce more security for people, or conversely, that the same amount of security could be afforded for less money. This of course serves to limit the attractiveness of social health insurance for consumers.

5.4 Limited willingness-to-pay of citizens

The call for reform of current social security systems frequently is based on the argument that they cannot be financed any longer. However, anything that does not exceed GDP can be financed in principle. The argument therefore must be watered down to the statement that the willingness-to-pay (WTP) of consumers is limited, and
mandated expenditure on any good or service in excess of that limit causes an efficiency loss. The problem with this argument has been that until recently, WTP for public goods in the health domain was not known.

In the case of Switzerland, some evidence has become available. In a discrete choice experiment involving 1,000 individuals in 2003, WTP for additional services to be provided (or rather, compensation required for accepting cutbacks) by social health insurance in exchange for an increased premium was measured (Telser et al., 2004; Zweifel et al., 2005).

Table 5: Compensation asked for cutbacks in Swiss social insurance (2003)

<table>
<thead>
<tr>
<th>Socioeconomic characteristics</th>
<th>Physicians selected on cost criteria (1)</th>
<th>Physicians selected on quality criteria (2)</th>
<th>Physicians selected on cost &amp; quality criteria (3)</th>
<th>Access to new therapies and drugs delayed by 2 years (4)</th>
<th>Reimbursement of generics only (5)</th>
<th>No reimbursement of drugs for minor complaints (6)</th>
<th>No small local hospitals (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total sample</td>
<td>103 (13.2)</td>
<td>53 (8.8)</td>
<td>42 (7.8)</td>
<td>65 (7.9)</td>
<td>3 (5.5)</td>
<td>-6 (5.3)</td>
<td>37 (5.7)</td>
</tr>
<tr>
<td>By region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German-speaking</td>
<td>88 (11.8)</td>
<td>38 (7.8)</td>
<td>26 (6.8)</td>
<td>56 (7.1)</td>
<td>5 (5.5)</td>
<td>-5 (5.3)</td>
<td>31 (5.2)</td>
</tr>
<tr>
<td>French-speaking</td>
<td>191 (76.3)</td>
<td>138 (58.5)</td>
<td>136 (56.9)</td>
<td>117 (45.4)</td>
<td>-14 (19.6)</td>
<td>-13 (19.2)</td>
<td>74 (31.0)</td>
</tr>
<tr>
<td>By average monthly income per household member</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; CHF 1500</td>
<td>67 (17.5)</td>
<td>44 (14.7)</td>
<td>35 (13.4)</td>
<td>52 (12.2)</td>
<td>-5 (10.0)</td>
<td>-2 (9.7)</td>
<td>28 (9.1)</td>
</tr>
<tr>
<td>CHF 1500 to 4000</td>
<td>108 (17.5)</td>
<td>56 (11.6)</td>
<td>42 (9.9)</td>
<td>66 (10.3)</td>
<td>9 (7.2)</td>
<td>-5 (6.8)</td>
<td>42 (7.8)</td>
</tr>
<tr>
<td>CHF 4000+</td>
<td>148 (55.8)</td>
<td>62 (29.9)</td>
<td>63 (29.7)</td>
<td>81 (29.4)</td>
<td>-14 (17.5)</td>
<td>18 (17.8)</td>
<td>33 (16.8)</td>
</tr>
</tbody>
</table>

Note: 1 CHF equals 0.7 € at 2003 exchange rates. Standard errors in parentheses.

Reading table 5 horizontally first, one notes that the amounts of compensation asked are consistently highest for consenting to a physician list based exclusively on cost criteria (col. 1). The sample average is as high as CHF 103 (€ 67 at 2003 exchange rates), or some 38 percent of the country’s average monthly premium of CHF 270 at the time. Still, the fact that it is finite speaks against the claim (often advanced by medical
associations) that free choice of physician is virtually priceless. Selecting physicians on quality or quality and cost (i.e. efficiency) criteria already requires a lot less compensation, viz. 20 and 16 percent of premium, respectively (col. 3). These premium reductions can be granted by current Managed Care alternatives available in Swiss social health insurance. A possible delay of access to new therapies and drugs by two years would meet with much more resistance; it would have to be compensated by no less than CHF 65 or 24 percent of premium on average. Limiting the drug benefit to generics only (col. 5) would have to be compensated by small amounts only that cannot be distinguished from zero. If the drug benefit were not to reimburse drugs used for the treatment of minor complaints, the Swiss on average would even be willing to pay a small amount [which again is not distinguishable from zero and likely reflects ‘warm glow’ (Andreoni, 1995)]. Finally, another cutback would be the concentration of dispersed existing hospital capacity in larger, centralized units (col. 7 of Table 5). In spite of the alleged superior efficiency of such units, this regulation would have to be compensated by CHF 37 or about 14 percent of the average monthly premium.

An argument against SHI in this context is preference heterogeneity. If preferences differ, the uniformity imposed entails an efficiency loss. A first sign of preference heterogeneity is the fact that compensation asked differs importantly between income classes. Reading table 3 vertically, one notes that for example, individuals belonging to the top income class demand a compensation of 220 percent the amount demanded by those of the lowest income class for voluntarily accepting a physician list based on cost considerations only (col. 1). Of course, wealthy individuals can always opt out by paying extra; however, poor individuals do suffer a loss because the required reduction by CHF 67 or 25 percent in premiums could in fact be achieved by at least one health insurer if premium regulation permitted it to pass on its savings from Managed Care to consumers (Lehmann and Zweifel 2004).

Preference heterogeneity is also reflected in amazingly large regional differences. In the case of accepting a physician list based on cost and quality criteria (col. 3), the French-speaking minority of Switzerland is so distrustful as to ask for a compensation of no less than CHF 136 per month, more than five times as much as the German-speaking
majority. Their WTP to avoid other restrictions is consistently more than twice as high as that of German-speaking Swiss.

**Conclusion 4:** There are several limits to social health insurance, ranging from the behavior of social insurers on to moral hazard effects, institutional design preventing correlations between risks to be accounted for, and to a willingness to pay for additional coverage that falls short of its additional cost.

**6. Summary and conclusions**

This contribution revolves about two related issues, viz., What are the reasons for the existence and growth of social health insurance?, and Are there limits to social health insurance? As to the reasons, demand for social health insurance may well reflect the demand for an efficiency-enhancing invention that overcomes certain market failures plaguing private insurance markets. In addition, equity considerations may also provide a powerful motive. On balance, however, the (scanty) available evidence points to a preponderance of public choice reasons. Social (health) insurance can be seen as an efficient instrument for gaining votes in the hands of politicians seeking (re)election (Conclusion 1 in the text).

Turning to the supply of social health insurance, two settings need to be distinguished. One is provision by competitive health insurers who are regulated with regard to premiums and most products, the other, by a monopolistic public scheme. A simple model generates the prediction that completing regulation by risk adjustment (whereby insurers having an above-average share of low risks must pay into a fund that subsidizes those having an above-average share of high risks) undermines incentives for product innovation. The basic reason is simple. Innovation tends to attract the young, who are deemed to be low risks in all existing risk adjustment schemes; it therefore induces a financial sanction. The monopolist insurer also pursues product innovation to the extent that it lowers insurance payments (which is of interest to political authorities); however, its incentives are weaker than the competitive insurer’s given reasonable parameter constellations (Conclusion 2).
For exploring the limits of social health insurance, the theory of the optimal health insurance contract serves as a point of departure. And indeed, this theory calls for a deductible to recover the administrative cost of providing health insurance. In addition, it may suggest partial coverage only in the case where the (marginal) utility of wealth is comparatively low in the state of sickness, causing material goods not to be very valuable. It specifically suggests a positive rate of coinsurance to combat moral hazard effects. However, there are additional limits to social health insurance. An important one derives from its institutional nature. From the point of view of risk-averse citizens, an ‘umbrella policy’ covering not only the risk of illness but also those of accident, disability, early death of the breadwinner, (unplanned) additions to the family, and insufficient income in old age could be of considerable advantage to the extent that these risks cause their assets health, wealth, and wisdom (skill capital) to be positively correlated. However, consumers may well shy away from creating a public monopoly insurer with the task and authority to cover all these risks jointly. They might be more inclined to entrust this task to a competitive insurer that can be exchanged for another if failing to deliver. These considerations put another limit on social health insurance.

Finally, political pressure to constrain social health insurance (and social security more generally) may reflect marginal willingness to pay on the part of citizens below marginal cost. Conversely, compensation asked for accepting restrictions in the domain of social health insurance (in the guise of reduced contributions) could be financed by health insurers through cost savings achieved. Recent evidence from Switzerland relates to this second approach. It suggests that if health insurers were permitted to fully pass on savings accruing e.g. in their Managed Care options, they could compensate the average consumer sufficiently to make this options attractive. In all, there are clear signs of social health insurance encountering several limits (Conclusion 4).

These limits will become more important in the future as the cost of health care increasingly occurs towards the end of human life, when they cannot be recouped by increased contributions any more. Moreover, social health insurance, by modifying the incentives of the great majority of health care providers of a country, induces the very change in medical technology that causes the cost of health care to increase so fast
(Zweifel, 2003). The challenge will be to devise contracts that create incentives for the consumers to make do with the second-latest medical technology when they are death-bound. However, competitive private rather than regulated social health insurers likely are better capable of meeting this type of challenge.

**Appendices: Comparative statics**

**A. Competitive social insurer**

The point of departure is equation (4) and (5) in the text, slightly rewritten,

\[(A.1) \quad e(\mu, i) \left\{ (P - \pi^L) - (P - \pi^L') \right\} - e(L, i) \left\{ \mu \pi^L + (1 - \mu) \pi^L' \right\} / \mu = \frac{i}{\mu(i, s)} \]

\[(A.2) \quad e(\mu, s) \left\{ (P - \pi^L) - (P - \pi^L') \right\} = \frac{s}{\mu(i, s)}.\]

The effect of an increase in regulation (possibly from a state of no risk-adjustment scheme) is to decrease the difference in expected margins of high and low risks,

\[(A.3) \quad \frac{\partial}{\partial \pi} \left[(P - \pi^L) - (P - \pi^L')\right] < 0 \Leftrightarrow \frac{\partial A}{\partial \pi} < 0.\]

For future reference, one also has

\[(A.4) \quad \frac{\partial}{\partial i} \left( \frac{i}{\mu(i, s)} \right) = \frac{\mu - i \cdot \partial \mu / \partial i}{\mu^2} = \frac{1 - e(\mu, i)}{\mu}.\]

Now, let the first-order conditions (A.1) and A.2) be subjected to a stock \(dR > 0\).

Written in matrix form, the comparative statics read, using (A.2) to (A.4),

\[
(A.5) \quad \left[ \begin{array}{cc}
\dfrac{e(\mu, i) - 1}{\mu} & 0 \\
0 & \dfrac{e(\mu, s) - 1}{\mu}
\end{array} \right] \left[ \begin{array}{c}
di \\
ds
\end{array} \right] = \left[ \begin{array}{c}
-e(\mu, i) \cdot \partial A / \partial R \\
-e(\mu, s) \cdot \partial A / \partial R
\end{array} \right] dR.
\]

Applying Cramer’s rule, one obtains
\[\frac{di}{dR} = \frac{1}{H} \begin{vmatrix} -e(\mu,i) \cdot \partial A / \partial R & 0 \\ -e(\mu,s) \cdot \partial A / \partial R & e(\mu,s) - 1 / \mu \end{vmatrix} < 0,\]

with \(H < 0\) symbolizing the determinant of the Hessian matrix. Applying Cramer’s rule once more yields

\[\frac{ds}{dR} = \frac{1}{H} \begin{vmatrix} e(\mu,i) - 1 / \mu & e(\mu,i) \cdot \partial A / \partial R \\ 0 & -e(\mu,s) \cdot \partial A / \partial R \end{vmatrix} < 0.\]

Thus, both innovative and risk selection effort are predicted to decrease provided \(e(\mu,i) < 1\), which looks like a reasonable assumption.

For comparison purposes, the predicted response to an increase in the amount of loss, \(dL > 0\), is also derived. In this case, the comparative static equations read,

\[\begin{bmatrix} 0 & \mu \\ 0 & -[e(\mu,i) - 1] / \mu \end{bmatrix} \begin{bmatrix} di \\ ds \end{bmatrix} = \begin{bmatrix} e(\mu,i) + e(L,i)(\mu \pi^b + (1 - \mu) \pi^i) \\ e(\mu,s)(\pi^b - \pi^i) \end{bmatrix} dL.\]

This can be solved to obtain (with \(e(\mu,s) = 0\))

\[\frac{di}{dL} = \frac{1}{H} \begin{vmatrix} e(\mu,i)(\mu \pi^b - \pi^i) + e(L,i)(\mu \pi^b + (1 - \mu) \pi^i) & 0 \\ -e(\mu,s)(\pi^b - \pi^i) & e(\mu,s) - 1 / \mu \end{vmatrix} = (1 / H) \begin{vmatrix} e(\mu,i)(\pi^b - \pi^i) + e(L,i) & [e(\mu,s) - 1] / \mu \\ e(\mu,i)(\mu \pi^b - \pi^i) + e(L,i)(\mu \pi^b + (1 - \mu) \pi^i) & e(\mu,s) - 1 / \mu \end{vmatrix} < 0 \text{ if } e(\mu,s) < 1.\]

**B. Public monopoly health insurer**

Slightly rewriting equation (9), one has
Subjecting this first-order condition to an impulse $dL > 0$, one obtains for the remaining decision variable $i$,

\[(B.4) \quad \{e(\rho, D) \cdot e(L, i)\} \{\mu \pi^b L + (1-\mu)\pi^s L\} - e(L, i)\{\mu \pi^b L + (1-\mu)\pi^s L\} dL = di / \rho.\]

After collecting terms, this becomes

\[(B.5) \quad \frac{di}{dL} = \rho \left\{ e(L, i) \left[ \frac{e(\rho, D)}{e(\rho, D) - 1} \right] \right\} \begin{cases} < 0 & \text{if } e(\rho, D) > 1 \\ > 0 & \text{if } e(\delta, D) < 1. \end{cases}\]

**References**


