The theory of social health insurance

Zweifel, P
The theory of social health insurance

Abstract

The objective of this text is to develop the theory of social health insurance (SHI; the expression used especially in the United States is “public health insurance,” which will be viewed as one variant of SHI here). While a good deal is known about the demand and supply of private insurance, the theoretical basis of SHI is much more fragile. Specifically, on the demand side, what are the reasons for social (or public) health insurance to exist, even to dominate private health insurance in most developed countries? With regard to supply, what do we know about the objectives and constraints of SHI managers? Finally, economists can predict properties of the equilibrium characterizing private health insurance (PHI). However, what is the likely outcome (“performance”) of SHI? At the normative level, one may ask, Should the balance be shifted from SHI to PHI?
The Theory of Social Health Insurance

Peter Zweifel*

University of Zürich, Socioeconomic Institute, Hottingerstr. 10 CH-8032, Zürich, Switzerland, pzweifel@soi.unizh.ch

Abstract

The objective of this text is to develop the theory of social health insurance (SHI; the expression used especially in the United States is “public health insurance,” which will be viewed as one variant of SHI here). While a good deal is known about the demand and supply of private insurance, the theoretical basis of SHI is much more fragile. Specifically, on the demand side, what are the reasons for social (or public) health insurance to exist, even to dominate private health insurance in most developed countries? With regard to supply, what do we know about the objectives and constraints of SHI managers? Finally, economists can predict properties of the equilibrium characterizing private health insurance (PHI). However, what is the likely outcome (“performance”) of SHI? At the normative level, one may ask, Should the balance be shifted from SHI to PHI?

*The author would like to thank an anonymous referee and Patrick Eugster (University of Zurich) for criticisms and suggestions and Boris Krey (University of Zurich) for providing the case studies in Section 3.
The objective of this text is to develop the theory of social health insurance (SHI; the expression used especially in the United States is “public health insurance,” which will be viewed as one variant of SHI here). While a good deal is known about the demand and supply of private insurance, the theoretical basis of SHI is much more fragile. Specifically, on the demand side, what are the reasons for social (or public) health insurance to exist, even to dominate private health insurance in most developed countries? With regard to supply, what do we know about the objectives and constraints of SHI managers? Finally, economists can predict properties of the equilibrium characterizing private health insurance (PHI). However, what is the likely outcome (“performance”) of SHI? At the normative level, one may ask, Should the balance be shifted from SHI to PHI?

Accordingly, the outline of this text is as follows. Section 2 starts by reviewing the conventional theory of demand for insurance and health insurance in particular. However, it also seeks to offer explanations of the demand of SHI, citing efficiency, public choice, and equity reasons. They may explain the existence (but not necessarily the prominence) of SHI. Section 3 is devoted to the supply of health insurance in general
and SHI in particular, which comprises more dimensions than just price and quantity. Section 4 reviews the properties of the optimal health insurance contract for providing a benchmark especially with regard to combating moral hazard. In Section 5, the question is asked whether there are factors limiting the apparently inexorable growth of SHI. Section 6 offers a summary and concluding remarks.
2

The Demand for Social Health Insurance

2.1 Theory of Insurance Demand

The standard way to present the theory of insurance demand is using a two-goods model, with wealth in the no-loss state and wealth in the loss state constituting the two goods. Here, a simpler alternative, based on the Von Neumann–Morgenstern (VNM, henceforth: risk-utility) function will be presented. In Figure 2.1, there are two levels of wealth, $W_l$ in the loss state and $W_n$ in the no-loss state. The associated utilities are $U[W_l]$ and $U[W_n]$, where $U[W_l] < U[W_n]$, the bracket to be interpreted to mean that the utility function $U(\cdot)$ is to be evaluated at the respective values of the argument. Expected utility is given by (see e.g., Zweifel and Breyer, 1997, ch. 6),

$$EU = \pi \cdot U[W_l] + (1 - \pi)U[W_n],$$

$$W_l := W_0 - L - P(I) + I,$$

$$W_n := W_0 - P(I),$$

(2.1)

with $\pi$ denoting the probability of loss ($0 < \pi < 1$), $P$ the premium, and $I$ the amount paid by insurance in the event of loss. In Figure 2.1,
the case \( \pi = 1/2 \) is shown. Clearly, the expected utility \( EU \) is associated with the expected value of wealth, \( EW \). It is a linear combination of utilities \( U[W_l] \) and \( U[W_n] \). It is well-known that linear combination of values lie on the connecting straight line.

Now consider an individual who has the possibility of escaping this risky prospect, in which a high value and a low value of wealth may be realized with a certain probability. Given that the alternative providing certainty would be financially equivalent \( (W = EW) \), a risk-averse decision maker would opt for it. This means that the risk-utility function must pass above the point \( EU \), e.g., through \( U[EW] > EU(W) \). The remainder of the risk utility function can be constructed as follows. The loss-state with probability \( \pi = 1 \) is no different from the certain unfavorable outcome, and likewise loss-state with probability \( \pi = 0 \) is equivalent to the certain favorable outcome. Therefore, the risk-utility function at \( \pi = 1 \) and \( \pi = 0 \) cannot differ from the extreme points of the linear combination (that defines \( EU(W) \)). On the whole, the risk-utility function must run concave from below, i.e. \( U''(W) < 0 \).

An interesting implication follows from indifference between risky and certain alternatives, which are depicted as a horizontal line through the point \( EU \), intersecting the risk-utility function at point \( Q \). The associated value of wealth (point \( C \)) is called the certainty equivalent.
of wealth. It shows that risk-averse individuals accept a reduction in their wealth if this permits them to escape the risky situation. The more marked the curvature of the risk utility function, the more risk-averse the individual considered, and the greater the difference between the expected value of wealth $EW$ and the certainty equivalent $C$. This difference can also be interpreted as a willingness-to-pay (WTP) for certainty.

Turning to insurance, a policy with full coverage in fact offers certainty in terms of wealth. Therefore, risk-averse individuals also have WTP for insurance if their assets are exposed to variability. In Figure 2.1, $P_{\text{max}}$ is the maximum total premium that such an individual is willing to pay. It consists of two components. First, $EL$ shows the expected value of the loss. This is also called the actuarially fair premium. The excess of $P_{\text{max}}$ over $EL$ is equivalent to the maximum loading for administrative expense and profit that an insurer offering the full coverage contract can charge the consumer depicted in Figure 2.1. Clearly, this loading also depends on the degree of risk aversion of the individual; without risk aversion, the risk utility-function would run linear, causing $P_{\text{max}}$ and $EL$ to coincide. Therefore, there would be no WTP for insurance.

### 2.2 The Demand for Health Insurance

The model of the preceding section is not satisfactory for health insurance because it is couched exclusively in terms of wealth. One approach would be to enter health status $H$ in the risk-utility function. A far easier alternative is to continue to work with a risk-utility function in terms of wealth only, but to make its shape depend on health status. First, the risk-utility function conditional on good health has a higher value than that conditional on bad health, i.e., $U_h(W) > U_s(W)$, for all levels of wealth $W$. Second, however, it is not so much the difference in levels but the difference in slopes that is crucial for the optimal amount of coverage. The argument will be only developed for the case where the premium is actuarially fair,

$$P(I) = \pi I.$$  (2.2)
Substituting Eq. (2.2) into Eq. (2.1), modified to comprise $U_h(W)$ for the healthy state and $U_s(W)$ for the sick state, and taking the first-order derivative with regard to insurance coverage $I$, one obtains

$$\frac{dEU}{dI} = \pi U'_s[W_l] (-\pi + 1) + (1 - \pi) U'_h[W_n] (-\pi) = 0. \quad (2.3)$$

Dividing this by $\pi \cdot (1 - \pi)$, one has

$$U'_s[W_l] = U'_h[W_n]. \quad (2.4)$$

Therefore, given actuarially fair premiums, the optimum for the potential buyer of health insurance is equality of the two marginal utilities of wealth. This makes sense, because as long as additional wealth is worth more in one state than the other, the consumer should reallocate wealth between the two states.

Turning to Figure 2.2 (panel A), the upper risk-utility function runs steeper throughout than the lower one, indicating that the marginal utility of wealth is higher in the healthy state than in the sick state for all values of $W$. The two parallel dashed lines show a possible solution that satisfies the equality of marginal utilities as given by Eq. (2.4). The optima are represented by points $Q$ and $R$ on the state-dependent risk-utility functions. They imply that optimally wealth should be higher in the healthy state than in the sick state. Therefore, insurance coverage should not be complete but contain a degree of cost sharing (which

![Fig. 2.2 Optimal degree of coverage in health insurance.](image-url)
of course can also be justified on other grounds, in particularly, moral hazard).

Turning to panel B, the situation is exactly reverse. Here, the lower-valued risk utility function runs steeper than the higher-valued one throughout. Applying the marginal utility criterion of Eq. (2.4) once more, one obtains optimal points $S$ and $T$. This time, these points indicate that wealth optimally should be higher in the sick state than in the healthy state. The interpretation is that possibly when an individual is ill, good accommodation, healthy food, and convenient clothing are more important than in the healthy state. For such an individual, health insurance, at least in the absence of moral hazard, should provide compensation for suffering. While this is a possibility, on the whole a higher marginal utility of wealth in the healthy state (the case of panel (A)) should be considered more realistic. Many consumption activities (eating, but also travel and entertainment) can only be enjoyed fully when in good health. Of course, there are additional considerations calling for a degree of cost sharing, and arguing against contracts that pay more than the medical expenditure needed for re-establishing health. Moral hazard effects need to be avoided, and separating contracts (see Section 2.3.1) call for offering less than complete coverage to favorable risks.

**Conclusion 1.** The theory of insurance demand predicts that risk-averse individuals derive benefit at least on expectation from health insurance, provided the premium does not contain an excessive loading for administrative expense and profit. To the extent that wealth is particularly important when ill, optimal coverage may even contain a compensation for suffering; however, this result may not hold under the influence of moral hazard.

### 2.3 Why Social Health Insurance?

Most developed countries have some kind of collective financing for health services, either through tax (e.g., the Health Service of 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 enrollment, 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 seeks to explain its existence in democracies on the basis of public choice models.

2.3.1 Efficiency reasons: Characteristics of private health insurance markets

SHI may be efficiency-enhancing if it mitigates or even eliminates possible market failures, viz. high cost of administration, asymmetry in the distribution of information, altruism and free riding, and optimal taxation.

(a) High cost of administration: The theory of the preceding section was couched in terms of actuarially fair premiums. This means that on expectation, insurance coverage is costless because the premium revenue is redistributed back entirely to consumers. In fact, however, premiums charged by private health insurers contain a considerable loading for acquisition, administrative expense, underwriting risk, and profit. One argument in favor of SHI is that thanks to its compulsory nature, it saves on acquisition cost and thanks to its uniformity, on administrative expense. Judging from insurers’ accounts (which however do not single out health insurance), these savings may amount to as much as 30 percent of premium revenue.

However, contributions to SHI come very close to being a tax. In many countries (such as Canada and Sweden), they are part of taxation; in other countries (such as Germany, the Netherlands, and the United States), they are part of the cost of labor and hence amount to a payroll tax. Even where social health insurers charge an actual premium (as in Switzerland), contributions cannot be reduced below the premium of the most efficient insurer, who must recover
The cost of mandated benefits. This means that there is an excess burden associated with these contributions, in analogy to income taxation. An estimate by Ballard et al. (1985) puts the excess burden at 30 percent of tax revenue in the United States, while a more recent estimate by Parry (2001) arrives at 26 percent specifically for payroll taxes in the United Kingdom. Therefore, social health insurance is unlikely to be efficiency-enhancing on net with regard to the loading.

(b) 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 has more precise information on his individual risk distribution than the insurer, the only possible Rothschild–Stiglitz equilibrium is a separating one in which the most unfavorable 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 an 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 favorable 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 these days since medical exams are used to determine the risk of an insured.

(c) Altruism and free riding: Altruistic rich members of a society may be willing to subsidize the provision of health care to the poor, especially if they are more interested in the health than in the subjective well-being of the poor (Pauly, 1970). Private charity fails to reach an efficient allocation since donations to the poor, whether in cash or in kind, have
2.3. Why Social Health Insurance?

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 (NHS) or SHI with compulsory membership and contributions according to ability to pay solve this free-rider problem of potential donors.

(d) 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 uniform contributions 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.

2.3.2 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 so that those persons who are endowed by nature with a lower stock of “health capital” and are thus already disadvantaged, have to pay a higher price for the same coverage on top of this. Behind the veil of ignorance, one would desire at least an equalization of the monetary costs of illness.
There are in principle two ways to achieve solidarity in health insurance (see Table 2.1). First, PHI premiums can be subsidized for those who would have to pay excessive contributions. The transfer could be on a current basis or 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 alternative is a monopolistic SHI scheme with open enrollment 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.

### 2.3.3 Public choice reasons

In PHI, redistribution occurs purely by chance, from consumers who did 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.
2.3. Why Social Health Insurance?

from high risks to low risks. In SHI, this redistribution not only affects wealth through its financing side but its benefit side as well, viz. medical services and health. This makes SHI 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. This has two possible 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.

(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. (2006)). Insurance coverage for everyone 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 and The Insurance Experiment Group (1993), ch. 11). However, there is also a price effect because health insurance boosts
the “true” 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. To the extent that rich individuals have higher true WTP to begin with, they benefit more strongly (in absolute terms) from this leverage effect of health insurance. Thus, the total redistributive effect of SHI is ambiguous.

(c) Redistribution of health: When it comes to health, altruism 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 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 GDP shares, 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 8 to 9 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 SHI schemes to benefit pivotal voter groups.
Conclusion 2. While the efficiency reasons for social (health) insurance have received much attention in the economics 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 factors (i.e., winning votes) may well be the crucial reason for the existence and even more the growth of social (health) insurance.
The Supply of Health Insurance

According to Conclusion 1 stated at the end of Section 2, governments (and public administration) can be seen as the suppliers of SHI. In systems of the National Health Service type, the government itself provides the insurance function while also acting as the organizer of medical care. This type will be called “public health insurer.” 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.” In both cases, the supply of health insurance has several dimensions. It can be characterized by the comprehensiveness and structure of the benefit package, the amount of effort devoted to risk selection, the price of coverage, the amount of integration of healthcare providers, and the market structure of health insurance.

3.1 Benefit Package

An unregulated private insurer has the option to specify its offer along three dimensions (Zweifel and Breyer, 1997, p. 159). First, it can decide to cover only certain types of services and leave out others, for instance,
3.1 Benefit Package

Type of Service

Amount of Benefits Type of Provider

Fig. 3.1 Differentiation of benefits.

to include inpatient and exclude outpatient care, which is not uncommon in low-income countries (Musau, 1999). Second, it can differentiate its offer by covering or excluding services offered by certain provider categories, for instance, including only physicians registered with a public agency and excluding those who are not. Third, it may determine the amount of the benefits paid in case of sickness. The compensation may state a certain quantity of services, the compensation per unit of consumption, or the limit up to which expenditures are refunded (see Figure 3.1).

There are many possible combinations between the three dimensions, creating opportunity for product innovation and the building of profitable market segments. The optimal choice is influenced by several factors listed in Table 3.1, which are discussed starting with the insurer’s point of view and moving toward demand-side considerations.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private health insurance</th>
<th>Community-based Health Insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk aversion of insurer</td>
<td>+/−</td>
<td>+/−</td>
<td>−−</td>
</tr>
<tr>
<td>Synergies among benefits</td>
<td>+</td>
<td>+</td>
<td>−−</td>
</tr>
<tr>
<td>Moral hazard</td>
<td>−</td>
<td>−</td>
<td>−−</td>
</tr>
<tr>
<td>Diversity of preferences</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Diversity of risks</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Emergence of new health risks</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Regulation</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Fraud and abuse</td>
<td>−</td>
<td>−</td>
<td>−−</td>
</tr>
</tbody>
</table>

Note: ↑ Reinforcement of relationship, ↓ Attenuation of relationship, n.a.: not applicable.
and regulatory and institutional factors that affect insurers’ decision making.

3.1.1 Risk aversion of insurer

The relevance of risk aversion for the behavior of insurers has been the subject of continued debate (Greenwald and Stiglitz, 1990, Chen et al., 2001). In industrial countries, owners of insurance companies can be assumed to hold fully diversified portfolios. As such, they are exposed only to nondiversifiable risk, which is reflected in the beta of the company (the slope of the regression linking the company’s expected rate of return to the expected rate of return prevailing on the capital market at large). Therefore, diversification is only in the interest of shareholders to the extent that it lowers the company’s (positive) value of beta. Management, being much less diversified in its assets, has an interest in diversification of its own. Therefore, the extent to which it actually engages in diversification of the underwriting portfolio is a question of corporate governance.

Assuming an interest in risk diversification caused by risk aversion, its impact on the benefit package can still go either way (see Table 3.1). To the extent that, e.g., inpatient and outpatient services constitute complements rather than substitutes, they are positively correlated. Including both in the benefits package then adds to the variance of liabilities ceteris paribus, which runs counter the interests of a risk-averse insurer. Benefits triggered by communicable diseases have the same effect, motivating their strict limitation. Even if there is negative correlation, it should be noted that risk diversification does not necessarily imply more complete benefit packages at the individual level since the insurer can offer different packages to different client groups.

To the extent that domestic investors in low-income countries cannot rely on a sufficiently developed capital market (or are prevented from full international diversification), their risk aversion is more likely to be relevant for management decisions. Management, finding itself in a similar situation, will tend to further reinforce this tendency (assuming corporate governance to be imperfect). In community-based health insurance, in particular, which amounts to a mutual insurer, owners
are individuals and households, whose degree of asset diversification is still far lower. This calls for an even keener interest in diversification. However, the low income level of enrollees may force most such schemes to stick to narrowly defined products in spite of a basic need for diversification (Musau, 1999).

A public health insurer is unlikely to be significantly risk averse with respect to its financial results. Its opportunities to shift the financial risk to the government – who can resort to printing money if necessary – and the responsibility for failure are numerous. Therefore, risk aversion cannot have much importance in determining the benefit package.

3.1.2 Synergies among benefits

Synergies denote economies of scope in production, distribution, and marketing that are unrelated to risk diversification effects. They cause insurers to benefit from offering a combination of benefits rather than a single benefit. In production, synergies arise when the costs of writing and executing contracts (specifically the processing of claims, cf. the term $\mu \pi$ in Eq. (3.2) of Section 3.3) do not rise proportionally with the number of benefits, resulting in decreasing expected unit cost. In distribution, the same channel may be used for selling additional products. In marketing, brand advertising benefits all the products sold by a given insurer.

In a public insurance system, synergies are not a very relevant criteria for a decision maker who aims at providing public and merit goods to the population (see Section 3.3). This objective tends to override the economic justification of extending benefits purely because of synergies.

3.1.3 Moral hazard

The effect of ex-post moral hazard (for a definition, see Section 3.3.6) on the benefit package can be illustrated as follows. Assume that consumers’ willingness to pay (WTP) out of pocket for a medical service or product is approximately given by the linear demand function $C''$ of Figure 3.2. In the case of health insurance with a 50 percent coinsurance rate, maximum WTP is doubled, from $C'$ to $C''$. More generally, the demand function is rotated outward to become the effective demand
The Supply of Health Insurance

function \( C''C \). The lower the rate of coinsurance, the more pronounced this rotation is. With no copayment at all (as is often the case with tax-funded schemes), the curve runs fully vertical from \( C \).

Therefore, the market equilibrium shifts from point \( E \) to \( F \), with a higher quantity of the service or product transacted. In terms of Eq. (3.1) of Section 3.3, the benefits to be paid (\( I \)) increase, resulting in an ex-post moral hazard effect. As will be argued in Section 3.3.6, a decrease in the rate of coinsurance causes both parts of the loading and hence the premium to increase. This creates a negative income effect (shifting the demand curve inward) that is neglected for simplicity.

The moral hazard effect is of relevance to the choice of benefit package because it comes to bear with each additional item in the package. The more complete the package, the larger the loading component in the gross premium and hence the larger the net cost of insurance. Therefore, moral hazard considerations should lead an insurer to exercise caution in expanding the package. Specifically, it would want to
add services characterized by low price elasticity of demand because the moral hazard effect is more limited in this case. In Figure 3.2, lower price elasticity means that for a given maximum WTP such as $C'$, the demand function runs steeper, causing point $C$ to shift toward the origin. This serves to reduce the difference between the true and the observed demand curve, and hence the size of the ex-post moral hazard effect.

Moral hazard may be less of a problem in community-based schemes (see Table 3.1), which usually consist of small risk pools. First, asymmetric information is less pronounced in a small (often rural) community, where each member of the pool can easily monitor the behavior of the others.

In a public insurance system, moral hazard sooner or later becomes an important consideration in the determination of the benefit package. The consumption of health care services usually entails little or no cost sharing for the user, which means that in Figure 3.2 the vertical observed demand function applies. Therefore, the public insurer must finance the maximum quantity $C$ times the unit price $CD$ for each benefit added. It is subject to the ex-post moral hazard effect to a higher degree than a private insurer, who would offer policies with varying degrees of cost sharing. Unless contributions (often levied in the guise of a payroll tax) or tax allocations are increased accordingly, the scheme ends up in deficit.

3.1.4 Diversity of preferences

The creation of a benefit package depends on its value to consumers. Consumers will demand a package that combines benefits to the extent that their marginal rate of substitution is equal on expectation. A unit of benefit will be added to the package until its ratio of expected marginal utility to the premium increase occasioned is equal across all benefits. This expected value importantly depends on the amount of risk aversion and the relevant probabilities of loss. Differences in loss probabilities are addressed below.

Diversity of preferences among the insured causes their optimality conditions to be satisfied at different (sometimes zero) levels of benefits.
In order to attract consumers, insurers will customize their products in an attempt to maximize expected profit. The diversity of preferences may relate to, e.g., the amount of the deductible, the rate of coinsurance, and the limits on benefits, as well as type of service (for instance alternative medicine) and type of provider. In this way, permanent innovation and adjustment to changing demand occurs. As a general rule, product differentiation is costly.

Public health insurers cannot accommodate different preferences almost by definition because their mission is to administer the entire population (or at least a great majority of it) a uniform product. The more preferences in fact differ, the more likely is it that a uniform national health insurance scheme burdens the country with a loss of efficiency (for some empirical evidence, see Section 5.4).

### 3.1.5 Diversity of risks

Diversity of risks (in the sense of differences in loss probabilities) promote a differentiation of degrees of coverage, combined with a differentiation of premiums. If insurers are unable to assess risks, a differentiation of premiums cannot take place, which encourages the purchase of excess coverage by high risks and reduced coverage by low risks. Therefore, the insurer runs the danger of incurring a deficit when expanding the benefit package under these conditions. The same argument holds when the insurer is prevented from differentiating premiums by a mandate to take on every applicant on the same conditions. When combined with asymmetric information, diversity of risks thus hampers the creation of comprehensive benefit packages (see Table 3.1).

For a public health insurer, uniformity of benefits is part of its mission because it acts on behalf of the government, whose likely objective is to provide citizens with a maximum of public and so-called merit goods. By assumption, public goods are enjoyed by everyone to the same degree; therefore, if the government views access to health care as a public good, its insurance branch must act accordingly, guaranteeing equal access through equal benefits. Diversity of risks can hardly be reflected in a diversity of (planned) benefits under these circumstances.
3.1.6 Emergence of new health risks

New health risks give rise to demand for an extension of the benefit package. However, even under competitive conditions insurers will not adjust to this demand immediately. First, they need time to assess the probability of loss $\pi$. Second, an extension of the benefit package calls for a premium adjustment, which in turn usually requires a cancellation of the policy. It takes new business to provide the insurer with the opportunity to test consumers’ WTP for the added benefit in the guise of a higher premium. Even under competitive conditions, new health risks will thus be covered only with a certain delay (see Table 3.1).

This is even more true of community-based schemes, which still have to deal with communicable diseases causing individual illness probabilities to be positively correlated. Extending the benefit package therefore may increase the risk of ruin, especially since these schemes operate in areas where close personal contact is very common (Nugroho et al., 2001).

A public insurer is called upon to cover emerging new risks because public health is at stake. While hardly concerned by the risk of ruin, it still has to be taken into account that the government possibly must cover high deficits.

3.1.7 Regulation

Regulation typically concerns not only premiums but also products because premium regulation can be subverted by product differentiation. Premium regulation typically prevents insurers from differentiating premiums according to true risk. A given uniform premium is associated with a contribution to expected profit in the case of a low risk but cause of an expected deficit in the case of a high risk. Therefore, it becomes vital for an insurer to attract as many low risks as possible. One way to achieve this is to modify the benefit package, excluding services that attract high risks. More generally, insurers will use benefits to compete with differentiated products since price competition is hindered by the regulator. In all, premium regulation in principle serves to increase the variety of benefit packages in the market, unless product regulation neutralizes this tendency.
Overall, regulation of insurance can be efficiency-reducing, in particular if it seeks to minimize the social cost of insolvency by avoiding insolvency altogether (see Appendix B and Table B.1). Typically, efficiency-enhancing regulation limits itself to mitigating the social costs of insolvencies while permitting them in principle. An overview is provided by Appendix Table B.2.

In most community-based schemes, the premium is determined by members themselves. The resulting premium is uniform; however, this triggers but little risk selection effort through product differentiation (see Section 3.2) because the risk pool is very homogeneous. Moreover, most schemes are local monopolies; therefore, they have little incentive to compete for members with differentiated benefit packages. An example in point is the Mburahati Health Trust Fund in Tanzania (Musau, 1999), which only offers coverage for outpatient care and a small contribution toward public hospital care.

Since public health insurance can be seen as being subject to a maximum degree of regulation (see Appendix Table B.1), it is also most strongly exposed to it in the determination of the benefit package. Expanding benefits is in the logic of a government who seeks to provide a maximum amount of public goods; therefore, a strong tendency in this direction can be expected.

3.1.8 Fraud and abuse

Fraud and abuse may occur at three levels. First, it constitutes an extreme form of moral hazard on the part of the insured, which however may be countered by the insurer by inspections and curtailment or even denial of benefits. Second, providers of services may act fraudulently; here, the countermeasure is to pattern their remuneration in a way to give them an incentive for honesty (revelation principle, see e.g. Laffont and Tirole (1993), ch. 1). Third, fraud and abuse may occur when health care providers affect their purchase. Being one step remote, it cannot easily be neutralized by the insurer unless competition between providers is strong.

Providers of medical supplies may ex-ante fraud physicians and hospitals, e.g., by offering money payments in exchange for their more
3.2 Risk Selection Effort

expensive products being used for treatment rather than cheaper products from competing suppliers. These products tend to be also of lower quality and quantity since corrupted suppliers have to recover their bribery payments through their sales margins. This results in insurable medical services being of lower quality at a given price. An insurer considering the extension of its benefits package thus has to take into account that such an addition may well be of lower quality, thus failing to induce much WTP in terms of higher premiums. This makes more comprehensive benefit packages rather unattractive (see Table 3.1). For instance, some private health insurers in Thailand decided to terminate coverage for ambulatory care because auditing the bills and checking for fraud became too costly (Health Systems Research Institute, 2002, p. 7).

A public health insurer in principle is affected by corruption in the same way as a private one in that it can offer only fewer services or lower-quality services for the amount of payroll tax or general tax received. This means that the benefit package is not as comprehensive as it could be. However, the list of benefits cannot easily be purged of those items whose suppliers had bought their slots. This serves to attenuate the negative relationship between benefits and fraud, at least as long as incurring a deficit is an option.

Conclusion 3. The comprehensiveness of the benefit package constitutes a first dimension of the supply of health insurance. It depends on at least eight factors, with moral hazard exerting an important limiting influence.

3.2 Risk Selection Effort

Most policy makers and even many economists believe that “skimming the cream,” i.e., making an effort to attract the favorable risks, is typical of private health insurers. However, upon closer examination, this belief is not justified. If health insurers were entirely free to grade their premiums according to risk, they would not want to invest in risk selection at all, and for the following reason (see also Appendix A). An unfavorable risk of course would be charged a high premium, whereas a favorable
Table 3.2 Factors affecting risk selection effort.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private health insurance</th>
<th>Community-based health insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk aversion of insurer</td>
<td>+ (n.a.)</td>
<td>+ ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Moral hazard</td>
<td>+ (n.a.)</td>
<td>+ ↓</td>
<td>n.a.</td>
</tr>
<tr>
<td>Size of the benefit package</td>
<td>+ (n.a.)</td>
<td>+ ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Diversity of risks</td>
<td>+ (n.a.)</td>
<td>+ ↓</td>
<td>n.a.</td>
</tr>
<tr>
<td>Access to risk information</td>
<td>+ (n.a.)</td>
<td>+ ↓</td>
<td>n.a.</td>
</tr>
<tr>
<td>Sellers’ concentration</td>
<td>− (n.a.)</td>
<td>− ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Regulation</td>
<td>+ (n.a.)</td>
<td>+</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Note: ↑ Reinforcement of relationship, ↓ Attenuation of relationship, n.a.: not applicable.

risk would demand and obtain a low premium. Given expected future health care cost, insurers would adjust premiums such that the expected contribution margin is equalized across risk groups. Under the pressure of competition, they simply cannot cross-subsidize one risk group to the detriment of another because the discriminated group can generate a more favorable offer from a competing insurer. For this reason, “n.a.” is entered in Table 3.2 above where appropriate in order to reflect the fully competitive unregulated benchmark, indicating that the factor considered is not effective. However, in the following it is assumed that premiums are regulated at least to some extent, imposing more uniformity than warranted in view of actuarial considerations and inducing an interest in risk selection on the part of competitive health insurers. A theoretical model analyzing both risk-selection and product-innovation effort can be found in Appendix A.

3.2.1 Risk aversion of insurer

If premiums have to differ from the expected value of the loss covered plus loading (see Eq. (3.1) of Section 3.3), the underwriting result of the insurer has excessive variance. The predicted response of management to this increased risk exposure depends on the same considerations as expounded in Section 3.1.1. If management has leeway to pursue its own interests, inducing risk-averse behavior, then it will undertake risk selection efforts because it can decrease their own risk exposure in
this way (see Table 3.2). However, this tendency can be neutralized by implementing a more or less elaborate risk-adjustment scheme (see e.g. Van de Ven and Ellis, 2000).

Community-based schemes are subject to the interest in risk selection because their member-owners certainly are much less diversified than the typical shareholder of an insurance company, which makes them particularly concerned about exposure to a risk that may ultimately spell insolvency. For a public insurer who wields a monopoly, risk selection is not relevant to begin with, motivating the “n.a.” entries in Table 3.2.

### 3.2.2 Moral hazard

A competitive health insurer would want to charge a high premium to consumers who are particularly susceptible to moral hazard (see Eq. (3.4) of Section 3.3.6). If this is not possible due to premium regulation, risk selection clearly is a substitute measure because it can be used to keep the high-moral hazard types out of the insured population.

### 3.2.3 Size of benefit package

With a very limited benefit package, differences in expected contribution margins between high and low risks typically are not all that large. This means that the incentive to engage in risk selection is not very marked either (see the model in Appendix A). Conversely, the more comprehensive the benefit package, the more health insurers are predicted to invest in risk election effort. This tendency is likely especially strong among community-based schemes because once they begin to offer more benefits, their risk exposure increases, and this increase can be counterbalanced by a more careful selection of risks.

### 3.2.4 Diversity of risks

Above all, diversity of risks means that the insured differ widely in terms of their expected value of loss, i.e., their illness probability and/or the amount of medical care utilized in the event of illness. The larger such discrepancies, the more does premium regulation (in the
limit: uniformity of premiums) induce excess variance in the underwriting result. A private health insurer is predicted to counter this by stepping up its risk selection effort. However, the same behavior is predicted for a community-based scheme (or in fact any nonprofit insurer) as long as running into deficit triggers a sanction of some sort. In their case this tendency is weaker because traditionally their insured population has been very homogenous to begin with (see Table 3.2).

### 3.2.5 Access to risk information

Risk selection is an attempt on the part of the health insurer to at least partially overcome an asymmetry of information resulting from the likely fact that the person to be enrolled knows more about his or her future health risks than does the insurer. However, genetic information may change that. In fact, the availability of such information permits the insurer to predict future healthcare expenditure of an individual with much greater precision. Moreover, refusal to provide genetic information serves as a signal suggesting that the person disposes of genetic information indicating he or she constitutes a high risk. This means that the effectiveness of risk selection effort is greatly enhanced by improved access to risk information of this type. Accordingly, risk selection becomes a more attractive alternative for health insurers.

### 3.2.6 Sellers’ concentration

The importance of sellers’ concentration can be seen from the following thought experiment (Wilson, 1977). If there were only two companies (A and B) in the market, risk selection would not make much sense provided the planning horizon of the two competitors extend beyond the current period. True, in period 1 insurer A may be able to filter out the unfavorable risks. However, it in fact would dump these risks on B, who in turn would have to resort to risk selection to stave them off in period 2. Thus, in period 3 these unfavorable risks would be seeking coverage with insurer A again. In the end, both A and B would lose from investing in risk selection. This consideration makes risk selection in concentrated health insurance markets less likely. This probably applies to community-based health insurance to a lesser degree because
their members also own the scheme, fully exposing them to the risk of insolvency that may be the consequence of a failure to carefully gauge potential clients. Of course, these considerations do not apply to a public insurer, who wields a monopoly.

### 3.2.7 Regulation

As stated in the introductory paragraph of Section 3.2, a health insurer who has the freedom to grade its premiums according to risk will tend to equalize expected contribution margins across risks. Unfavorable risks, while expected to cause high healthcare expenditure, also pay a high premium, whereas favorable risks must be attracted by low premiums that reflect their low future cost. Arguably, it is premium regulation, seeking to relieve the high risks from “excessive” premiums, that induces risk selection by health insurers (Pauly, 1984). A means-tested subsidy paid out to potential purchasers of health insurance with low incomes could provide an alternative. In this way, this counterproductive side effect of premium regulation (which is to be expected regardless of for-profit status) can be avoided (see Zweifel and Breyer (2005) for details).

**Conclusion 4.** The amount of risk selection effort is a second dimension of supply of health insurance. Induced by premium regulation, it is of great concern to policy makers. Its extent depends on at least seven factors, the only mitigating one being a high sellers’ concentration among insurers.

### 3.3 Loading: The True Price of Insurance

Since premiums are in part paid back to consumers in the guise of benefits, they do not reflect the price of insurance. Rather, the true price of insurance is that part of the premium that is not used to pay benefits, the so-called loading for administrative expense and profit. In more formal terms, insurers pay an indemnity $I$ to cover a loss against a premium. The gross premium can be divided in a net premium ($\pi \cdot I$), with probability of loss $\pi$ depending negatively on preventive effort on the one hand and a loading on the other. The net premium covers
the expected amount of benefit to be paid. The loading can be further subdivided in a component that is a per-unit amount \( \mu \) associated with claims processing. The higher the likelihood of a claim being presented, the more often an administrative process is triggered. The other component is a multiple \( \lambda \) of expected benefits net of copayment (symbolized by a rate of coinsurance \( c \) for simplicity), reflecting acquisition cost, a risk premium, and profit. Therefore, a viable insurance contract must be priced to contain the following elements (Zweifel and Breyer (1997), ch. 6.2)

\[
P(I) = \text{net premium} + \text{loading} \\
= \pi(V) \cdot (1 - c) \cdot I + \mu \cdot \pi(V) + \lambda \cdot \pi(V) \cdot (1 - c) \cdot I,
\]

(3.1)

where

- \( P \): Premium,
- \( \mu \): Loading factor for variable administrative expense,
- \( \pi \): Loss probability, probability of illness; \( 0 < \pi < 1, \pi'(V) < 0 \),
- \( V \): Preventive effort (unobservable),
- \( c \): Rate of coinsurance; \( c < 1 \),
- \( \lambda \): Loading factor for acquisition cost, risk premium, and profit,
- \( I \): Benefit paid in the event of illness.

This equation needs to be completed by the following consideration. The more complete coverage, denoted by \( I \), the weaker in general are insured’s incentives for prevention \( V \).\(^1\) Taking into account this ex-ante moral hazard effect, the amount of loading can be written as

\[
\text{Amount of loading} = \mu \cdot \pi(V(I)) + \lambda \cdot (1 - c) \cdot \pi(V(I)) \cdot I
\]

(3.2)

The question arises immediately whether the concept of loading has any relevance to a public health insurer. It does, and for two rather different reasons. First, a public scheme also has its administrative expense, which rises as the frequency of claims \( \pi \) increases. This frequency depends on preventive effort \( V \) precisely as with any private insurer, and \( V \) in turn is again negatively related to coverage

\(^1\)Under certain circumstances the incentives for prevention are higher when coverage increases. \( V \) responds positively to an increase in \( I \) when the insured earns a high wage, is risk averse and/or enjoys a generous sick leave. This can be the case in developed countries (Zweifel and Manning (2000, p. 417))
Table 3.3 Factors affecting the net price of health insurance (loading).

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private insurance</th>
<th>Community-based health insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative expenses</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(including capital charge)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinsurance</td>
<td>+/-</td>
<td>+/- ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Pool size</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
</tr>
<tr>
<td>Benefit package</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Copayments and caps</td>
<td>-</td>
<td>-</td>
<td>- ↓</td>
</tr>
<tr>
<td>Moral hazard</td>
<td>+</td>
<td>+</td>
<td>+ ↑</td>
</tr>
<tr>
<td>Quality and proximity of health care services</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Regulatory framework</td>
<td>+/-</td>
<td>+/- ↓</td>
<td>+/- ↑</td>
</tr>
<tr>
<td>Fraud and abuse</td>
<td>+</td>
<td>+</td>
<td>+ ↑</td>
</tr>
</tbody>
</table>

Note: ↑ Reinforcement of relationship, ↓ Attenuation of relationship, n.a. not applicable.

\[ I \] (the ex-ante moral hazard effect). The term \( \mu \cdot \pi(V(I)) \) of Eq. (3.2) therefore applies to public insurance as well. Second, although a public insurer need not charge for acquisition cost, risk bearing, and profit, it gives rise to a “loading” that is very similar to the second term of Eq. (3.2). The larger the expected value of benefits to be paid net of coinsurance \( ((1 - c) \cdot \pi \cdot I) \), the higher must be the rate of tax levied on labor income or on sales. Now as is well known, taxes cause inefficiencies because they reduce the volume of transactions; some contracts that would have been mutually beneficial are not struck under the influence of tax. These inefficiencies easily amount to 20 percent of transaction value (see e.g. McMaster, 2001) and thus are of a comparable magnitude as \( \lambda \) in Eq. (3.2).

In all, the expression for the loading given by Eq. (3.2) can be applied to public health insurance as well, at least to a first approximation. The “loading” may of course differ depending on the type of taxation used to fund the scheme. The amount of loading is influenced by several factors listed in Table 3.3.

### 3.3.1 Administrative expense

Administrative expense must be recovered before the insurer reaches the break even. They are added to the expected loss. The loading
factor $\mu$ reflects these expenses and thus importantly determines the amount of loading (see Eq. (3.2)). It depends on possible economies of scale, implying that a critical number of contracts and transactions may be necessary to reach minimum average cost. The loading factors also include capital utilization costs, and surcharges for uncertainty about future cost inflation in the healthcare sector and about the loss probability $\pi$.

Community-based schemes are known for their low administrative expenses because they do not employ much staff. Moreover, the staff they do have is voluntary in the main. This serves to keep the loading factors at a low value. In fact, members bear part of the costs of organization by spending time and effort to decide on the product to be offered and premium to be charged.

Public health insurance constitutes a monopoly, which means that marketing and advertising expenses are reduced. On the other hand, a monopoly reduces the pressure to minimize cost. On the whole, the relationship may be comparable to that in private competitive health insurance.

### 3.3.2 Reinsurance

Generally, reinsurance is an expense that reduces the expected value of profit (if the premium exceeds the actuarial value of losses ceded; see Doherty and Tinic (1981)). It is therefore similar to administrative expense, causing the loading to increase, ceteris paribus. The benefit of reinsurance is that it improves the solvency of the insurer, permitting a lower value of the loading factor $\lambda$. Still, if additional capital is available at lower cost than reinsurance, it is preferable for an insurer to rely on the capital market rather than taking out reinsurance.

Reinsurance can be beneficial to community-based health insurers, whose pool size usually is insufficient for the law of large numbers to come to full effect. According to the law of large numbers, insurers are able to estimate $\pi$ and hence the expected value of benefits to be paid more precisely when the number of risks increases. Ceteris paribus, this facilitates the attainment of a given level of solvency. In addition, the typically nondiversified individual (member) owners of such schemes will gain from the lower variance of the surplus (assets minus
3.3. Loading: The True Price of Insurance

Liabilities) generally afforded by reinsurance. This benefit in terms of variance reduction, however, has to be weighed against the reinsurance premium. Therefore, low-cost reinsurance may become a precondition for the viability of community-based health insurance, which most often do not have access to capital markets.

Reinsurance will hardly be an issue for a public health insurer. Its large risk pool allows to minimize per-capita reserves (see Section 3.3.3), to which reinsurance contributes. In addition, these reserves are usually provided by the government as a lender of last resort; ultimately, the taxpayers act as the reinsurers of the public health insurer. Compared to a private insurer, these savings on reinsurance entail a cost advantage of the public monopolist (see Table 3.3).

3.3.3 Pool size

A large number of insured of similar type allows to estimate the unknown parameters $\pi$ and $I$ with greater precision, due to the law of large numbers. Therefore, the insurer does not have to carry as much reserves per unit risk for attaining a given level of solvency (Dror and Preker, 2002, p. 135). The pertinent loading factor $\lambda$ becomes smaller, resulting in a smaller total loading.

On the other hand, a large pool size shields the individual insurance buyer from social control through other members. This control likely refers to the benefits claimed ($I$) rather than preventive behavior and hence $\pi$. Increased pool size thus strengthens ex-post moral hazard but leaves ex-ante moral hazard unaffected. The second term of Eq. (3.2) increases, indicating that the amount of loading increases.

In the case of a community-based health insurer, the trade-off between the two influences can be studied. For instance, the Dana Sehat schemes in Indonesia are organized in several thousand independent groups, with approximately 50 to 100 families in each group. Families are homogenous with regard to household size and income, and due to the community environment, behavior is closely monitored. Although the total number of Dana Sehat participants is very large (7 million people in Indonesia), moral hazard can be controlled effectively, resulting in a small loading in spite of small pool size. Taiwan’s “Farmer’s
health insurance” provides a counter example. There, a risk pool typically comprises a few thousand individuals (Bureau of National Health Insurance, 2003). This could lead to a lower value of $\lambda$; however, greater pool size also calls for more complex management, and social control is undermined. Although information about the total loading is not available, it is likely to be higher than in Indonesia.

Public health insurance schemes have too large risk pool sizes for moral hazard effects to be mitigated by social control anymore. Therefore, expanding the pool even more causes the loading contained in the contribution to unambiguously decrease due to the law of large numbers.

3.3.4 Benefit package

An extension of the benefit package increases the likelihood of some claim being submitted. Therefore, the probability of loss $\pi$ increases even without any behavioral modification on the part of the insured (moral hazard effects are dealt within Section 3.3.6). Likewise, payment may occur under additional titles, resulting in an increased value of payments $I$. Therefore, the amount of loading must increase according to Eq. (3.2).

This argument holds also for community-based and public health insurance (see Table 3.3).

3.3.5 Copayments and caps

Copayments and caps have three effects on total loading. First, they serve to limit ex-post moral hazard. Copayments increase the net price of medical care to consumers, causing them to lower the quantity demanded, while caps increase the net price to its full market value when the threshold quantity is exceeded. Therefore, the value of payments $I$ decreases on average and with it the amount of loading. Caps have the additional feature of excluding very high values of $I$, thus reducing also the (semi) variance of $I$ and hence the loading factor $\lambda$.

Second, copayments relieve the insurer of part of the payment in the advent of illness. As shown in Eq. (3.5) in Section 3.3.6, an increase in the rate of coinsurance $c$ serves to lower the total amount of loading.
Copayments and caps thus unambiguously serve to reduce the amount of loading.

The same arguments hold for community-based schemes. They have even greater force for public health insurance, where the initial rate of copayment is zero, resulting in maximum ex-post moral hazard effects. Indeed, according to Eq. (3.5), the amount of loading reacts most strongly to a variation in the rate of coinsurance $c$ when $(1 - c) = 1$, i.e., when $c = 0$ initially.

### 3.3.6 Moral hazard

Moral hazard increases the consumption of health care services by the insured and thus causes additional costs to the insurer compared to the situation without insurance. It is a common phenomenon in the insurance and healthcare industry. It is convenient to distinguish between ex-ante and ex-post moral hazard. Ex-ante moral hazard refers to the probability of illness $\pi$. This probability depends on related preventive effort on the part of the insured, denoted by $V$.

While preventive effort can hardly be observed in the context of health behavior, it generally decreases when the amount of coverage offered is extended. Ex-ante moral hazard thus results in a positive relationship between $\pi$ and the amount of insurance coverage $I$.

Indeed, because of ex-ante moral hazard an increase in $I$ not only is associated with a higher gross premium, but a higher amount of total loading as well. For convenience, Eq. (3.2) is repeated here:

\[
\text{Amount of loading } = \Lambda = \mu \cdot \pi(V(I)) + \lambda \cdot (1 - c) \cdot \pi(V(I)) \cdot I. 
\]  
(3.3)

The derivative of this expression with respect to $I$ (neglecting possible effects of $I$ on the loading factors $\mu$ and $\lambda$) is given by

\[
\Lambda'(I) = \mu \cdot \pi'(V) \cdot V'(I) + (1 - c) \cdot \lambda \cdot \pi'(V) \cdot V'(I) \cdot I \\
\quad -\mu \cdot \pi'(V) \cdot V'(I) \\
\quad +\lambda \cdot (1 - c) \cdot \pi(V(I)) > 0. 
\]  
(3.4)
With \( \pi' \) and \( V' \) negative, the first term is positive. For the same reason, the second term is positive as well, and the third term is positive by definition. In analogy to the development in Zweifel and Breyer (1997, p. 183), the loading usually increases progressively in \( I \), if \( \pi''(V) > 0 \) (prevention becoming less effective at the margin) in addition to \( V'(I) < 0 \).

According to Eq. (3.4), some health insurance benefits may be more affected by ex-ante moral hazard than others because preventive effort \( V \) responds more strongly to an increase in \( I \). Conversely, this effect may be mitigated to some extent if health insurance is provided through the employer, who can at least monitor prevention at the workplace. This difference would be reflected in a more moderate increase of the loading (as well as the gross premium) when coverage becomes more complete or more comprehensive.

Summing up, ex-ante moral hazard likely causes an increase in the total loading, which may be even progressive in benefits \( I \). There do not seem to be strong reasons to modify this argument for community-based schemes. With regard to public health insurance, the government’s objective of maximizing the provision of public and merits goods (see Section 3.1.4) frequently militates against imposing a copayment. However, this implies that any increase in benefits must go along with a maximum increase in the loading because of ex-ante moral hazard. In Eq. (3.4), the amount of loading reacts most strongly to an increase in benefits if \((1 - c) = 1\), i.e., when \( c = 0 \).

Turning to ex-post moral hazard, this means the tendency of the insured to demand more medical care (or care of higher quality or provided by a more expensive provider) after the onset of an illness. The effect of ex-post moral hazard was illustrated in Figure 3.2 of Section 3.1.3; there, the role of coinsurance played a crucial role. It remains to be shown that a decrease in copayment also increases the amount of loading.

For this, a slightly different interpretation of the variable \( I \) is needed. Now \( I \) becomes the amount of benefits that is actually claimed (rather than promised in the contract), which depends on the rate of coinsurance. Therefore, \( I \) has to be replaced by \( I(c) \) in Eq. (3.3) above, resulting in the derivative (note that now preventive effort \( V \) is
3.3. Loading: The True Price of Insurance

3.3.1 Quality and proximity of healthcare services

Healthcare services of high quality have a direct effect on the total loading because the benefits actually claimed typically are more expensive (see the effect of a high value of $I$ in Eq. (3.3)). High quality of services may also aggravate ex-post moral hazard effects. This can be illustrated using Figure 3.2 of Section 3.1.3 again. Maximum true WTP for such services must be very high, causing the observed demand function to run steeply. In this case, ample insurance coverage (low $c$) results in a marked discrepancy between true and observed WTP. Graphically, the distance between quantities A and B becomes larger. In terms of Eq. (3.5) above, a decrease of the rate of coinsurance $c$ would cause benefits claimed to increase very strongly. With $I'(c)$ large – equivalent to a steep demand function – the loading must increase more strongly...
with a decrease in $c$. Therefore, the loading depends positively on the quality of medical services in general.

Increasing proximity of services causes the cost of access and hence total cost of utilizing medical care to fall. Therefore, the amount of services claimed $I$ increases, and with it the amount of loading also (see Eq. (3.5)).

Most members of community-based schemes are located far away from high-quality healthcare service providers. Any increase in the proximity of a health care provider therefore is likely to have a considerable effect on the cost of access, inducing a particularly marked increase in utilization. However, these schemes benefit from a degree of mutual monitoring of their members that does not prevail in the context of a private insurer. Therefore, the amount of loading may not respond more strongly to an increase in proximity than in industrial countries.

Increased quality and proximity also drive up the loading component in contributions to public health insurance; Eq. (3.5) applies once more (see Table 3.3).

### 3.3.8 Regulatory framework

The types of regulation of relevance in this context are again premium and product regulation. If designed to guarantee solvency, premium regulation typically amounts to an increased safety loading, which is reflected in $\lambda$. Conversely, if regulation is consumer-orientated, it may result in increased transparency for consumers, enhancing demand and resulting in a larger risk pool. This means that the reserves held per unit risk can be reduced, causing $\lambda$ to be smaller. Turning to product regulation, this implies that certain procedures in loss settlement have to be followed, presumably at an increased cost to the insurer. This drives up the value of the other loading factor, $\mu$. Therefore, the overall effect of regulation on the loading is ambiguous, although in the case of U.S. automobile regulation, Frech III and Samprone Jr. (1980) found that regulation had a demand-decreasing net effect, pointing to a positive relationship between regulation and loading.

In community-based schemes, insurance packages and the premium rate are strictly regulated by the members themselves. This regulation
does not aim at creating reserves through a loading surcharge on the
risk premium; rather, the insured must come up with additional con-
tributions (often in kind) in the event that the scheme runs a deficit.
The downside of the reduced loading is an increase in the residual
asset variance for members; however, risky insurance is associated with
reduced WTP.

Public health insurance is usually governed by an elaborate regula-
tory framework (in Section 5.2, the view will be expounded that public
insurance is at the high end of a scale, depicting increasing regulatory
intensity). This adds to administrative expense and hence the “load-
ing”; the total amount of loading may still be low due to savings on
the cost of acquisition.

3.3.9 Fraud and abuse

Fraud and abuse are closely related to the institutional framework.
In Section 3.1.8, emphasis was on the corruption possibly occurring
between suppliers of medical inputs and physicians and hospitals. At
this juncture, fraud and abuse by the insured are taken up and their
impact on the loading discussed.

Fraud and abuse are an extreme form of moral hazard. In the case
of ex-ante moral hazard, preventive effort $V$ could be said to turn nega-
tive, implying that insured’s behavior increases the probability of illness
to 1. A negative value of $V$ may well be induced by insurance; in terms
of Eq. (3.4) of Section 3.3.6, $V'(I)$ is strongly negative as well. This
means that the amount of loading must increase very rapidly with any
increase in $I$.

Fraud can also occur ex-post, e.g., in the guise of conjuring with
providers to overstate medical bills. Again, this is an extreme form of
ex-post moral hazard that is encouraged by a vanishing rate of coin-
surance (or more generally, the absence of cost sharing). For as soon
as the insured have to pay parts of the medical bill out of pocket, they
have an incentive to resist fraudulent overbilling. In general terms, the
relationship between the degree of cost sharing $c$ and benefits claimed
$I$ is strong in the presence of fraud. For the insurer, the term $I'(c)$ in
Eq. (3.5) takes on a very large value (in absolute terms), indicating that
the total amount of loading must increase strongly with a decrease in cost sharing when fraud is prevalent.

As argued in Section 3.1.3, moral hazard in community-based schemes, and as such also any extreme form of it, is mitigated due to the characteristics that prevail in rural communities (close to full information). Therefore the amount of loading due to fraud and abuse should not increase much in this variant of health insurance.

A public health insurer is under reduced pressure to control fraud and abuse; contrary to private insurers, it does not have to compete for customers through a favorable benefit-cost ratio (to which a low amount of loading contributes).

Conclusion 5. The third dimension of supply of health insurance is the loading contained in the premium, which constitutes the net price of health insurance. It depends on at least nine factors, with copayments and caps an important mitigating one.

3.4 Vertical Restraints/Vertical Integration

Two forms of vertical restraints (in the extreme: full vertical integration) can be distinguished, insurer-driven and provider-driven. A third form of integration is not vertical but lateral. It occurs when a firm with main activities outside the sector takes up business in health insurance or the provision of health care. It will be dealt with only in passing.

3.4.1 Insurer-driven vertical integration

A private insurer can limit its activities to the refunding of medical expenditures incurred. This amounts to a total absence of vertical restraints, let alone vertical integration. Such a policy is costly to the insurer, however, if the providers of medical care have some monopolistic power. In that event, insurance coverage drives up providers’ markup over marginal cost. This can be illustrated by Figure 3.3, which builds on Figure 3.2 of Section 3.1.3.

The added feature of Figure 3.3 is two marginal revenue functions ($MR$). Without insurance coverage, the health care provider faces the $MR$ function derived from the true demand function ($MR_1$). The
quantity satisfying the optimality condition, “marginal revenue equals marginal cost” (of healthcare services) is $A$. Accordingly, the monopoly price is $P^*$, which already contains a markup over marginal cost. With insurance, the $MR$ function becomes $MR_0$, associated with the observed demand function. The new optimal quantity of services provided is $B$, consistent with a higher monopoly price at $P^{**}$, reflecting an increased markup over marginal cost. In this situation, the moral hazard effect of insurance not only consists of an increased quantity consumed ($B > A$), but also higher prices ($P^{**} > P^*$). Since this boosts payments $I$, the amount of loading and hence the price of insurance increases as well (see Eq. (3.4) of Section 3.3.6). One rationale of insurer-driven vertical integration is to avoid this extra moral hazard effect, given by $(P^{**} - P^*)$. For example, the insurer might employ the health care provider, with the employment contract stipulating fees as low as marginal cost (of course, wage income paid would have to contain a fixed component to make up for the associated loss of revenue on the provider’s part).

Fig. 3.3 Effect of insurance coverage on monopolistic pricing.
In more general terms, the provision of health insurance and of healthcare services may be viewed as two parts of a system. The extra moral hazard effect then amounts to an externality within the system that the insurer may seek to mitigate by imposing vertical constraints on service providers. To be successful, it must itself have a degree of monopoly power (see Section 3.4.1.1). Therefore, the objective of the insurer becomes to avoid a double monopoly markup, or double marginalization (Waldman and Jensen, 2001, p. 468f). The solution can be a two-part remuneration scheme. First, the provider agrees to charge a price equal to marginal cost, and second the insurer pays a fixed amount sufficient to motivate the provider to sign the contract. In the extreme case, the insurer can opt for fully integrating service providers to avoid this and other externalities. The different possibilities form a continuum between independent provision and full vertical integration (see Figure 3.4).

For example, when full integration would be inefficient, the insurer may limit itself to ownership of hospitals while contracting with ambulatory care providers. It also can mix insurer-managed plans with plans that are governed by contractual relationships devoid of vertical restraints. The imposition of restraints can finally be delegated to, e.g.,

![Diagram of vertical integration/restraints](image)

Fig. 3.4 Forms of vertical restraints and integration imposed by the insurer.
### 3.4. Vertical Restraints/Vertical Integration

Table 3.4 Factors affecting insurer-driven vertical integration.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private health insurance</th>
<th>Community-based health insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market power of the insurer</td>
<td>+</td>
<td>+ †</td>
<td>+</td>
</tr>
<tr>
<td>System efficiency gains to be realized</td>
<td>+</td>
<td>+</td>
<td>+ †</td>
</tr>
<tr>
<td>Management know-how of insurer</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Contestability of healthcare markets</td>
<td>+</td>
<td>+ †</td>
<td>+</td>
</tr>
<tr>
<td>Potential to increase entry barriers to competitors</td>
<td>+</td>
<td>+ n.a.</td>
<td></td>
</tr>
<tr>
<td>Contestability of health insurance market</td>
<td>−</td>
<td>− †</td>
<td>n.a.</td>
</tr>
<tr>
<td>Lack of capital of insurer</td>
<td>−</td>
<td>− †</td>
<td>− †</td>
</tr>
<tr>
<td>Opportunistic behavior and fraud on the part of insurers</td>
<td>−</td>
<td>− †</td>
<td>− †</td>
</tr>
<tr>
<td>Cartelization of service providers</td>
<td>−</td>
<td>− †</td>
<td>− †</td>
</tr>
<tr>
<td>Legislation prohibiting vertical restraints</td>
<td>−</td>
<td>− †</td>
<td>−</td>
</tr>
</tbody>
</table>

Note: † Reinforcement of relationship, † Attenuation of relationship.

A medical association, with the likely result that individual provider behavior is not very effectively restrained.

The factors encouraging and hampering vertical integration by the insurer are listed in Table 3.4.

#### 3.4.1.1 Market power of the insurer

This amounts to a necessary condition for the imposition of vertical restraints. If one of many insurers were to impose vertical restraints, a given service provider would always have the opportunity to strike a contract with a competitor that does not seek to impose such constraints. Moreover, as long as these constraints do not amount to exclusive dealings, failure to sign up with a particular insurer has negligible consequences for a service provider. Therefore unless the insurer considered wields a degree of market power, service providers do not need to accept any vertical restraints.

Market power of community-based health insurers typically is high because they as a rule wield a monopoly in the rural area they serve. On
this score, their degree of market power would certainly enable them ceteris paribus to impose vertical restraints.

A public health insurer, being a monopolist, can impose strong vertical restrictions on providers in terms of prices and products delivered if not prevented by legislation. There is a risk of abusing market power; in particular, purchasing prices may be set so low as to drive foreign suppliers of drugs and privately funded hospitals out of the market. This risk is higher under a public insurance scheme than under a competitive private insurance system. Grant and Grant (2002), citing an unpublished text, refer to the example of a sub-Saharan African country where payments by national health insurance are so low that service suppliers have to heavily rely on unofficial charges for finance. The authors also use data from Transparency International (several years), which shows that up to 80 percent of recent transactions with health workers in certain countries involves an unofficial fee or bribe.

3.4.1.2 System efficiency gains to be realized

The double marginalization problem noted above is not the only within-system externality that can be mitigated by vertical restraints. One that is also discussed in the industrial organization literature (Carlton and Perloff, 1999, ch. 12) is the risk of the distributor delivering sub-standard quality, with adverse reputation effect on the producer. In the present context, this translates into physicians and hospitals skimping on quality in the treatment of patients enrolled with a particular insurer. The solution to this problem can be the creation of a quality assurance scheme by the insurer.

Another problem that is more peculiar to the healthcare sector is fraud. As emphasized by Ma and McGuire (1997), the insurer has to rely on a report provided by the physician to be able to establish the appropriateness of treatment. The typical vertical restraint used here is a clause to the effect that service providers are to offer additional information in case of ambiguity.

A third within-system externality, of particular relevance to health care, is the “medical technology race.” Given that insurance coverage is complete and density of supply high, service providers cannot
compete much by price and location. An important remaining parameter of competition is medical technology. However, for the insurer it suffices to have few specialized providers offering the most advanced technology for diagnosis and treatment of a given health condition. This implies that a technology race among the providers who are contractual partners amounts to a source of inefficiency. To avoid it, the insurer may assign providers to certain health conditions, at the same time guaranteeing them a minimum number of cases per period. Such a commitment can be supported by a premium reduction offered to enrollees in return for a restricted choice of provider, as is often the case with managed care contracts.

Community-based schemes also face a double marginalization problem. In the rural areas where they operate, an individual physician or hospital may be a local monopolist. The fact that they contract with nonprofit institutions is of limited relevance as soon as these providers must recover their cost. Quite likely any patients treated free of charge or at a reduced fee are those without any insurance coverage at all. The deficit incurred must be neutralized by higher fees from those who do have insurance protection, viz. members of community-based schemes. Provision of substandard quality therefore can be an issue since these providers are also monopsonists in their local labor markets. This induces them to pay a comparatively low wage, making them unlikely to attract the most skilled healthcare workers. With regard to fraud, community-based health insurers may benefit from the nonprofit status of especially hospitals; however, public hospitals have a tradition of cheating to ease bureaucratic processes. The technological race between competing providers can be excluded from consideration since community-based insurers are localized primarily in rural areas of low-income countries, where local monopolies prevail.

Another source of efficiency gain is mode of payment. In many rural areas of low-income countries, service providers are still paid in kind. However, service providers generally prefer to receive cash. This has led some schemes to use so-called moneylenders as intermediaries who transform in-kind contributions into cash, to be paid to providers. In return, hospitals in particular have been willing to accept prospective payment for treating scheme members, which constitutes a vertical restraint.
A public health insurer, being protected by a monopoly, is under reduced pressure to reap any system efficiency gains through vertical restraints. Therefore, this particular motivation is seen as being of reduced importance compared to competing private insurers (see Table 3.4).

### 3.4.1.3 Management know-how of insurer

Of course, ample management know-how helps to successfully negotiate and monitor vertical restraints. This is especially true of full vertical integration, which presupposes knowledge on the part of the insurer on how to efficiently run provider facilities.

Management know-how is very scarce in community-based schemes, making vertical restraints less likely than conventional, often not fully specified contracts with service providers. For public health insurance, it may be roughly comparable to that of private health insurers operating in the same country.

### 3.4.1.4 Contestability of healthcare markets

Contestable markets are characterized by an actual or potential influx of suppliers, with the potential influx becoming effective as soon as incentives to enter become strong enough. As the experience of Managed Care Organizations in the United States suggests, newcomers to the market for medical services (e.g., young physicians) are more likely to participate, i.e., to accept the corresponding vertical restraints.

Having their centers of activity in rural areas, community-based schemes cannot count much on the contestability of the healthcare markets they deal with. If at all, service providers move from the countryside to the cities. Therefore, chances for these schemes to find partners who accept vertical constraints are rather slim.

To a public health insurer, increased contestability of healthcare markets certainly facilitates vertical restraints. However, public administrators still have to seek out those alternate providers that are available; their incentive to undertake this effort may be undermined by the monopoly status of the scheme.
3.4. Vertical Restraints/Vertical Integration

3.4.1.5 Potential to increase entry barriers to competitors

One motivation for vertical restraints and integration can also be to foreclose the insurance market to potential entrants.\(^2\) For, a new health insurer has to establish contractual relationships with providers to build a delivery system. By tying up scarce supply of healthcare services, incumbent insurers thus can indirectly preclude the entry of new competitors. Given the complexity of healthcare services and its high human capital content, controlling a part of healthcare supply can become a more effective barrier than closing the insurance market itself. On the other hand, vertical restraints can be disrupted by an outsider willing to offer a compensation high enough to make the healthcare supplier leave the vertical arrangement. However, such payment tends to be above the level a newcomer is willing to pay ("natural asymmetry," as Carlton and Perloff (1999, p. 357) put it).

Community-based schemes benefit from a different type of barrier to entry, which obviates the use of vertical integration to protect their market from outside competition. This follows from a likely analogy to credit markets. There, most community schemes are set up along kinship lines, at least in rural areas. In the case of Nigeria, more than 95 percent of borrowing and lending occurs within a given community scheme that usually coincides with a tribe. This suggests that a challenge to an incumbent community-based scheme would have to surmount a high barrier in the guise of kin relationships.

To a public health insurance scheme, the potential of vertical integration to reinforce entry barriers confronting competitors has no relevance because entry by a competitor is prohibited by law (see Table 3.4).

3.4.1.6 Contestability of health insurance markets

When insurance markets are and remain contestable, incumbent insurers will be strapped for resources in defending their position, being mainly absorbed with assuring their survival in the insurance market itself. In addition, when insurers have to compete because entry

\(^2\) For a discussion of the issues in the case of health care, see Preker et al. (2000).
or exit barriers are low, profitability is driven down to the competitive return; funds and management time will be too scarce to engage in the imposition of vertical restraints or even full vertical integration.

With regard to community-based health insurance, barriers to entry emanate mainly from the characteristics of informal markets. Many health insurers who might consider entry do not accept in-kind payment of the premium. This payment may not only take the form of cattle but also the provision of bonded labor and the cession of land rights. Thus, there are no barriers to entry that hamper the imposition of vertical restraints by community-based health insurance, ceteris paribus.

In the case of a public health insurer, the contestability of the market for health insurance has again no relevance since the law makes that market not contestable to begin with.

3.4.1.7 Lack of capital of insurer

This is another impediment especially to integration. Often, full vertical integration (but less so vertical restraints) requires a capital investment on the part of the firm acquiring control. If internal finance is available, management enjoys some leeway in deciding about such an investment, monitoring by the owners of the firm being incomplete. Lacking internal finance, the integrating firm has to convince banks and investors that vertical integration will improve profitability and that the debt can be repaid.

Community-based schemes are organized as mutuals and thus do not sell tradable shares of ownership. This precludes external equity finance, except through increasing membership. However, this alternative frequently runs into problems because the scheme may lose its homogeneity and hence an important cost advantage, as argued in Section 3.3.3. Finance through, e.g., banks is also difficult because the scheme cannot offer marketable collateral. However, in some cases lateral integration may help. Citing the experience of communities in Bangladesh, Desmet et al. (1999) argue that community-based credit schemes, which many individuals are already involved in, may provide the entry point to finance health insurance. But on the whole,
3.4. Vertical Restraints/Vertical Integration

lack of capital constitutes an important impediment to integration for community-based health insurers.

Turning to public health insurance, lack of capital hampers vertical integration as well because the scheme is not permitted to accumulate funds or issue debt for such purpose. Initiatives of this type would be interpreted as a sign of for-profit orientation.

3.4.1.8 Opportunistic behavior and fraud on the part of insurers

Insurers with a reputation of opportunistic and fraudulent behavior have difficulty striking contracts calling for vertical restraints. By engaging in opportunistic behavior, insurers inflict damages on providers, albeit at the expense of their own reputation and credibility. This reduces their chances of successfully arranging vertical restraints with providers. Insurers must therefore first establish their credit and payment reputation among providers in order to win them over for vertical restraints.

Fraud seems to be a minor issue in community-based schemes because service providers wield a local monopoly in many cases. If found cheating, the insurer therefore stands to lose the one available provider in the region. Since this constitutes an effective sanctioning mechanism, it should be easier to agree on vertical restraints (see Table 3.4).

Opportunistic behavior and fraud can also occur with a public insurer, undermining the willingness of service providers to enter vertical agreements. However, this effect is attenuated by the consideration on the part of providers that they have no choice but to sign up if they want to profit from the demand-enhancing effect of insurance coverage (see Section 3.3.6).

3.4.1.9 Cartelization of service providers

On the provider side, cartelization makes the imposition of vertical constraints more difficult. First, the cartel is a means for providers to jointly increase their incomes. An insurer seeking to negotiate a vertical restraint must beat this benchmark. Second, a cartel must impose
discipline on its members to be successful, notably with respect to restricting output. Restrictions on output however conflict with the integrating firm’s desire to avoid double marginalization, which may result in the imposition of a minimum volume of sales. In the present context, a medical association would like to see its members keeping to a low volume of treatments to support higher fees. However, an insurer may want to contract for a minimum volume of services at a fixed fee to avoid upward pressure on fees induced by insurance coverage (see Figure 3.3 of Section 3.4.1). These intentions are in conflict.

To community-based schemes, cartelization of healthcare providers has little relevance. In rural areas of low-income countries, providers are sufficiently protected from competition through mere distance. They can therefore do without the protection afforded by a cartel.

For a public health insurance scheme, cartelization of providers constitutes an obstacle to vertical restraints and integration in much the same way as for a private insurer. However, since the cartel has no one else to contract with, it may agree to a uniform set of vertical agreements to secure the viability of the system (and its demand-enhancing effect) as a whole.

\[\text{3.4.1.10 Legislation prohibiting vertical restraints}\]

Restrants can be entirely impossible when there is legislation prohibiting vertical restraints and integration in the healthcare sector altogether. For example, medical practices and/or hospitals must not be owned by individuals without a medical degree in several industrial countries. At the very least, medical management must lie in the hands of physicians.

For a community-based insurer, there seem to be few legal impediments to vertical integration. In fact, they were able to closely cooperate with missionary hospitals in several countries such as Uganda, Kenya, and Indonesia.

A public health insurer must presumably respect legislation concerning vertical integration in the same way as a private insurer does since the objective of this legislation is to secure the independence of the comparatively small businesses of healthcare providers.
Conclusion 6. The amount of vertical integration is the fourth dimension of the supply of health insurance. Its insurer-driven variant depends on at least ten factors, with the avoidance of double marginalization effects being an important motive.

3.4.2 Provider-driven vertical integration

The second type of vertical integration is provider-driven. The typical case would be a hospital chain that seeks to avoid double marginalization in its dealings with insurers that wield a degree of market power. The chain may also view an insurer as a sales channel, where promotional effort is decisive for the market success of its products. If insurers provide an insufficient amount of advice to future patients, client matching suffers, with unfavorable effects on hospitals’ reputation. A competing insurer could free-ride on these efforts by letting the other make promotional effort while selling its own policy at a lower premium. Such free riding would of course undermine insurer’s incentive to provide advice. The solution to the problem can be the assignment of exclusive territories to insurers or even exclusive dealings (Carlton and Perloff, 2000, pp. 403–405).

In general, the factors encouraging provider-driven vertical restraints and integration (listed in Table 3.5) are the same ones hampering their insurer-driven counterparts (listed in Table 3.4 in Section 3.4.1). With regard to public health insurance, however, provider-driven vertical integration is regarded as not applicable throughout (resulting in the “n.a.” entries in the last column of Table 3.5). The reason is that a hospital or a group of physicians will find it impossible to impose rules on a public agency, e.g., with regard to the amount of contribution to be paid by the insured. For full integration, they would even have to acquire property in the agency, which is not imaginable according to known legal codes.

3.4.2.1 Market power of service provider

As in the case of insurer-driven vertical constraints and integration, market power is a necessary condition for success. This condition
Table 3.5 Factors affecting provider-driven vertical integration.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private health insurance</th>
<th>Community-based health insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market power of service provider</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>System efficiency gains to be realized</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Management know-how of provider</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Contestability of insurance market</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Potential to increase entry barriers to competitors</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Contestability of healthcare markets</td>
<td>-</td>
<td>-</td>
<td>n.a.</td>
</tr>
<tr>
<td>Lack of capital of service providers</td>
<td>-</td>
<td>-</td>
<td>n.a.</td>
</tr>
<tr>
<td>Market power of insurer</td>
<td>-</td>
<td>-</td>
<td>n.a.</td>
</tr>
<tr>
<td>Cartelization of insurers</td>
<td>-</td>
<td>-</td>
<td>n.a.</td>
</tr>
<tr>
<td>Legislation prohibiting vertical restraints</td>
<td>-</td>
<td>-</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Note: ↑ Reinforcement of relationship, ↓ Attenuation of relationship, n.a.: not applicable.

usually is not satisfied by a single physician but may be met by a physician network, or a hospital with a large catchment area.

In the rural areas where community-based schemes are typically active, notably hospitals have the market power to impose vertical restraints on insurers or to integrate insurance altogether. An example is provided by the Kisiizi hospitals of Uganda.

### 3.4.2.2 System efficiency gains to be realized

The possible efficiency gains are the same as those discussed in Section 3.4.1.2. It is conceivable that an insurer has enough market power to increase premiums independently of the cost incurred from paying service providers. This again results in a double marginalization, hurting healthcare providers this time.

Skimping on quality by the insurer is also possible in the guise of delayed reimbursement of patients, but also of having unjustified recourse to small print in its insurance policy. However, it is not quite
3.4. Vertical Restraints/Vertical Integration

clear whether the loss of reputation caused falls on the service provider rather than the insurer. In the latter case, there is no externality affecting the healthcare provider.

In the same vein, fraud by the insurer (in particular, failure to pay in the event of insolvency) might constitute a source of within-system inefficiency. The ensuing loss of reputation is more likely to fall on the insurer than the service provider, however.

Negative external effects because of insurers engaging in a technological race do not seem to be an issue either.

Up to this point, incentives for healthcare providers to integrate health insurance into their operations seem to be rather weak. However, provider-driven insurance schemes may have some cost advantages compared to a non-integrated competitor since they already have some relevant risk information about the insured. This is an efficiency gain accruing to healthcare providers.

Healthcare providers and in particular hospitals dealing with community-based schemes must take into account double marginalization since a given scheme usually is the monopoly supplier of health insurance in its region. This consideration speaks in favor of vertical restraints or even full integration. On the other hand, the possibility of such a scheme to deliver substandard quality of service is rather remote. After all, the insured own the scheme themselves, and it is they who would suffer from a lower quality of service than contracted for (Musau, 1999). Also, hospitals are confronted with fraudulent behavior on the part of community-based insurers, as evidenced by the case study of Chogoria Hospital in Kenya. Here, schemes running group policies let non-members (who initially were not identifiable as such at the point of service) present themselves for treatment, creating bad debts for the hospital (Musau, 1999). A technological race is not an issue, most community-based schemes lacking the resources for building up elaborate administrative capacity.

3.4.2.3 Management know-how of provider

Management know-how is another factor, facilitating the implementation of vertical restraints and especially vertical integration.
The lack of management know-how is particularly marked in community-based schemes, leading to even less vertical restraints/integration between health providers and insurers.

### 3.4.2.4 Contestability of insurance market

If the market for health insurance is contestable, a healthcare provider considering vertical integration has a chance to strike an agreement with newcomers. This serves to increase the likelihood of successfully imposing vertical constraints.

As already outlined in Section 3.4.1.4, community-based schemes do not face much contestability of their markets. A newcomer would have to incur extremely high investments to match the advantages of social control enjoyed by them (see Table 3.5).

### 3.4.2.5 Potential to increase entry barriers to competitors

Vertical restraints and integration can also serve a strategic purpose by raising the entry barrier, e.g., to a new hospital. The same applies to physician networks that set up an insurance scheme to the disadvantage of outside physicians.

Hospitals dealing with community-based insurers, being local monopolies, could in principle attempt to protect their markets through integrating the CBI scheme operating in their catchment area. However, the little evidence available suggests that the main motive for provider-driven vertical integration is the prospect of eliminating within-system inefficiencies (see Section 3.4.2.2).

### 3.4.2.6 Contestability of healthcare markets

Providers find it difficult to integrate insurers if their market is contestable. In analogy to the arguments proffered in Section 3.4.1.6, resources must be spent on defending their position in the market, leaving little room for investing in vertical restraints and integration.

Most healthcare providers doing business with community-based health insurers are located in poor rural areas. This means that even if there should be any monopoly rents, their amount must be very
limited. Therefore, the incentive for a new competitor to break into such a market is weak, resulting in a small degree of contestability.

3.4.2.7 Lack of capital of service providers

Especially physician networks may lack capital because their joint liability status impedes their access to capital markets. In a deregulated, competitive market, for-profit hospitals and especially hospital groups may offer an investment with favorable hedging properties. With a measure of independence from the capital market and hence comparatively low beta, they can raise capital at a lower cost than other industries.

Lack of formal capital is a great problem in the case of healthcare providers dealing with community-based insurers. Located in rural areas, neither physicians nor hospitals have easy access to domestic capital markets. In addition, with intermediation by moneylenders incomplete, healthcare providers have difficulty raising internal finance.

3.4.2.8 Market power of insurer

Insurers with market power require ample compensation to let themselves be constrained or integrated.

In community-based schemes, market power of insurers is high since they usually are the only supplier of health insurance coverage. Ceteris paribus, a healthcare provider considering vertical integration would meet with some difficulties.

3.4.2.9 Cartelization of insurers

The costs of negotiation are particularly high in the case of cartelization because all members of the cartel must usually be included.

With regard to community-based schemes, cartelization is of little relevance for two reasons. First, the fact that they often operate along kinship lines makes it more difficult to reach horizontal agreements. Second, as stated in Section 3.4.1.1, community-based schemes usually constitute a monopoly, causing them to have little interest in the protection from competition afforded by a cartel. In sum, this results in an attenuation of cartelization as a factor influencing vertical restraints.
3.4.2.10 Legislation prohibiting vertical restraints

There may be legislation prohibiting medical providers to own an insurer. However, no instance is known to this author, relating to either industrial countries, or community-based schemes in particular.

3.4.3 Actual examples of vertical integration

As Tables 3.4 and 3.5 and their discussion show, there are factors facilitating and hampering both insurer- and provider-driven vertical integration. This leads to the expectation that depending on the mix of these influences, imposition of vertical restraints and attainment of full vertical integration does occur.

Table 3.6 below contains evidence on some of the existing variants of insurer- and provider-driven vertical integration as well as lateral integration. It relates to the competitive case, and community-based schemes, illustrating that the factors discussed above may result in all three types of integration in both settings.

Conclusion 7. Vertical integration constitutes the fifth dimension of the supply of health insurance. Its provider-driven variant depends on the same ten factors ranging from administrative expense to fraud and abuse. Those facilitating insurer-driven integration usually hamper provider-driven integration and vice versa.

3.5 Market Structure

Market structure has several dimensions, among the more important being the number of buyers and sellers and the amount of product differentiation (Carlton and Perloff, 1999, ch. 1). The number of buyers has not been an issue in health insurance markets, even in countries where employers are involved in its provision. With regard to product differentiation, it can be said that its degree increases with the number of sellers unless economics of scope are very marked (see below). Often, the amount of vertical integration is also seen as a dimension of market structure. However, in view of
its great importance for the organization of the healthcare sector as a whole, vertical integration is discussed separately (see Section 3.4). Thus, the number of sellers (and with it, the degree of their concentration) will be retained as the principal dimension of market structure.

One particular aspect of market structure that will be left out from this exposition is the legal form of the insurance company. Originally, most health insurers were mutuals, presumably because a reasonable degree of homogeneity of risks could be attained in this way. Homogeneity of risks ensures that the variance of total claims to be paid does not increase without bounds when more risks are added (Malinvaud, 1972, Appendix). A finite variance in turn implies that the expected value of the loss can be estimated with increased precision (a decreased standard error according to the Law of large numbers),
permitting the insurer to hold less reserves per unit risk while holding its probability of insolvency constant (Cummins, 1991). However, mutuals are at a disadvantage when it comes to raising capital for expanding their risk pool because they do not issue tradable ownership shares.

For this reason, the preferred legal form of insurers has become the publicly traded stock company in industrial countries. Yet, the mutual form is alive and even thriving in the guise of community-based health insurance in low-income countries. In the wake of development, with increasing demand for capital to finance expansion, these schemes may change their legal form to become stock companies. However, assessing the conditions governing such a transition is not the aim of this section. For this reason, it is taken as given that for the foreseeable future private health insurers (which need not be stock companies either) and community-based schemes will continue to coexist in low-income countries.

Focusing on the degree of concentration on the main descriptor of market structure, some important factors influencing it are listed in Table 3.7. The discussion starts with factors that relate to the demand side and then shifts to the supply side. Table 3.7 has no entries for public health insurance for the simple reason that the scheme is assumed to be a monopoly under all circumstances.

### 3.5.1 Diversity of preferences

With greater diversity of preferences, a larger set of differentiated insurance products is necessary for matching supply and demand. This creates potential for niche products written by specialized insurers, and therefore a greater number of companies, ceteris paribus. However, the theory of consumer demand also says that diversity of preferences becomes effective only if incomes are sufficiently high. With a very small income, the attainable consumption set in attribute space is too restricted to permit choices that lie far apart. Therefore, the number of profitable product varieties (and usually firms) is low when income is low.
In the case of community-based schemes, there is the countervailing effect of lacking access to the capital market, which limits the size of the unit and its geographical expansion. The balance of the two influences is an open issue.

3.5.2 Economies of scale

In the case of an insurer, the size of its risk pool may be the source of economies to scale, defined as decreasing unit cost as a function of the number of individuals insured. Thanks to the Law of large numbers, a larger pool size enables the insurer to reduce its reserves per unit risk without increasing its risk of insolvency (Cummins, 1991, Table I). This means that the premiums of a large insurer contain a smaller amount of loading (see Section 3.3.3), which results in a lower premium for a given amount of expected benefits paid. A large insurer could therefore gain even more market share, with the so-called natural monopoly as a possible outcome.

However, a growing pool within a given country may require the acceptance of less favorable risks, with the consequence of a rise in the expected value of the benefit to be paid. Also, a larger pool can be associated with a loss of social control among the insured, encouraging moral hazard. According to Eq. (3.2) of Section 3.3, both effects cause the amount of loading to increase, thus counteracting economies of scale. There does not seem to be very much empirical evidence on this issue in the domain of insurance, let alone health insurance. In fact, the available evidence points to constant rather than increasing returns to scale (see e.g. Fecher et al., 1991). Absent economies to scale, however, there is no reason to expect a particularly high degree of concentration on private insurance markets, at least for this reason.

Fujita et al. (1999), argue that economies of scale occur due to positive spatial externalities. In the present context, this may explain why health insurers in low-income countries concentrate mainly in urban areas. Strong centripetal forces that draw businesses closer to one another (because firms may want to share a customer base or local services, have access to trained and experienced labor) outweigh weaker centrifugal forces that drive businesses farther apart (because
firms compete for labor and land). The first set of influences constitute spillover effects resulting in economies to scale in the guise of lower costs of administration and advertising. As such, they encourage concentration.

Fujita et al. (1999), although not focusing on community-based schemes, also provide an intuition as to why these are concentrated in rural areas. There, strong centripetal forces (such as the ability to serve certain customers and the acceptance of informal market behavior like barter) outweigh the weaker centrifugal forces (such as small customer base, bad infrastructure, and an underdeveloped capital market). Economies of scale thus may occur due to the first set of influences, serving to lower unit costs given the market characteristics of community-based health insurance.

3.5.3 Economies of scope

Economies of scope prevail in insurance if the cost of providing an extra unit of coverage in one line of business decreases as a function of the volume written in some other line. In the context of health insurance, economies of scope may operate at two levels. First, the health insurance line may benefit from other business activities of the same firm. For instance, it may be possible to market health insurance through the existing distribution network for selling, e.g., banking services. The tendency toward an increased degree of concentration in the health insurance market is indirect and hence not very marked in this case. Also, the limited amount of available empirical evidence suggests that economies of scope at this level are not important (see e.g. Suret, 1991).

Second, however, health insurers A and B may realize that while their products are differentiated, the expenses for marketing and administering those of A increase less than proportionately when the quantity of B’s products is increased as well. The amount of loading hence would increase less than proportionately with the expected volume of benefits combined, providing a powerful motive for a merger of the two companies. With economies of scope (often also called synergies) of this second type, there is a tendency toward concentration,
3.5. Market Structure

Table 3.7 Factors affecting the degree of concentration of health insurance sellers in markets for private health insurance.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Private health insurance</th>
<th>Community-based health insurance</th>
<th>Public health insurance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of preferences</td>
<td>–</td>
<td>–</td>
<td>n.a.</td>
</tr>
<tr>
<td>Economies of scale</td>
<td>+/−</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Economies of scope</td>
<td>+</td>
<td>+</td>
<td>n.a.</td>
</tr>
<tr>
<td>Barriers to entry</td>
<td>+</td>
<td>+ ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Barriers to exit</td>
<td>–</td>
<td>– ↑</td>
<td>n.a.</td>
</tr>
<tr>
<td>Antitrust policy</td>
<td>–</td>
<td>–</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

Note: ↑ Reinforcement of relationship, ↓ Attenuation of relationship.

which however does not have to be accompanied with a smaller number of product varieties. More generally, the number of product varieties sold in the market does not vary in step with the number of firms in this case.

This argument holds for community-based health insurance as well (see Table 3.7).

3.5.4 Barriers to entry

High barriers to entry exist when a newcomer to the market must make large investments that cannot be recuperated if entry fails (high sunk costs). Barriers to entry thus cause the degree of concentration to be higher than it would otherwise be. They are clearly relevant in the case of health insurance markets, where a newcomer usually needs to launch an extensive advertising campaign to gain even a small share of the market. This investment cannot be recuperated if the newcomer should decide to withdraw later in time.

A small number of sellers make the negotiation and monitoring of collusive agreements less costly. For this reason, concentration poses a threat to price and product competition also in insurance markets. However, collusive agreements can be destabilized by the emergence of an additional competitor. This destabilization is less likely to occur when there are high barriers to entry. Therefore, barriers to entry not only increase the degree of concentration but may also reinforce the anti-competitive effects that usually accompany a high degree of concentration.
Barriers to entry in community-based health insurance are reinforced by the informal nature of the market (e.g., not all insurance companies are willing to accept payment in kind). Furthermore the relationship between the insurance scheme and its members usually develops over a long period of time (which also helps to minimize moral hazard effects). A newcomer to this market thus would have to make a substantial non-recuperable investment to acquire this experience. This constitutes a barrier to entry facilitating concentration in the community-based segment of the market for health insurance.

3.5.5 Barriers to exit

When challenged by a newcomer, one or several of the incumbents may consider exiting from the market rather than defending their position. However, exit is not an attractive alternative if it entails the loss of investments that cannot be recuperated (i.e., constituting sunk costs). For instance, a sales force specialized in health insurance is not an asset anymore once the firm leaves the market; even with economies of scope, it has a reduced value, e.g., in selling life insurance. Barriers to exit thus keep the degree of concentration lower than it would otherwise be. However, through their stabilizing effect, they still help to preserve collusive agreements, reinforcing the anticompetitive effect of concentration. Bailouts of ailing companies also modify the opportunity cost of leaving the market, thus creating a barrier to exit.

Markets in which community-based schemes operate may be characterized by very high barriers to exit. These schemes benefit from advantages due to their favorable reputation and established social control mechanisms (limiting in particular ex-post moral hazard, see Section 3.3.6), which are lost if an exit from the market occurs. Again, this constitutes a factor that contributes to a lower degree of concentration, ceteris paribus.

3.5.6 Antitrust policy

In many countries, merger projects must be submitted to antitrust authorities. Mergers that would result in a notable increase in the level of concentration are subject to scrutiny according to the rules followed
both by the U.S. Federal Trade Commission and the Commission of the European Union. Up to this point, few mergers of health insurers have been blocked. This does not mean that antitrust policy does not have an impact on concentration. Indeed, the mere risk of having a merger proposal rejected may well keep concentration at a lower level than would otherwise obtain.

Mergers of community-based schemes are very rare, but not because of effective antitrust policies. Arguably antitrust policies do not take effect at all in these schemes, which consist of small groups, whose members share common characteristics like close family and long-run community relationships. Mergers thus come at the cost of increased heterogeneity, which seems to greatly outweigh their benefits. The literature on credit markets offers evidence on the importance of market segmentation along geographic and kinship lines. Udry (1993, p. 95) discovered that loans between individuals in the same village or kinship group accounted for 97 percent of the value of transactions. Hardly any loans were provided to outside communities, as information about repayment possibilities and village sanctions as a mechanism for contract enforcement were lacking. Similar evidence on informal credit markets is reported in a case study of rural China (Feder et al., 1993).

**Conclusion 8.** Market structure as indicated by sellers’ concentration constitutes the sixth dimension of the supply of health insurance. It depends on at least six factors, with barriers to entry exerting a positive and barriers to exit a negative influence.
The Design of an Optimal Health Insurance Contract

The efficiency reasons given above for the existence of SHI with compulsory membership can be 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 toward 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 expenditure.

(a) Administrative expense: Copayment provisions can be called for to save administrative expense 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 expense is proportional to the expected volume of health expenditure, a feature of the optimal insurance contract is a fixed deductible (Arrow, 1963).
(b) **Non-insurable loss:** Illness typically involves not only monetary 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-monetary and monetary losses. In particular, if marginal utility of wealth is lower in case of illness than in good health (due to reduced ability to enjoy expensive types of consumption), optimal health insurance does not fully reimburse the monetary 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 convex 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 more price elastic the demand for the particular type of medical services, the higher the optimal copayment rate (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 9. The optimal health insurance contract suggests several reasons for stopping short of full coverage. While administrative expense 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 appears 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. 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, 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 social health insurance as traditionally understood. When risk adjustment is imposed in order to “marry” uniform premiums in the face of differences in expected cost with competition, insurers considering innovation fear the financial sanction that goes along with attracting young clients (a more formal analysis is provided in Appendix A). 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 (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 positively correlated,
an unfavorable situation for risk-averse individuals. Therefore, insurance as a system should at least mitigate those 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 5.1). Not one out of a total of six correlation coefficients is significantly negative.

Also, social health insurance fails to make up for shortfalls in the benefits of private insurance. While employee benefits for old age, which

<table>
<thead>
<tr>
<th></th>
<th>OAS</th>
<th>MCHI</th>
<th>MCSM</th>
<th>MA</th>
<th>WC</th>
<th>DI</th>
<th>UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>OAS*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health†</td>
<td></td>
<td>0.82*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCHI</td>
<td></td>
<td></td>
<td>-0.29</td>
<td>-0.31</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCSM</td>
<td>-0.55*</td>
<td>-0.40</td>
<td>0.33</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td></td>
<td>-0.41</td>
<td>0.16</td>
<td>0.93*</td>
<td>0.77*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Workers compensation</td>
<td>-0.64*</td>
<td>-0.65*</td>
<td>0.31</td>
<td>0.84*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disability DI</td>
<td></td>
<td>-0.41</td>
<td>0.16</td>
<td>0.93*</td>
<td>0.77*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Unemployment UI</td>
<td>0.02</td>
<td>0.28</td>
<td>0.29</td>
<td>0.63*</td>
<td>0.24</td>
<td>0.70*</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPI</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social health insurance SPI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social accident insurance SAI</td>
<td>0.45</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment insurance UI</td>
<td>0.91*</td>
<td>0.67*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.76*</td>
<td>0.65*</td>
<td>0.83*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Correlation coefficient significant at percent level or better.

† MCHI = Medicare Hospital Insurance, MCSM = Medicare Supplementary Medical, MA = Medicaid.

Table 5.2 Correlations of trend deviations in the benefits of U.S. and German private and social insurance.

<table>
<thead>
<tr>
<th></th>
<th>Social health</th>
<th>Social accident</th>
<th>Social disability</th>
<th>Social unemployment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United States,</strong></td>
<td><strong>OAS</strong></td>
<td><strong>MCHI</strong></td>
<td><strong>MCSM</strong></td>
<td><strong>MA</strong></td>
</tr>
<tr>
<td><strong>1974–1992</strong></td>
<td>Life</td>
<td>0.74*</td>
<td>0.47*</td>
<td>−0.33</td>
</tr>
<tr>
<td></td>
<td>Health</td>
<td>0.67*</td>
<td>0.62*</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Disability</td>
<td>−0.35*</td>
<td>−0.13</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>Germany (West),</strong></td>
<td><strong>SPI</strong></td>
<td><strong>SHI</strong></td>
<td><strong>SAI</strong></td>
<td><strong>UI</strong></td>
</tr>
<tr>
<td><strong>1975–1993</strong></td>
<td>Life</td>
<td>0.27</td>
<td>0.26</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Health</td>
<td>0.79*</td>
<td>0.56*</td>
<td>0.92*</td>
</tr>
<tr>
<td></td>
<td>Accident</td>
<td>−0.41</td>
<td>−0.15</td>
<td>−0.41</td>
</tr>
<tr>
<td></td>
<td>Liability</td>
<td>0.43</td>
<td>0.16</td>
<td>0.54*</td>
</tr>
</tbody>
</table>

*Correlation coefficient significant at percent level or better. Source: Zweifel (2000b).

are counted as social insurance, are negatively correlated with PHI (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 5.2). Conversely, again not a single out of 20 coefficients of correlation is significantly negative in the case of Germany (3 out of 21 are negative in the case of the United States). 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, 3 out of 10 correlation coefficients are significantly negative, pointing to at least some diversification effects within private insurance (not shown). 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.

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 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 1993, 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., 2006).

Reading Table 5.3 horizontally first, one notes that the amounts of compensation asked are consistently highest for consenting to a physician list based exclusively on cost criteria (column 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. 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 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.
Table 5.3 Compensation asked for cutbacks in Swiss social insurance (2003).

<table>
<thead>
<tr>
<th>Socioeconomic characteristics</th>
<th>Physicians selected on cost criteria</th>
<th>Physicians selected on quality criteria</th>
<th>Physicians selected on cost &amp; quality criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amounts in CHF/month</td>
<td>Access to new therapies</td>
<td>Reimbursement of generics only</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>No small local hospitals for minor complaints</td>
</tr>
<tr>
<td>Access to new therapies delayed by 2 years</td>
<td>106 (13.2)</td>
<td>53 (8.8)</td>
<td>42 (7.8)</td>
</tr>
<tr>
<td>Reimbursement of generics only</td>
<td>88 (11.8)</td>
<td>38 (7.8)</td>
<td>26 (6.8)</td>
</tr>
<tr>
<td>No reimbursement of drugs for minor complaints</td>
<td>191 (76.3)</td>
<td>128 (58.5)</td>
<td>136 (56.9)</td>
</tr>
</tbody>
</table>
| Note: 1 CHF equals 0.70 at 2003 exchange rates. Standard errors in parentheses.
and likely reflects “warm glow” (Andreoni, 1995)). Finally, another cut-back would be the concentration of dispersed existing hospital capacity in larger, centralized units. 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. 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. Of course, wealthy individuals can always opt out by paying extra; however, poor individuals do suffer a loss because the 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, 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 speakers.

**Conclusion 10.** 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 WTP for additional coverage that falls short of its additional cost.
Summary and Conclusions

This text revolves about two related basic 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 2 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 at 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 supervisors); however, its incentives are weaker than the competitive insurer’s given reasonable parameter constellations (Appendix A).

The supply of insurance has five dimensions, (1) comprehensiveness of the benefit packages, (2) amount of risk selection efforts deployed by insurers, (3) the amount of loading contained in the premium, (4) amount of vertical integration, and (5) degree of sellers’ concentration. Each dimension depends on several factors, among which moral hazard effects (that however in their turn can be influenced by the design of the contract) are among the most prominent (Conclusions 3 to 8).

The importance of contract design motivates a review of the theory of the optimal health insurance contract, which also serves as a point of departure for exploring the limits of social health insurance. 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 (Conclusion 9). 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 WTP 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 10).

These limits will become more important in the future as the cost of health care increasingly occurs toward 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.
Among the five dimensions of supply distinguished in the body of Section 3, only two are retained here for simplicity. Moreover, many of the influences listed particularly in Tables 3.1 through 3.5 are neglected for simplicity. The decision situation of an insurer under the pressure of competition will be analyzed first, followed by that of a public monopoly insurer.

A.1 Competitive Health Insurer

A competitive health 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 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 \( \bar{P} \)) is regulated to be uniform and constant (in spite of lowered expected losses thanks to innovate effort to keep things simple). 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 efforts 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 read as (see Zweifel and Eisen, 2003, ch. 5.5.2 for a similar model),

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

\[
\text{with } \frac{\partial \mu}{\partial i} < 0, \quad \frac{\partial \mu}{\partial s} < 0, \quad \frac{\partial L}{\partial i} < 0.
\]

Neglecting boundary solutions, the first-order conditions for an optimum are given by

\[
\frac{\partial E\Pi}{\partial i} = \frac{\partial \mu}{\partial i} \left\{ \bar{P} - \pi^h L \right\} + \mu \left( -\pi^h \frac{\partial L}{\partial i} \right) - \frac{\partial \mu}{\partial i} \left\{ \bar{P} - \pi^l L \right\} + \{1 - \mu\} \left\{ -\pi^l \frac{\partial L}{\partial i} \right\} - 1 = 0, \tag{A.2}
\]

\[
\frac{\partial E\Pi}{\partial s} = \frac{\partial \mu}{\partial s} \left\{ P - \pi^h L \right\} - \frac{\partial \mu}{\partial s} \left\{ \bar{P} - \pi^l L \right\} - 1 = 0. \tag{A.3}
\]
Focusing on Eq. (A.2) first, and multiplying it by \((i/\mu)\), one obtains
\[
\frac{\partial \mu}{\partial i} \frac{i}{}\left\{ (\bar{P} - \pi h L) - (\bar{P} - \pi l L) \right\} \]
\[
- \left( \frac{\partial L}{\partial i} \right) \left\{ \mu^* \pi h L + (1 - \mu^*) \pi l L \right\} / \mu^* = \frac{i^*}{\mu^*}. \quad (A.4)
\]

The first term on the 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 insured 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 \(\bar{P}\), 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, 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 RHS of Eq. (A.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. The RHS of Eq. (A.4) is nothing but the marginal cost of innovation, again distributed over the high risks.

Turning to Eq. (A.3) now and multiplying it through by \((s/\mu)\), one obtains
\[
\left( \frac{\partial \mu}{\partial s} \right) \frac{s}{\mu} \left\{ (\bar{P} - \pi h L) - (\bar{P} - \pi l L) \right\} = \frac{s^*}{\mu^*}. \quad (A.5)
\]

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. (A.5)). However, its marginal cost can again be distributed over the number of high-risk insured $\mu$.

For comparative-static analysis, the point of departure are Eqs. (A.4) and (A.5), slightly rewritten,

$$e(\mu, i) \left\{ (P - \pi h L) - (P - \pi l L) \right\} - e(L, i) \left\{ \mu \pi h L + (1 - \mu) \pi l L \right\} / \mu = \frac{i}{\mu(i, s)}, \quad (A.6)$$

$$e(\mu, s) \left\{ (P - \pi h L) - (P - \pi l L) \right\} = \frac{s}{\mu(i, s)}. \quad (A.7)$$

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

$$\frac{\partial}{\partial R} \left[ (P - \pi h L) - (P - \pi l L) \right] := \frac{\partial A}{\partial R} < 0. \quad (A.8)$$

For future reference, one also has

$$\frac{\partial}{\partial i} \left( \frac{i}{\mu(i, s)} \right) = \frac{\mu - i \cdot \partial \mu / \partial i}{\mu^2} = \frac{1 - e(\mu, i)}{\mu}. \quad (A.9)$$

Now, let the first-order conditions (A.6) and (A.7) be subjected to a shock $dR > 0$. Written in matrix form, the comparative statics read, using (A.6) to (A.9),

$$\begin{bmatrix} [e(\mu, i) - 1]/\mu & 0 \\ 0 & [e(\mu, s) - 1]/\mu \end{bmatrix} \begin{bmatrix} di \\ ds \end{bmatrix} = \begin{bmatrix} -e(\mu, i) \cdot \partial A/\partial R \\ -e(\mu, s) \cdot \partial A/\partial R \end{bmatrix} dR. \quad (A.10)$$

Applying Cramer’s rule, one obtains

$$\frac{di}{dR} = \frac{1}{H} \begin{vmatrix} -e(\mu, i) \cdot \partial A/\partial R & 0 \\ -e(\mu, s) \cdot \partial A/\partial R & e(\mu, s) - 1/\mu \end{vmatrix} < 0, \quad (A.11)$$
with $H < 0$ symbolizing the determinant of the Hessian matrix. Applying Cramer’s rule once more yields

$$\frac{ds}{dR} = \frac{1}{H} \left| \begin{array}{cc} e(\mu, i) - 1 & e(\mu, i) \cdot \partial A / \partial R \\ \mu & -e(\mu, s) \cdot \partial A / \partial R \end{array} \right| < 0.$$  \hspace{1cm} (A.12)

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

### A.2 Public Monopoly 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 (2004); 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. (2003)). Here, a much simpler alternative will be presented that has the advantage of facilitating comparisons with Section A.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 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 as

$$EU = \rho(D) \cdot U(D), \quad \text{with} \quad \frac{\partial U}{\partial D} D < 0.$$  \hspace{1cm} (A.13)

If there is a deficit ($D < 0$), then the official has an increase in utility if $D$ increases toward zero ($\partial U / \partial D$); if the scheme has a surplus
Health Insurer Behavior in Terms of Innovation and Risk Selection Effort

$(D > 0)$ 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 $\partial U/\partial D = \pm 1$ 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 high 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 $(D < 0, \partial U/\partial D = 1)$, one can write the public health insurer’s objective function as

$$\max_{i} EU = \max_{i} \rho(D)D.$$  \hspace{1cm} (A.14)

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

$$\frac{\partial EU}{\partial i} = \frac{\partial \rho}{\partial D} \cdot \frac{\partial D}{\partial i}D + \rho \cdot \frac{\partial D}{\partial i} = 0.$$  \hspace{1cm} (A.15)

This yields

$$\frac{\partial \rho}{\partial D} \cdot \frac{D}{\rho} = -1, \quad \text{or} \quad e(\rho, D) = -1.$$  \hspace{1cm} (A.16)

Note that $e(\rho, D)$ is a constant by assumption. Therefore, the choice of $i^{**}$ by the public health insurer is completely arbitrary. However, since $i$ impinges on its budget, $i^{**} = 0$ is the dominant solution. By way of contrast, $i^{**} > 0$ generally pertains in the case of the competitive health insurer.
The main motive to regulate health insurance is to (1) eliminate the social costs caused by insolvency by preventing insolvency, or (2) mitigate the social costs caused by insolvency while accepting the possibility of insolvency (Zweifel and Eisen, 2003, ch. 8.1). Indeed, individuals losing their health insurance protection may face hardship and poverty that affect society as a whole. The two issues are taken up below.

(1) Regulations designed to eliminate insolvencies also seek to avoid instability in insurance markets that may occur due to adverse selection processes. Typically, they are very comprehensive and detailed because current operations of insurers must be monitored to attain the objective. However, this type of regulation generates inefficiency because it prevents insurers from adopting least-cost solutions. Thus, regulation aimed at avoiding insolvency under all circumstances may not maximize social welfare. Once private insurance schemes are fully regulated – such that, e.g., prices, quantity and quality, of private insurance products are determined outside the market mechanism – resource allocation is likely to deteriorate. In other words, wrong insurance product pricing, wrong insurance packages, and reduced competitive behavior may lead to an inefficient and inequitable allocation. Table B.1 below provides
Table B.1 Regulations that tend to lower efficiency.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Imposed premiums</td>
<td>Lack of incentive signals</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Undermines price competition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Premium fails to reflect expected costs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disturbs balance of underwriting and investing activities</td>
</tr>
<tr>
<td>2</td>
<td>Obligation to provide specific products, approval of product</td>
<td>Restricts product competition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Does not reflect individual benefit-cost estimates</td>
</tr>
<tr>
<td>3</td>
<td>Rules on active/passive ownership (vertical integration)</td>
<td>Prevents insurers from finding the optimal degree of vertical integration</td>
</tr>
<tr>
<td>4</td>
<td>Obligation to provide certain benefits and/or to insure certain risks</td>
<td>Can make insurance not viable</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Does not reflect individual benefit-cost estimate</td>
</tr>
<tr>
<td>5</td>
<td>Separation of lines of business</td>
<td>Loss of synergy effects both for insured and insurer (allocation of reserves is not optimal)</td>
</tr>
<tr>
<td>6</td>
<td>Budget approval</td>
<td>Hampers product innovation</td>
</tr>
<tr>
<td>7</td>
<td>Rules on investments</td>
<td>May prevent insurers from obtaining maximum expected return for a given volatility</td>
</tr>
<tr>
<td>8</td>
<td>Subsidies and tax exemptions in favor of insurers</td>
<td>Justified if insurers provide a public good (e.g., cohesion of society)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Induces overconsumption of insurance</td>
</tr>
<tr>
<td>9</td>
<td>Obligation to contract with providers</td>
<td>Lowers pressure on providers to reach efficiency</td>
</tr>
</tbody>
</table>

an overview of regulations that tend to lower efficiency, along with a short explanation. For example, budget approval (item 6) stifles product innovation because, apart from possible delays, the insurer runs the risk of having the cost of innovation disapproved.

(2) However, regulation can also be designed to reduce social cost once insolvency occurs by making insurers bear that cost. One way to internalize it is to require the deposit of reserves, another, the establishment of a guaranty fund financed jointly by the insurers (items B and F of Table B.2). These measures go a long way toward eliminating hardship of insured in the advent of insolvency. Even these regulations are not without their cost, however, because, e.g., the reserves deposited could usually be invested at a higher rate of return. In addition, there is a direct cost of administering these regulations. On the whole, however, regulations motivated by the objective of internalizing the social cost of insolvency seem to have a better chance of being efficiency-enhancing.
Table B.2 Regulations that tend to enhance efficiency.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Licenses for insurers</td>
<td>Serves to lower probability of insolvency</td>
</tr>
<tr>
<td>B</td>
<td>Minimum capital</td>
<td>Serves to lower probability of fraud</td>
</tr>
<tr>
<td>C</td>
<td>Minimum liquidity</td>
<td>Serves to lower probability of insolvency</td>
</tr>
<tr>
<td></td>
<td>requirements</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Reinsurance schemes</td>
<td>Serves to lower probability of insolvency</td>
</tr>
<tr>
<td>E</td>
<td>Provision of a guarantee fund</td>
<td>Serves to lower probability of insolvency</td>
</tr>
<tr>
<td>F</td>
<td>Industry-wide insolvency fund</td>
<td>Serves to lower probability of insolvency</td>
</tr>
<tr>
<td>G</td>
<td>Provision of information to regulators and consumers</td>
<td>Serves to increase transparency</td>
</tr>
<tr>
<td>H</td>
<td>Agreed-on accounting procedures, internal and external auditing</td>
<td>Serves to increase transparency</td>
</tr>
<tr>
<td>I</td>
<td>Mandatory risk adjustment scheme among insurers in the presence of adverse selection</td>
<td>Avoids cream-skimming by insurers</td>
</tr>
</tbody>
</table>

Finally, insurance regulation may have the objective of creating demand for private coverage, which is seen as a precondition for an expanded provision of private health care and the reaping of efficiency gains associated with it (Griffin, 1989, p. 23).


Transparency International (several years). [www.transparency.org].


