Research Memorandum

**No 171**

# Regulated Competition in Health Insurance Markets

**Rudy Douven**[1]

CPB Netherlands Bureau for Economic Policy Analysis, The Hague, November 2000

[1] I would like to thank Ed Westerhout for his inspiring discussions and suggestions during the project. Furthermore, I would like to thank Leon Bettendorf, Erik Schut and the participants in a seminar at the Ministry of Health, Welfare and Sports for helpful comments on an earlier version of this paper.

CPB Netherlands Bureau for Economic Policy Analysis
Van Stolkweg 14
P.O. Box 80510
2508 GM The Hague, The Netherlands

Telephone +31 70 338 33 80
Telefax +31 70 338 33 50

ISBN 90-5833-051-6

CONTENTS Page

## 1. Introduction

The optimal design of health insurance policies is a major challenge for many governments. Such policies must make trade-offs between risk sharing on the one hand and agency problems on the other. A central theme in insurance is that increasing the generosity allows for a better dispersion of risks but also leads to higher expenditures because individuals are induced to consume more care. A similar theme is at work in the relationship between governments and health care insurers. The more generous the government subsidizes health care plans the weaker are the incentives for the insurers to undertake efficiency activities and the higher health care expenditures will be (see also Cutler and Zeckhauser (2000) and Newhouse (1996)).

Traditionally, health care insurers are considered to play two different roles. First of all, they supply insurance contracts, which serve to reduce the variability of financial outcomes due to medical risks by pooling a large number of people. Second, health care insurers should play a role as countervailing power to the suppliers of medical services. Especially in the field of medical care where asymmetric information contributes to a dominant position of suppliers, insurers may be better equipped than patients to bargain for a high quality of medical services at low cost.

For a long time, the organization of health care insurance in many countries concentrated on the first of these two aspects. For example, in the Netherlands, health care insurance for basic cure services such as services delivered by general practitioners, medical specialists and pharmaceuticals is administered by a large number of sickness funds. These sickness funds were used to operate under a scheme of full cost reimbursement: for each guilder paid to their insured, sickness funds received a guilder from the sponsor of the scheme, *i.e.* the Dutch government.

Recently, things have begun to change. Since 1992, Dutch government policies aim at a gradual reduction of reimbursement rates. These policies aim to induce insurers to undertake actions that raise the efficiency of the delivery of medical services. Thus, insurers should pay more attention to their role of countervailing power party instead of their role of supplier of health care insurance.

Although our paper will concentrate on the Netherlands, more countries employ similar reimbursement and/or risk adjustment schemes. In the United States, the Federal government announced in 1999 its intentions to employ risk adjustment to pay HMO's that enroll individuals in medicare (Cutler and Zeckhauser (2000)). In other countries (*e.g.* Israel and Colombia) risk adjustment premium subsidies are such that community-rated premiums should be zero for all enrollees (See Van de Ven and Ellis (2000)).

This paper explores the effects of (im)perfect risk adjustment schemes and the effects of two types of reimbursement schemes, retrospective equalization and retrospective compensation, that characterize the financing of Dutch sickness funds. It analyses the effects of these schemes on the efficient delivery of medical services and on health care premiums. It finds that policies that foster competition among sickness funds not only increase the efficiency that features the delivery of medical services but also may increase risk-type specific selection activities by insurers. Furthermore, it finds that premiums are influenced by several opposite effects. Hence, the recent policy shift in the Netherlands may be a mixed blessing. Whether it is beneficial or harmful depends on the exact working of the several institutions in the playing field.

The paper adopts a static model that fully integrates the behaviour of insurers with that of consumers of health care services, *i.e.* patients, the Dutch government and the providers of care. Insurers have a number of instruments at their disposal to influence outcomes. In particular, they can choose the intensity of efficiency activities and selection activities. In addition, they can set their premiums. Our model recognizes the heterogeneity of consumers observed in reality by distinguishing high and low risks (as is done in Rothschild and Stiglitz (1976)). Furthermore, high and low risks may face different premium and selection elasticities. Insurers can explore selection activities separately for the two groups of risks. However, legally they are not allowed to risk differentiate their premiums.

The model is inspired by the overview of Van de Ven and Ellis (2000).[2] The model is related to the models in Chalkley and Malcomson (1998a, 1998b) and Ellis (1998) and Ellis and McGuire (1990). Whereas most of these models use an integrated approach for modeling the health care insurer and the health care provider, we make an explicit distinction between these two types of institutions. Surprisingly enough there are no models in the health care literature which take the perspective of the health care insurer as point of departure, who has to cope with at least four different actors simultaneously: the consumers, the government, the health care provider and other insurers. Besides price competition among insurers we also consider the possibility of (risk-) selection competition among insurers and, furthermore, model the consequences of competition on the efficiency in the provision of care. In our model the government has also a prominent role, since we model explicitly a prospective and two retrospective payment schemes. This enables us to study the consequences of risk adjustment and reimbursing health plans on the trade-off efficiency in production versus selection (Newhouse (1996)). The literature lacks, as far as we know, an explicit model of these phenomena.

The structure of the paper is as follows. Section 2 gives the institutional background by describing the reimbursement schemes of Dutch sickness funds. Section 3 sets out the formal model. Section 4 explains the outcomes of the model in the Nash equilibrium.

[2] The model is part of a larger modelling project which covers the complete Dutch health care sector. See CPB(1999) for an overview of this project.

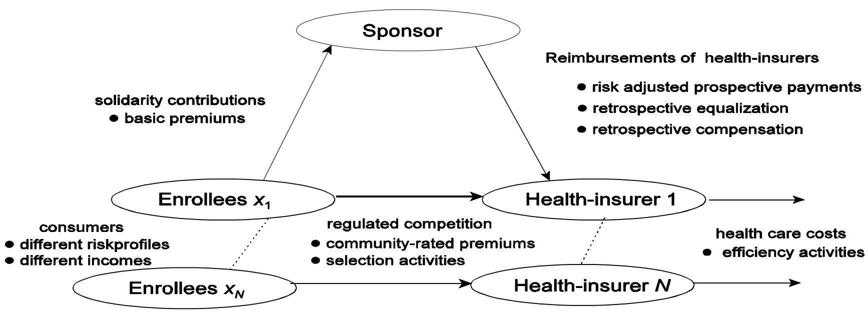
Section 5 discusses various assumptions of the model and section 6 contains concluding remarks.

## 2. Sickness funds in the Netherlands

Figure 1 shows the public scheme for health insurance as administered by sickness funds in the Netherlands. This public scheme regulates insurance for those with labour income below a certain threshold (about 30.000 euro). Insurance is obligatory for those who are eligible and covers about two-third of the Dutch population (about 10 million people). The scheme covers health expenditure on basic cure services like for example hospital care, care delivered by general practicioners and pharmaceuticals. It is administered by a large number of independent sickness funds (28 in 1999). Consumers are allowed to switch yearly from insurer to another, and acceptance is obligatory. Each consumer faces equal benefit packages, as designed by the government, and two types of premiums: a basic premium and a community-rated premium.

The basic premium is uniform across sickness funds and is a function of income of the insured. This contributes to income solidarity.[3] In figure 1 these payments reflect the solidarity contributions of the consumer to the sponsor, *i.e.* the Dutch government. The sponsor collects the solidarity contributions of consumers and reimburses these contributions across sickness funds. The sponsor splits these reimbursements into a prospective and retrospective part.

*Figure 1* *The health insurance system as administered by Dutch sickness funds.*




[3] This element is not central in a scheme of regulated competition however. Indeed income solidarity can also be achieved in other ways, like through the tax system.

- Risk adjusted prospective payments

For each enrollee an insurer receives prospective payments which depends on the risk characteristics of the enrollee, *i.e.*, the sponsor performs risk adjustment. Risk adjustment implies that the sponsor uses consumer information to calculate the expected health care expenditures of individual consumers. The idea is that the better the match is between risk characteristics and insurer reimbursements, the less incentives sickness funds will have to apply risk-selection. Currently, the risk-adjusted payments equals the risk-adjusted predicted per capita costs at the national level minus a fixed amount that is identical for all persons (see Van de Ven and Ellis (2000)). Risk adjusters that are used include age, sex, disablement and region-characteristics of the consumer.

Prospective payments cover only a part of the realised health care costs and control only very imperfectly for risk differentials between different insurers. Currently, the risk-adjusters on which the prospective payment scheme is based explain only about 10 percent of the variation in health care costs (van Vliet *et al.*, 1999). Hence, without any retrospective element, insurers that are unlucky to have a high proportion of bad-risk consumers, would face higher losses. Moreover, any existing information gap about individual health characteristics of the consumers between the insurers and the government might induce insurers to apply risk-selection. Therefore, sickness funds receive also retrospective payments.[4] Two types of payments are distinguished.

- Retrospective equalization

Retrospective equalization is a form of risk sharing among insurers. The idea is to compensate losses of insurers by gains of other insurers. For example, at the end of the year it may become visible that realised health care costs are unevenly distributed across insurers, due to failures in the risk adjustment scheme of the sponsor or simply due to bad luck of some insurers. The exact percentage of the amount of losses or gains which will be equalized is determined each year by the sponsor. Retrospective equalization is budget neutral and, therefore, involves no extra sponsor subsidies to the insurers.

- Retrospective compensation

Retrospective compensation implies that the sponsor additionally subsidizes part of the difference between realised health care costs and the prospective and retrospective equalization payments. This payments scheme compensates high-cost insurers more than low-cost insurers.

[4] In the Netherlands a distinction is made between four categories: variable costs of hospital care, fixed costs of hospital care, medical specialists and other health provisions. Each category has its own form of risk sharing.

So far we have explained the upper part of figure 1; basic premiums of the consumer flow through the sponsor to the health insurer. The main point is that these flows are regulated by the sponsor. This enables the sponsor to introduce all types of solidarity issues into the insurance system, such as risk and income solidarity across consumers and cost solidarity across insurers.

However, in order to stimulate insurers to implement more efficient activities to control costs, the Dutch government introduced, about eight years ago, elements of regulated competition into the system. This is visible in the lower part of figure 1. Health insurers are also allowed to raise community-rated premiums.

Community-rated premiums are set by individual sickness funds. However, sickness funds are not allowed to differentiate across different risk categories. All enrollees contracted by the same insurer pay identical community-rated premiums. Community-rated premiums have to be paid by the enrollee directly to the insurer. Since community-rated premiums are allowed to differ across insurers, insurers can use this premium to attract consumers. This element of price competition is expected to increase efficiency efforts of insurers. The argument is that less efficient insurers will produce higher costs and, therefore, will have to set higher community-rated premiums than efficient insurers. Consumers will choose their most favourable insurer since they are allowed to change from their insurer yearly. Other people argue that efficiency activities will not get off the ground, for example due to the strong bargaining position of health care providers, and that insurers will increase their selection activities to attract more profitable consumers.

In the next section the above plan will be modelled explicitly. The model is built from the perspective of the health care insurer who may pursue efficiency and selection activities and in setting a community-rated premium. The underlying assumptions of the model are that insurers face price and (risk-) selection competition and have full information (*i.e.* and possibly more information than the government) when determining their optimal behaviour.

## 3. The model

The model considers a population of $x$ consumers, who differ with respect to their medical risk. We distinguish a population of low risks, $x_L$ , and a population of high risks, $x_H$. We assume that there are $N$ insurers in the health care market. Each insurer $i \in \{1,..,N\}$ attracts $x_i$ consumers of which some are low risk consumers, $x_{iL}$ , and some are high risk consumers, $x_{iH}$:

$$x := \sum_{i=1}^{N} x_i := \sum_{i=1}^{N} x_{iL} + \sum_{i=1}^{N} x_{iH} := x_L + x_H \tag{1}$$

Insurers are obliged to accept any person who wants to enter. Hence $x$, $x_L$ and $x_H$ are given for insurer $i$. Insurers have the disposal of the following instruments:

- Each insurer $i$ may invest in efficiency activities, $e_i$ , by inducing health care providers to keep costs down without stinting on quality of service.[5]
- Each insurer $i$ is allowed to choose a community-rated premium, $p_i$. This premium may differ across insurers but is equal for every, low or high risk, enrollee contracted by insurer $i$.
- Each insurer $i$ may explore two types of selection activities. Selection activities on low risk types, $s_{iL}$, and on high risk types, $s_{iH}$ . Selection activities may correspond to all type of activities to attract consumers, such as marketing activities or less ethical selection activities such as creaming (see Ellis (1998)).[6]

### 3.1. Health care costs

Health care costs differ among consumer-type. We assume that for each type of consumer, the insurer faces the following health care costs:

$$z_{ik} = z_k^0 - \alpha_k \left( e_i^\gamma + \frac{\beta}{N-1} \sum_{j \neq i} e_j^\gamma \right) , \qquad (2)$$

for $k$=$H,L$, $i, j$ =1,...,$N$, $0 \leq \alpha_k$ , $0 \leq \beta \leq$ N-1 , $0<\gamma<1$.

This equation postulates that health care costs for an enrollee of risk type $k$ belonging to insurer $i$ equals $z_{ik}$. Health care costs depends on two terms:

- Fixed costs

The fixed costs $z_k^0$ are assumed only to depend on the type of risk and are equal for all insurers. In the absence of efficiency activities, low risk enrollees will be assumed less costly than high risk enrollees: $z_L^0 < z_H^0$.

- Efficiency and health care costs

Carrying out efficiency activities, $e_i$ , lowers health care costs for each consumer. The impact of efficiency activities upon health care costs will depend on $\alpha_k$ ,$\beta$ and $\gamma$. These parameters reflect the institutional setting of insurers. $\alpha_k$ indicates to what extent

[5] We follow here the definition of Newhouse (1996). Efficiency in production refers to least cost treatment of a patient's medical problem, holding quality constant.

[6] It may also correspond to general quality activities of insurers such as services. Note that dumping is not possible; the whole population has to be ensured by law. Furthermore, the model does not consider skimping (*i.e*. reducing quality of medical treatment by providers for profit reasons).

efficiency activities may yield different gains for high and low risks.[7] $\beta$ reflects the spillover gains of efficiency activities of insurers. For example in a situation where a hospital provides similar services to consumers of different insurers, spillover effects are likely to occur since health care providers are not likely to distinguish enrollees of different insurers. If efficiency activities of an insurer have a substantial impact on health care costs of enrollees of other insurers then $\beta$ is closer to $N$-1. This situation is typical for an environment where consumers have a large freedom of choice with respect to health care providers, irrespective of their choice of insurer. However, at the other extreme, there are independent (HMO type of) insurers, who explore their own hospitals, specialists and general practitioners. In such an institutional setting $\beta$ will be much smaller, but $\alpha_k$ may be higher. If the ties among health care insurers and health care providers are more intense and exclusive, building up efficiency activities are likely to be more fruitful and may be less costly.[8] The extreme case here is $\beta$=0. This corresponds to a situation where insurers and health care providers are completely vertically integrated. Remark that the impact of $\beta$ is divided by the number of insurers minus one. This choice is arbitrary and mainly simplifies the presentation of outcomes. It takes care of the fact that health care costs of a consumer are independent of the number of insurers, if each insurer aims for an equal amount of efficiency activities. Finally, $\gamma$ is assumed to be between $0<\gamma<1$, indicating that marginal efficiency activities are decreasing in the intensity of efficiency behaviour.

In the above equation we implicitly assume that efficiency activities of the insurer may reduce health care costs. More explicit models of the relationship between insurer and health care providers are numerous in the literature. Chalkley and Malcomson (1998) use contract theory, in the style of Laffont and Tirole (1993), and take also the consequences for quality of services into account. Ellis (1998) and Ellis and McGuire (1990) consider different payment systems for the provider and model explicitly the bargaining situation between the health care provider and the patient.

[7] Although the impact of efficiency may differ for different risk types it is not allowed for the insurer to differentiate efficiency activities across different risk types. We assume that health care providers follow their ethical code and treat patients on basis of their diagnosis and independent of the fact whether they are low or high risks.

[8] There is still a large dispute going on here. Glied (2000) summarizes empirical research on managed care. She concludes that managed care plans reduce the rate of health care utilization somewhat. Less evidence, however, exists on their effect on overall health care costs and cost growth. A recent study of Altman, Cutler and Zeckhauser (2000) finds that treatment intensity differs hardly between HMO's and an indemnity plan, but HMO's charge a lower price.

### 3.2. Attracting consumers

We assume that an insurer attracts consumers on basis of their premium and their selection activities.[9] We assume that the more selection activities (in high or low risk) undertakes or the lower the insurer sets the premium, the more consumers will be attracted. In the model insurer *i* attracts consumers of risk type *k* by the following equations:

$$x_{ik} = x_{ik}^{0} - \zeta_k x_k (p_i - \frac{1}{N-1} \sum_{j \neq i} p_j) + \delta_k x_k (s_{ik}^{\eta} - \frac{1}{N-1} \sum_{j \neq i} s_{jk}^{\eta}) \qquad (3)$$

for $k = L,H$, $i, j = 1,...,N$ , $0 < \eta, \delta_k, \zeta_k < 1$.

The equations shows that insurers compete with three different type of instruments to attract consumers. Insurers can attract low-risk type consumers by undertaking selection activities $s_{iL}$, or high risk types by undertaking selection activities $s_{iH}$. The third instrument to attract consumers is the community-rated premium $p_i$. The equation posits that insurers can undertake selection activities as a substitute for premium adjustments. We will now explain each term in the equation separately.

- The constant term.

If selection activities across insurers are equal, $s_{iL} = s_{jL}$ and $s_{iH} = s_{jH}$ for $i,j \in \{1,...,N\}$, and also community-rated premiums are equal across insurers, $p_i = p_j$ , we assume that each insurer has a certain number of enrollees: $x_{ik}^{0}$, with $\sum_{i=1}^{N} \sum_{k=H,L} x_{ik} = x$. Remark that the equation is constructed such that for any choice of $s_{iL}$, $s_{iH}$ or $p_i$ the sum $\sum_{i=1}^{N} \sum_{k=H,L} x_{ik} = x$ always holds.

- Community-rated premiums

Everything else equal, if the premium of insurer *i* is higher than the average premium of the other insurers, insurer *i* will lose market share on the insurance market. A one unit change of the community-rated premium implies that a share $\zeta_L$ of the population of risk type *L* (and a share $\zeta_H$ of the population risk type *H*) will change from insurer. Low- and high risks may react differently on premium changes as is reflected in different shares $\zeta_L$ and $\zeta_H$. For example, high risks may attach more weight to their health status, and thus may generate a higher search intensity, than low risks.[10]

[9] If insurers and health care providers are integrated then the distance to the provider may play a crucial role for the consumer. Therefore Ellis (1998) considers also travel time. However, travel time plays a much smaller role if consumers may choose their provider irrespective of their choice of insurer. The latter resembles more the case in the Netherlands.

[10] The plausibilty of this reasoning should follows from the fact that in our model all consumers are equal, except with respect to their medical risk. For example, it is not inconsistent with a recent empirical

- Selection activities

Given equal premiums among insurers, the equations in (3) specify that more low (high) risks consumers are attracted if selection activities $s_{iL}$ ($s_{iH}$) of insurer $i$ are higher than the average amount of selection activities of the other insurers. The amount of consumers an insurer will attract depends on the parameters $\delta_k$ and $\eta$. For example, a low $\delta_k$ or $\eta$ implies that consumers are insensitive to these type of activities. $\delta_L$ and $\delta_H$ may differ. Again, different type of risks may react differently on similar type of selection activities. Everything else equal, high risks are likely to react more sensitively than low risks to health care activities arranged by insurers. We assume $0<\eta<1$, which indicates that selection activities feature decreasing returns to scale.

Note that premiums are modelled through a linear relationship, whereas changes in selection activities are modelled as decreasing returns to scale. This choice has important consequences for the final outcomes. It is made to keep the model analytically tractable and, moreover, yields a unique Nash equilibrium.

### 3.3. The role of the sponsor

As we have discussed in section two, the sponsor collects solidarity contributions from the consumers. Considering the viewpoint of the insurers, they will receive prospective and retrospective payments from the sponsor.

#### 3.3.1. Prospective payments

The prospective payments, $b_i$, are modelled as follows

$$b_i = b_{iL} + b_{iH} = \mu_L x_{iL} + \mu_H x_{iH} , \tag{4}$$

for $i = 1,...,N$, $0< \mu_L < \mu_H$ .

For each enrollee insurers receive a lump sum subsidy depending on its risk characteristics. For a low risk enrollee an insurer receives $\mu_L$ and for a high risk enrollee $\mu_H$. We assume that subsidies for high risks are higher than for low risks.

A description of the retrospective payments is more complicated. Assume that (e.g. after a year) the realised total health care costs for each insurer is known. Then, first retrospective equalization, and next, retrospective compensation takes place by the

study of Beeson Royalty and Solomon (1999), who found lower price elasticities for specific groups of people, such as older workers, employees with longer tenure, and enrollees who insure a family member with a chronical medical condition.

sponsor. The sponsor calculates the retrospective payments by taking into account the realised health care costs and the prospective payments.

### 3.3.2. Retrospective equalization

If the realised health care costs for insurer *i* are

$$c_i \ = \ c_{iL} \ + \ c_{iH} \ = \ z_{iL} x_{iL} \ + \ z_{iH} x_{iH} \ , \tag{5}$$

then, after subtracting the prospective payments, each insurer faces health care costs of $c_i$ - $b_i$. In the sequel we will call this difference between medical costs and prospective payments net costs for the insurer. After retrospective equalization insurer *i* receives from the sponsor

$$b_i^* \ = \ b_i \ + \ \nu \, x_i \left( \frac{(c_i - b_i)}{x_i} - \frac{\sum_{j=1}^{N} (c_j - b_j)}{\sum_{j=1}^{N} x_j} \right) \ , \tag{6}$$

for $i$ =1,...,$N$ and $0 \leq \nu \leq 1$.

The first term, $b_i$ , are the prospective payments from the sponsor and the second term represents the amount due to retrospective equalization. The first term between brackets reflects the average net costs of the *i*th insurer per enrollee. The last term between brackets denotes average net costs per enrollee, where the average is taken over all insurers. Since the average is taken over the sum of low and high risks together the specification allows for cross-subsidies between low and high risks. If insurer *i*'s average net costs are above total average net costs, the *i*th insurer will receive from the sponsor this difference times his number of enrollees times the equalization parameter $\nu$. If insurers net costs are below average the insurer has to pay a similar amount to the sponsor. Retrospective equalization is thus a zero-sum activity: the payments made by low-cost insurers compensate exactly the subsidies to high-cost insurers. Since retrospective equalization is budget neutral for the sponsor it is often called a form of risk sharing among insurers.

### 3.3.3. Retrospective compensation

After retrospective equalization, retrospective compensation is carried out by the sponsor:

$$b_i^{**} \ = \ b_i^* + \theta (c_i - b_i^*) \ , \tag{7}$$

for $i = 1,...,N$ and $0 \le \theta \le 1$.

According to this expression, insurer *i* receives (pays) an additional subsidy if, after equalization, the operating profit is negative (positive). In this formula the sponsor payments to the insurer, after equalization, are enlarged (reduced) with a fraction of the realised losses (profits). This fraction $0 \le \theta \le 1$ is determined, together with the equalization parameter, $\nu$, every year by the sponsor. In contrast to retrospective equalization, retrospective compensation is not budget neutral.

### 3.3.4. Solidarity premiums

The sponsor finances its payments by levying premiums that are related to income of consumers, but unrelated to their medical risk. As our model does not distinguish consumers, with respect to income, this amounts to a flat solidarity premium by consumers. The premium per consumer that balances the budget of the government is as follows: $1/x \sum_{i=1}^{N} b_i^{**}$.

Total consumer premiums sum premiums to the sponsor and premiums paid to the insurer. Hence, the expression for total premiums per consumer reads as follows:

$$p_i^c = p_i + \frac{1}{x}\sum_{i=1}^{N} b_i^{**} \tag{8}$$

Note that the premium $p_i^c$ differs for consumers who bought their health insurance from different insurance companies due to the fact that community-rated premiums may differ. The premium $p_i^c$ is considered a measure of consumer welfare.

## 3.4. The profit and objective function of the insurer

We define the profit function of insurer *i* as follows:

$$W_i = b_i^{**} + p_i x_i - (z_{iL} x_{iL} + z_{iH} x_{iH}) - (s_{iL} + s_{iH}) - e_i \tag{9}$$

for $i=1,...,N$.

This profit function, $W_i$, subtracts total costs from total returns. Negative entries are realised health care costs of the low and high risks ($z_{iL} x_{iL} + z_{iH} x_{iH}$), the investment costs in selection for low and high risks ($s_{iL} + s_{iH}$) and the investment costs in efficiency activities ($e_i$). Positive entries are the subsidies from the sponsor $b^{**}_i$ and community-rated premiums collected directly from the consumer ($p_i x_i$).

Although making profits from medical insurance is an important objective, we must not ignore the fact that many insurance companies operated for a long time under full cost reimbursement.[11] An insurance company may also want to attract consumers to make (additional) profits on supplementary, or other type of, insurance. Cross-selling may be an important issue here. Attracting more consumers means also more status (and thus bargaining power) by becoming a big player in the field of health insurance. More consumers makes it also possible for insurers to make larger investments in risk-share capital. Furthermore, a risk averse insurer may want to attract more enrollees to reduce the variation of its expected health care costs.

We model this latter aspect by defining an objective function, $V_i$ , in which besides profits the insurer maximizes the number of enrollees. The maximization problem is now defined as:

$$Max_{e_i,\, s_{iL},\, s_{iH},\, p_i} \quad V_i \;=\; W_i \;+\; \varphi\, x_i \,, \tag{10}$$

where $0 \le \varphi$. Note that if $\varphi = 0$, then insurers fully maximize their profits from medical insurance.

We assume that insurers act rationally and optimize the objective function with respect to the four instruments, assuming a Nash equilibrium between insurers.

### 3.5. Solution mechanism

Technically the optimization problem results in $N$ objective functions and $4N$ instruments to optimize. We compute the first-order conditions by substituting equation (1)-(7) in the objective function $V_i$ , $i$=1,...,$N$. This yields $V_i$ ($e_1$,...,$e_N$, $s_{1L}$, ...,$s_{NL}$, $s_{1H}$, ...,$s_{NH}$ , $p_1$,..., $p_N$). Next, we compute for $i$=1,...,$N$: $\partial V_i / \partial e_i = 0$, $\partial V_i / \partial s_{iL} = 0$, $\partial V_i / \partial s_{iH} = 0$, $\partial V_i / \partial p_i = 0$, given the actions of the other insurers. This yields a closed system of 4 times $N$ equations, which can be solved. This will result in a Nash equilibrium. To obtain an analytical solution we assume symmetry between insurers. The symmetry argument implies that $x^0_{iL} := x_L / N$ and $x^0_{iH} := x_H / N$. Furthermore, in equilibrium the optimal outcomes $\hat{e}_i, \hat{s}_{iL}, \hat{s}_{iH}, \hat{p}_i$ are equal for all insurers $i$=1,...,$N$. Inspection of the

[11] Besides profits, other objectives may have played (and may still play) an important role. There may be a social objective such as aiming at a desired level of care. See Schut (1995), for a historical overview of the Dutch insurance market.

corresponding second-order conditions (not shown here) reveals that the system of first order conditions indeed depicts a maximum.[12]

## 4. Explaining the Nash equilibrium of the model

In this section we describe and explain the Nash equilibrium of the problem. In the following sections we will describe the optimal Nash rule for each instrument in turn and discuss derived terms such as insurer profits and total consumer premiums.

### 4.1. Optimal efficiency rules

The optimal investment rule for efficiency activities by insurers reads as follows:

$$\hat{e} := \hat{e}_1 = \dots = \hat{e}_N = \left( \frac{\gamma}{N}(\alpha_H x_H + \alpha_L x_L)(1-\theta)(1-\nu(1-\frac{1}{N}-\frac{\beta}{N})) \right)^{\frac{1}{1-\gamma}} \tag{11}$$

From this equation, it follows that optimal efficiency investments ($\hat{e}$) decrease with $N$. The larger the number of insurers, the fewer enrollees are available for each insurer. Hence, the smaller are the gains from pursuing medical efficiency and the fewer efficiency activities will be carried out. Efficiency investments are increasing in $\alpha_H x_H + \alpha_L x_L$ and $\gamma$. Like before, the more enrollees $x_L$ (or $x_H$), the higher the profits from an extra unit of efficiency. A similar reasoning holds for $\alpha_L$ (or $\alpha_H$) and $\gamma$. The higher the rate of retrospective compensation $\theta$, the smaller the gains that an individual insurer receives from pursuing medical efficiency. This results in lower efficiency activities by the insurer.

The term $(1-\nu(1-1/N-\beta/N))$ is of special interest.[13] Since, in a Nash equilibrium we assume the actions of the other insurers as given, an insurer knows that possible losses will be covered by other insurers. The higher the equalization parameter $\nu$ the more an insurer will be covered by other insurers and thus the less risk an insurer bares.

[12] Here we do not go deeper into the analytical computations which are rather lengthy and cumbersome. One can show that the Nash equilibrium produces a unique internal solution for a given range of parameter values, which are obtained by computing the second order conditions of the maximization problem. These computations can be retrieved by the author upon request.

[13] The fact that the impact of equalization equals $(1-\nu(1-1/N-\beta/N))$ can be understood by rewriting the term between brackets in equation (6). In contrast with retrospective compensation, retrospective equalization always yields some marginal return. Compare the two exteme cases: {$\theta$=1, $\nu$=0} and {$\theta$=0, $\nu$=1} In the first case marginal returns from investing in efficiency is zero, whereas in the latter marginal returns depend on $1/N+\beta/N$.

Equalization implies in most cases cost solidarity across insurers which is a glaring contrast with competition. A remarkable result is that the more spillover effects between insurers are in place (the higher β), the lesser equalization has an impact on efficiency activities. The reason is that if an insurer undertakes efficiency activities then all insurers will profit from these activities. Due to these profits less money has to be split up across insurers. In the extreme case β=*N*-1 efficiency gains are equal for every insurer and, thus, the impact of retrospective equalization is zero.

Substituting the optimal efficiency rules in equation (2) generates realised health care costs:

$$\hat{z}_k := \hat{z}_{i,k} = z_k^0 - \alpha_k (1 + \beta) \hat{e}^{\gamma} , \tag{12}$$

for $k$=$H$,$L$ and $i$ =1,...,$N$.

We define also net realised health care costs per risk type. These costs subtract the prospective payments of the sponsor from the realised health care costs:

$$\bar{z}_k := \hat{z}_k - \mu_k, \tag{13}$$

for $k$=$H$,$L$.

## 4.2. Optimal selection rules

In the Nash equilibrium, the following selection rules, for low and high risks, are optimal

$$\hat{s}_L := \hat{s}_{iL} = \left( \delta_L \cdot \frac{x_L x_H \eta}{\zeta_L x_L + \zeta_H x_H} \cdot \left( \frac{1}{x_H} \cdot \frac{x}{N} + (1-\theta)(1-\nu)\,\zeta_H (\bar{z}_H - \bar{z}_L) \right) \right)^{\frac{1}{1-\eta}} \tag{14}$$

$$\hat{s}_H := \hat{s}_{iH} = \left( \delta_H \cdot \frac{x_L x_H \eta}{\zeta_L x_L + \zeta_H x_H} \cdot \left( \frac{1}{x_L} \cdot \frac{x}{N} + (1-\theta)(1-\nu)\,\zeta_L (\bar{z}_L - \bar{z}_H) \right) \right)^{\frac{1}{1-\eta}} \tag{15}$$

for $i$ =1,...,$N$.

Although we make a distinction between selection activities for low and high risks, this does not mean that general selection activities, intended for low and high risks simultaneously and which do not discriminate among risk types, are not incorporated in the model.[14] The costs of general selection activities may fall for a part under $\hat{s}_L$ and for the remaining part under $\hat{s}_H$. How large these parts are may depend on the distribution of high ($x_H$ ) and low risks ($x_L$) in the population. The first terms between the inside brackets ($1/x_H$ and $1/x_L$) in equation (14) and (15) may determine the size of the two parts.

We start with discussing the case $\zeta_L = \zeta_H$ , $\delta_L = \delta_H$ and $\bar{z}_H = \bar{z}_L$. In this case high and low risks are similar for an insurer and we assume that an insurer will undertake only general selection activities; the costs of general selection activities are $\hat{s}_L + \hat{s}_H$. As discussed before, $\hat{s}_L$ and $\hat{s}_H$ may be different because the distribution of high and low risks may differ. General selection activities will increase if the scale of the insurer increases ($x/N$ has a positive effect on both $\hat{s}_L$ and $\hat{s}_H$ in (14) and (15)). The size of general selection activities is also related to the substitution effect, implied by equation (3), between investing in selection or attracting consumers by changing the premium. The more consumers are less sensitive to price (the lower $\zeta_L x_L + \zeta_H x_H$) and the more sensitive they are towards selection activities (higher $\eta$ and/or higher $\delta_L, \delta_H$ simultaneously) then general selection activities are increasing as well.

Risk-type specific selection activities, *i.e.* that discriminate among risk types, such as exploring marketing activities or concentrate advertisements only on the favourable risks (also called cream skimming or cherry picking) come into the picture if one of the equalities $\zeta_L = \zeta_H, \delta_L = \delta_H$, and $\bar{z}_H = \bar{z}_L$ does not hold. We consider now one inequality at a time (and for reasons of exposition assume that the other two equalities still hold).

If $\zeta_L \neq \zeta_H$ then low (high) risks are more price elastic than high (low) risks. However, as long as both risks have equal net costs ($\bar{z}_H = \bar{z}_L$) there is no need for the insurer to undertake risk-type specific selection activities. General selection activities however may change due to the substitution effect, as described above.

If $\delta_L \neq \delta_H$ then low (high) risks are more sensitive to selection activities than high (low) risks. It is therefore profitable for the insurer to concentrate more on the low (high) risks. This aspect distinguishes risk-type specific selection activities from general selection activities. The larger the difference between $\delta_L$ and $\delta_H$ is, the more risk type specific selection activities will take place (and the less general selection activities will take place).[15]

[14] Examples of general selection activities are the placement of a bilboard with only the name of the insurance company or an improvement of services of the insurance company.

[15] We discuss here only directions and do not answer the question how the insurer divides $\hat{s}_L + \hat{s}_H$ between general selection activities and risk type specific selection activities. This depends also on factors outside the model.

If $\bar{z}_H \neq \bar{z}_L$ then we have a situation where the sponsor performs *imperfect risk adjustment.* First we discuss, however, the case of $\bar{z}_H = \bar{z}_L$, *i.e.* the sponsor performs *perfect risk adjustment.* In that case in equation (14) and (15), the last term equals zero. Perfect risk adjustment by the sponsor implies that prospective payments price the heterogeneity of different risks perfectly. Since net health care costs are exactly equal for high and low risks, insurers are indifferent with respect to medical risk to insuring one type of risk or the other. Thus, discriminating between the two types of risk with respect to medical risk makes no sense. Note however that $\bar{z}_H = \{\hat{z}_H - \mu_H\}$ and $\bar{z}_L = \{\hat{z}_L - \mu_L\}$ may be different from zero. Hence, sponsor subsidies may deviate a lot from health care costs, as long as the deviations for the two types of risks are equal.

We now return to the case $\bar{z}_H \neq \bar{z}_L$. In that case net health care costs for high (low) risks are higher than for low (high) risks. The larger this difference the more $\hat{s}_L$ and $\hat{s}_H$ will deviate from each other and thus the more risk-type specific selection activities will occur.[16] Note that the size of the price elasticities $\zeta_L$ and $\zeta_H$ plays a role as well in the ultimate size of risk-type specific selection activities. For example consider the case where the low risks are the good risks ($\bar{z}_H > \bar{z}_L$ ). If now the price elasticity of high risks increases, it becomes even more attractive for an insurer to concentrate on low risks. Besides the fact that low risks are more profitable, for medical reasons, they will also change less quickly from insurer once they are contracted.

Note that our selection equations perfectly show how complicated it will be in practice to isolate the various reasons why insurance companies undertake, for example, selective advertisement strategies. Besides the size of (unpriced) risk heterogeneity many other aspects play an important role, such as the differences in size of the high an low risk population ($x_L$ and $x_H$), differences in price sensitivity ($\zeta_H$ and $\zeta_L$), differences in selection sensitivity issues ($\delta_H$ and $\delta_L$) and the type of payment system of the sponsor.[17]

Besides perfect risk adjustment, there are several other policy options to reduce selection activities for reasons of unpriced risk heterogeneity. Both, increasing

[16] Theoretically, two types of bad risks can be distinguished. Risks which are simply bad in the sense that they are less profitable than high risks and risks which are really bad in the sense that they imply losses for the insurer. In the latter case, the optimal selection rule towards bad risks may even become negative. Note, however, that this latter possibility does not satisfy the second order conditions of the Nash equilibrium.

[17] For example, if an insurance company chooses to place a bilboard on the the street with young healthy people on the background then this choice may be due to the fact that young healthy people are good risks. However, it may also be due to the fact that young healthy people react more sensitively to these advertisements. The size of risk-type specific selection activities (or the number of bilboards on the streets) depends also on the price elasticities of these people and of the type of prospective and retrospective payments the government applies.

retrospective equalization and compensation reduces the risk-type specific selection strategies for insurance companies. This observation follows directly from equation (14) and (15). A higher θ or ν diminishes the effect of imperfect risk adjustment by the sponsor. Retrospective equalization equalizes risks across insurers and retrospective compensation transfers part of the risk to the sponsor (and thus both options decrease the marginal returns from selection). However, both policies may imply a mixed blessing. As follows from equation (11), high values for θ and ν will diminish efficiency activities by insurers. This marks the trade-off efficiency in production versus selection, as indicated by Newhouse (1996).

### 4.3. The optimal rule for the community-rated premium.

The optimal rule for the community-rated premium is

$$\hat{p} = \hat{p}_i = \frac{x}{N(\zeta_L x_L + \zeta_H x_H)} - \varphi + \frac{1-\theta}{\zeta_L x_L + \zeta_H x_H}\left(\zeta_L x_L \bar{z}_L + \zeta_H x_H \bar{z}_H + \frac{\nu x_L x_H}{x}(\zeta_H - \zeta_L)(\bar{z}_L - \bar{z}_H)\right) \tag{16}$$

for $i$ =1,...,$N$.

In (16) we distinguish the competition effect or mark-up $x/N(\zeta_L x_L + \zeta_H x_H) - \varphi$ and costs wich are related to medical costs. The latter are represented in the third term in (16).

More sensitive consumers (higher $\zeta_L$ and $\zeta_H$ ) will intensify competition among insurers which will lead to lower premiums. Scale effects play a role as well. If the number of insurers increases, more competition will prevail and premiums will fall. If insurers obtain more money by activities not related to medical care (see 3.4.), φ may play a substantial role. If these activities are very profitable, and insurers receive for each enrollee more money outside the medical insurance market, then the mark-up may even become negative.

The third term in (16) is related to the supply of health care. The more retrospective compensation takes place (higher θ), the higher are the retrospective payments the insurer will receive from the sponsor and, thus, the lower the community-rated premium needs to be. The costs are, furthermore, a weighted average of the net health care costs of low and high risks, where the weight are proportional to $\zeta_H x_H$ and $\zeta_L x_L$. To understand this, consider for a moment the case where high risks are more price sensitive to premium changes than low risks ($\zeta_H > \zeta_L$). Thus, a marginal reduction in the community-rated premium will attract more high risks. Combine this with a case of imperfect risk-adjustment where high risks are the bad risks ($\bar{z}_H > \bar{z}_L$), then a marginal reduction of the community-rated premium will be, relatively, less profitable for an insurer. As a consequence the optimal community-rated premium will be relatively high in the Nash equilibrium. Finally, the last term in the equation turns up because retrospective equalization allows for cross subsidies between low and high risks. It depends on the positiveness of the retrospective equalization parameter ν. If we consider the similar case as discussed before, then the community-rated premiums will be adjusted downward since the relatively higher costs of the high risk population will, partly, be equalized by other insurers. The main point of allowing cross-subsidies between high and low risks is that it smoothes differences in total health care costs across insurers.

## 4.4. Insurer profits

The profits of the insurer can be obtained by substituting the optimal outcomes (11)-(16) in the profit function (9). This yields $\hat{W} = \hat{W}_i =$

$$= \frac{x}{N}\left( \frac{x}{(\zeta_L x_L + \zeta_H x_H)\,N} - \varphi + \frac{(1-\theta)(1-\nu)(\zeta_H - \zeta_L)(\bar{z}_H - \bar{z}_L)x_L x_H}{x(\zeta_L x_L + \zeta_H x_H)} \right) - (\hat{s}_L + \hat{s}_H) - \qquad (17)$$

for $i = 1,...,N$.

The amount of insurer profits depends on several effects. First, the number of enrollees times the mark-up yields a part of the profits. As a result, profits rise if the number of enrollees per insurer rises. Also, if consumers are more price sensitive (high $\zeta_L$ and $\zeta_H$ in the first term) then competition among insurer intensifies and profits will decline. Second, profits are affected if both hold: $\zeta_L \neq \zeta_H$ and risk adjustment is imperfect. This is reflected in the third term between brackets. The ultimate effect of this term depends upon a combination of different effects. For example, consider the case where $\bar{z}_H > \bar{z}_L$ and $\zeta_H < \zeta_L$. In that case insurers are less eager to adjust premiums downwards since lowering premiums will attract relatively more the loss-making high risks. This effect leads to profits and becomes larger if asymmetry between the two types of risks (in net costs, price sensitivity or population) increases.

Competition assures that selection investments return as costs in the insurer profit function. Symmetry implies that there are no specific gains for an individual insurer resulting from selection. In the model selection is a zero-sum game; all investments finally result in a public loss.

In the profit function efficiency plays a similar role as selection. Efficiency costs represent investment costs which have to be paid by the insurer. Opposite to selection, however, efficiency lowers total health care costs. In general, the reduction in health care costs does not show up in the profits but is fully reflected in the premium and thus all the proceeds of efficiency go to the consumers.[18]

To get a better understanding of the model, it is also interesting to consider a simple case where retrospective equalization and compensation is absent and $\zeta = \zeta_L = \zeta_H$ . This leads to the following expressions for the premium and profits:

$$\hat{p} = \hat{p}_i = \frac{1}{N\zeta} - \varphi + \frac{x_L}{x}\hat{z}_L + \frac{x_H}{x}\hat{z}_H \tag{18}$$

$$\hat{W} = \hat{W}_i = \frac{x^T}{\zeta N^2} - (\hat{s}_L + \hat{s}_H) - \hat{e} \tag{19}$$

for $i = 1,...,N$.

This case shows more clearly a general aspect of the model. Since all insurers are symmetric, the premium of each consumer depends, besides a mark-up (the competition effect), on a weighted average of the health care costs for low and high risks. The profits for each insurer are simply the mark up per consumer times his number of enrollees minus the costs for investments in efficiency and selection. If there are many insurers and $\zeta$ is relatively high then the mark up will be low.

## 4.5. Total consumer premiums

Total consumer premiums are

$$\hat{p}^c := \hat{p}_i^c = \hat{p} + \frac{1}{x}\sum_{i=1}^{N} b_i^{**} = \hat{p} + \frac{x_L}{x}(\mu_L + \theta\bar{z}_L) + \frac{x_H}{x}(\mu_H + \theta\bar{z}_H) \tag{20}$$

for $i = 1,...,N$.

Each consumer has to pay a community-rated premium and a solidarity contribution to the sponsor.

[18] This is the case if risk-adjustment is perfect or $\zeta_L = \zeta_H$. Theoretically one can argue that this is a rather extreme effect of the model. Competition assures that insurers receive no gains from efficiency. As selection, efficiency activities are for the insurer 'necessary' losses to survive in the market. The reason for this outcome lies in our assumption in equation (3); premiums are modelled linear and selection activities are modelled as decreasing returns to scale.

The first term reflects the community-rated premium. The second and third term reflect the prospective and retrospective payments to the sponsor. First, each consumer contributes a proportional share, related to the number of high and low risks in the population, of the prospective payments, $\mu_L$ and $\mu_H$, to the sponsor. A similar share, depending on the amount of retrospective compensation, is related with respect to net health care costs.

## 5. Asymmetry, information and dynamics

The outcomes in the model are based on some important assumptions, such as symmetry across insurers and full information of insurers. Also, the model is static; it does not consider dynamic aspects. In the following three subsections we will explore some of the possible consequences if we release these assumptions.

### 5.1. Retrospective equalization, compensation and asymmetry

The symmetric outcomes in the Nash equilibrium show that enlarging equalization (increasing $\nu$) or enlarging retrospective compensation (increasing $\theta$) lowers efficiency activities and, thus, raises health care costs and, thus, (individual and total) premiums. The impact of retrospective compensation on total selection activities is fairly neutral. Depending on the crookedness of the risk adjustment scheme, decreasing $\theta$ in equation (14) and (15) implies that risk-type specific selection activities on low risks and high risks alter, albeit in a different direction. The symmetry assumption, however, may leave some important practical issues of the picture.

In the following example we will highlight the asymmetric consequences of retrospective equalization and compensation. Consider a simple case where each insurer $i$ faces an equal number of enrollees. Assume that insurer $i$ faces total costs, $\hat{z}_i$, where the $\hat{z}_i$'s are mutually independent distributed with mean $E(\hat{z}_i)=\hat{z}_i$ and variance $VAR(\hat{z}_i)= \sigma^2$, for $i$=1,...,$N$. The mean costs across insurers differ. This reflects the idea that insurers behave asymmetrically, due to for example different initial conditions, and thus may face different health care costs. In the sequel we will only concentrate on the consequences of retrospective equalization and compensation. Therefore, we substitute the distributions in equation (6) and (7). This yields, after straightforward calculation, that health care costs minus total sponsor subsidies for insurer $i$, are distributed with:

$$\begin{aligned} E(z_i - b_i^{**}) &= (1-\theta)(z_i - \nu(z_i - \frac{1}{N}\sum_{j=1}^{N} z_j), \\ VAR(z_i - b_i^{**}) &= \sigma^2 \left[ (1-\theta)^2((1-\nu+\frac{\nu}{N})^2 + \frac{\nu^2(N-1)}{N^2}) \right] \end{aligned} \tag{18}$$

for $i$=1,...,$N$.

From the Nash equilibrium of our model follows that health care costs for insurer $i$ diminish with (1-θ)(1-ν-ν/$N$). Since marginal efficiency activities are decreasing in the intensity of behaviour, this results in lower efficiency activities if retrospective equalization and compensation increase. Now compare the set of mean costs $\{z_i, i=1,...,N\}$ and the set of mean costs of insurers in (18). The asymmetric aspect of the example, however, shows that the set of mean costs in (18) becomes more dense. Since health care costs of an insurer are likely to be closely related to the height of the premium we expect that an increasing density in health care costs will be correlated with an increasing density in premiums. This effect by itself is likely to weaken competition across insurers. Another element is related to the variance. Reducing retrospective equalization and compensation by the sponsor increases the variances in (18) and thus raises the risk that insurers bare.

Average health care costs for the sponsor are distributed as:

$$E\left(\frac{1}{N}\sum_{i=1}^{N} b_i^{**}\right) = \frac{\theta}{N}\sum_{i=1}^{N} z_i \quad \text{and} \quad VAR\left(\frac{1}{N}\sum_{i=1}^{N} b_i^{**}\right) = \frac{\sigma^2}{N}, \tag{19}$$

for $i$=1,...,$N$. Note that since retrospective equalization is budget neutral, only retrospective compensation has an impact on the mean costs that the sponsor bares. The big difference between the variance in (18) and (19) is that the latter variance is divided by $N$. Transposing risk from the sponsor to the insurers enlarges the risk for the individual (risk averse) insurer (consumer). Indeed the sponsor pools the whole population, whereas each individual insurer pools only a part of the population.

To conclude, the effect of reducing retrospective equalization and compensation on premiums may be at least fourfold. First, reducing retrospective compensation implies that part of the health care costs passes on to the insurer which will increase the community-rated premiums. In the case of imperfect risk adjustment and different price sensitive consumers this may exert an upward effect on total consumer premiums. Second, as the Nash equilibrium also shows, it may exert a downward effect on premiums since more efficiency activities may take place. Third, the variability of the community-rated premium across insurers may increase. This observation follows from the fact that reducing retrospective equalization and compensation makes the distribution of mean health care costs of insurers less dense, which may result in increasing variability in the community-rated premium.[19] Fourth, the consequences of a rising risk by insurers (through the increasing variances) may be passed on to consumers by adding a risk premium to the community-rated premium.

[19] Data in the Netherlands from recent years shows that a reduction in retrospective equalization and compensation is correlated with an increasing variability in community-rated premiums.

## 5.2. Uncertainty and the role of information

The model is on the basis of full information of the insurer over the consumers, the sponsor and the health care providers. The assumption of full information is, strictly speaking, not necessary. The model also applies to the case where insurers, retrospectively, make an assessment of the situation, using all relevant information, about the consumers, the sponsor and the health care providers. On the basis of this assessment insurers will select their optimal rules as discussed in section 4. Ex-post, however, it may turn out that an insurer may have misjudged the situation. This misjudgement may be caused by uncertainty or incomplete information about many factors, which include misjudgement about the behaviour of consumers, health care providers and sponsor.

Whereas the consequences of possible failures in risk adjustment are discussed extensively in the literature (Newhouse (1996), Van de Ven and Ellis (2000)), the model presents another argument about the information asymmetry between the insurer and the sponsor. This is the argument of timing. The insurer has time to consider to invest in efficiency activities after the sponsor has announced its rules for the prospective and retrospective payments. Perfect risk adjustment by the sponsor now implies not only full information at time of the announcement, but also full (forward looking) information about the activities of the insurer in the future. In the following equation (which follows directly from (10)) we present this aspect more sharply:

$$\{\hat{z}_H - \mu_H\} = \{\hat{z}_L - \mu_L\} \quad \Rightarrow$$
$$z_{0L} - \alpha_L\,(1+\beta)\,\hat{e}^{\gamma} - \mu_L = z_{0H} - \alpha_H\,(1+\beta)\,\hat{e}^{\gamma} - \mu_H \tag{20}$$

In order to perform perfect risk adjustment the sponsor should have knowledge, not only about $\gamma$, $\alpha_L$ and $\alpha_H$ , but also on $\hat{e}$. While in practice the sponsor may have some knowledge about the institutional characteristics of the insurer (the parameters $\gamma$, $\alpha_L$ and $\alpha_H$) obtained from the past, the information about possible efficiency activities ($\hat{e}$) will always be uncertain.

In the model insurers can also discriminate perfectly between high risk and low risk consumers. However, elements of imperfect information of the insurer can be handled within the model as well. If the insurer fails to identify consumers correctly, for example due to adverse selection, this can be modelled by adding uncertainty to the model.[20]

One problem which arises in practice is that it is often difficult to distinguish between a situation where an insurer faces uncertainty or a situation where an insurer produces inefficiently. While retrospective equalization and compensation are primarily

[20] This is a challenge for future research. In Cutler and Reber (1998) the trade-off between competition and adverse selection is explored more deeply.

designed by the sponsor to subsidize the insurer in the former case, the sponsor also subsidizes the insurer in the latter case. In the Netherlands there is now a growing concern for this aspect which results in adjusting the equalization and compensation parameter downwards.[21] Moreover, in the coming years, realised healthcare consumption or costs in the previous year will be included as a risk adjuster in the formula that defines prospective payments. This new policy has advantages. Since previous year consumption, or healthcare costs, is highly correlated with this year's costs, this may imply a better match between prospective payments to insurers and their risk characteristics. Inclusion of this variable may thus help to reduce the variability of healthcare costs and, as a result, decrease the incentives for risk-selection behaviour. However, this new policy may also reduce the incentives for efficiency behaviour. In particular, the return on efficiency investments is low if insurers expect prospective payments next year to decrease should they succeed in lowering healthcare costs this year. This problem, however, is in essence a dynamic problem. This is the topic of our next subsection.

### 5.3. Dynamics

The inclusion of dynamics in the underlying model may be important for several reasons. Activities by insurers such as selection may have long term consequences. For example, once bad risks enter an insurance pool they are likely to stay there for a long time and, thus, may reduce profits for an insurance company over a longer period. This is in contrast with general selection activities such as advertising where the long term consequences are less clear. Long term consequences are likely to play a role in the minds of insurers if they have to make a judgement between the consequences of putting one unit of money into risk-type specific selection activities or into general selection activities. Insurers who ignore this aspect may be saddled up with a pool of bad risks and, finally, may even end up in bankruptcy.[22]

Another interesting aspect of dynamics is the yearly "game" between the sponsor and the insurers.

Although insurers may have an information advantage when selecting their efficiency activities after the sponsor has announced its new payment system for the coming year, a problem for the insurers arises when the sponsor updates the risk formula every year.

[21] This does only hold for those categories where these effects are most likely to take place. For example, fixed costs of hospitals, such as capital costs, are often specified in long term contracts and thus difficult to influence by the insurer. Therefore, for fixed costs in hospitals still a retrospective compensation parameter of 0.95 applies in the Netherlands.

[22] An example here is the Swiss health insurance market (see Beck (1999)).

Insurers will realise that efficiency gains of a previous year will become incorporated in the sponsor's subsidies of next year. The underlying static model is not adequate to foresee how the sponsor and insurers have to react optimally to this dynamic issue. The most likely result is that insurers will diminish their efficiency activities in the current year.

In a dynamic competitive environment, efficiency activities are only profitable if individual insurer gains (investment profits minus investments costs) outweigh the gains of other insurers. In the underlying Nash equilibrium, insurers invest in efficiency activities if their gains outweigh their costs. However, due to the existence of efficiency spill-overs in the health insurance markets it may happen in our model that, although a single insurer makes profits from his efficiency activities, other insurers make more profits. The insurers who obtain more profits may use these profits in the next period to obtain a competitive advantage (*e.g.*by lowering the community-rated premiums). Thus, while in our static model insurers are still willing to invest in efficiency activities they may not do so in a dynamic model. This especially will be the case if spill-overs are large. A phenomenon which is typically for an environment where consumers have free rein in choosing their healthcare providers. To analyse these aspect more profoundly, a dynamic model would be more appropriate here, but our first guess is that in a dynamic context investments in efficiency in such an environment will be lower.

## 6. Conclusions

This paper has explored the economic effects of financing schemes for insurer activities. The economic effects concern primarily health care institutions where consumers have a large freedom of choice with respect to health care providers and insurers. These type of situations apply for many countries in Europe and, by way of an example, we consider in the paper the sickness funds in the Netherlands. The situation in the Netherlands is particularly interesting since currently the government is implementing health care policies which aim at increasing the risks of the insurers by cutting down on various reimbursement schemes. The idea behind these policies is to create a more efficient delivery of health care services by intensifying competition among health care insurers.

We show that reimbursement schemes that foster the efficient delivery of medical services may exert various effects on the premiums across insurers. The model shows that premiums may fall, due to the efficient delivery of care. We discussed, however, also the possibility that premiums may rise, since increasing the risk of individual insurers may force them to include a risk premium. Effects on the premium related to the heterogeneity of the population are found to be mixed. There may be an upward effect on the premium if bad risks are more price-sensitive than good risks. If we release the symmetry assumption in our model then we expect that the variability of community-rated premiums is likely to increase as well, since cutting down on the reimbursement schemes emphasizes the differences in health care costs across insurers.

Another important finding of the paper is that transposing risk from the sponsor to the insurer may increase risk-type specific selection activities of the insurer since bad risks will be less profitable than good risks. The magnitude of this effect depends on the quality of the risk adjustment scheme and the degree of retrospective reimbursements the government applies. Since, in practice, we are a long way from having a perfect risk adjustment scheme we should be careful not to underplay the possible magnitude of these effects. If risk adjustment is not perfect, risk-selection can be diminished by introducing a certain amount of retrospective reimbursement. Both, retrospective equalization en compensation will work. However, the higher the amount of retrospective reimbursement the lower will be the incentives for efficiency behaviour of insurers. The magnitude of risk-selection will also depend on the profitability of pursuing efficient delivery of medical services. In situations where insurers have contracts with almost all health care providers, a single insurer may find it very difficult to undertake efficiency activities which only reduce its own health care costs while leaving health care costs of his competitors untouched. Undertaking efficiency activities in a competitive environment makes no sense if health care costs of other insurers would be reduced in an equal (or higher) amount.

Our model shows that more competition is likely to reduce total health care costs, premiums and insurer profits. Two opposite effects are related to the number of insurers. One effect is that a larger number of insurers amounts to fewer efficiency and selection activities, which results in higher community-rated premiums. The opposite effect is that a larger number of insurers intensifies competition which exerts a downward effect on the premium. More efficiency will take place but all the proceeds of efficiency will end up with the consumers. Thus, in such a world, insurers will face higher efficiency and selection costs and lower profits.

Selection in the health care literature often pays only attention to risk-selection on the basis of medical costs. However, our model studies also the consequences of more general forms of selection activities which are unrelated to medical costs, such as advertising or improving insurer services. Our model shows that the amount of general selection activities increases if the scale of the insurer increases, if consumers become more sensitive to these types of selection activities, and if consumers become less sensitive to price.

Interesting are also the results of the model related to the (price) sensitivity of consumers. More price sensitive consumers will intensify competition among insurers and thereby exert a downward pressure on the premium. Since insurers will more concentrate on prices (premiums) they will reduce also the amount of selection activities (due to the substitution effect between prices and selection). A policy which increases (price) sensitivity of consumers appears to be a sensible option. A word of caution is, however, suggested here by the model as well. Policies which result in an overall increase in the price sensitivity of all consumers are beneficial. However, policies that only reach particular groups of the population may not have a desirable effect. The model shows that if bad risks are more price sensitive than good risks, then a sluggishness in adjusting premiums downward may be the undesired result. Insurers will simply not be very eager to cut premiums since it will attract relatively more bad risks. More empirical research is necessary to measure the importance of asymmetric price sensitivity and how price elasticities are related to risk-selection activities by insurers.

Of course many important health care issues fall outside the scope of the theoretical model. We mention two important issues. First, asymmetric information. Asymmetry lurks everywhere, among insurers and between insurers and other players in the health care field. Second, dynamic issues. Dynamic issues are extremely important in health care. The static model in this paper may be used as a starting point for constructing a dynamic model.

## 7. References

Altman, D., D.M. Cutler and R.J. Zeckhauser (2000), Enrollee mix, treatment intensity, and cost in competing indemnity and HMO plans, NBER Working Paper 7832.

Beck, K. (1999), Risk adjustment and competition in Swiss health insurance market-developments, shortcomings and alternatives, Internal paper, CSS Insurance, Lucerne.

Beeson Royalty, A. and N. Solomon (1999), Health plan choice; price elasticities in a managed competition setting, *Journal of Human Resources* 34(1), pp. 1-41.

Chalkley, M. and J.M. Malcomson (1998a), Contracting for health services when patient demand does not reflect quality, *Journal of Health Economics* 17(1), pp. 1-19.

Chalkley, M. and J.M. Malcomson (1998b), Contracting for health services with unmonitored quality, *Economic Journal* 108, pp. 1093-1110.

CPB/SCP (1999), Ramingsmodel Zorgsector (in Dutch), Final report 2nd phase, CPB, The Hague.

Cutler, D.M. and S.J. Reber (1998), Paying for health insurance: the trade-off between competition and adverse selection, *Quarterly Journal of Economics* 113(2), pp.433-466.

Cutler, D.M. and R.J. Zeckhauser (2000), The anatomy of health insurance, *Handbook of Health Economics*, (eds. J.P.Newhouse and A.J. Culyer), chapter 11, pp. 563-643.

Ellis, R.P.(1998), Creaming, skimping and dumping: provider competition on the intensive and extensive margins, *Journal of Health Economics* 17, pp. 537-555.

Ellis, R.P. and T.G. McGuire (1986), Optimal payment systems for health services, *Journal of Health Economics* 9 (4), pp. 375-396.

Glied, S. (2000), Managed Care, *Handbook of Health Economics*, (eds. J.P.Newhouse and A.J.Culyer), chapter 13, pp. 707-753.

Laffont, J.J. and J. Tirole (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.

Newhouse, J.P. (1996), Reimbursing health plans and health providers: efficiency in production versus selection, *Journal of Economic Literature* 34, pp. 1236-1263.

Rothschild M. and J. Stiglitz (1976), Equilibrium in competitive insurance markets: An essay on the economics of imperfect information, *Quarterly Journal of Economics* 90(4), pp. 629-650.

Schut, F.T. (1995), Competition in the Dutch health care sector, dissertation, Erasmus University Rotterdam.

Van de Ven, W.P.M.M. and R.P. Ellis (2000), Risk adjustment in competitive health plan markets, *Handbook of Health Economics*, (eds. J.P.Newhouse and A.J. Culyer), chapter 14, pp. 755-845.

Van Vliet, R.C.J.A., L.M. Lamers and W.P.M.M. van de Ven (1999), Verbetering van de incentive-structuur op de ziekenfondsmarkt, Internal Report (in Dutch), iBMG, Erasmus University Rotterdam.

**Abstract**

The efficient delivery of medical services may be pursued by intensifying competition among health care insurers. This paper develops a model of regulated competition among health care insurers. It shows that increasing competition may foster efficiency-raising activities, reduce insurer profits and lower health care costs. However, it may also increase the variability of consumer premiums and increase risk-type specific selection activities by insurers as the government will generally lack information on the risk characteristics of the insured.



*keywords*: Health insurance, Regulated competition, Nash equilibrium, Risk adjustment.

*JEL classification:* I11; I18; D8