



BGPE Discussion Paper

No. 22

**Double Moral Hazard and Outcome-based
Remuneration of Physicians**

Udo Schneider

March 2007

ISSN 1863-5733

Editor: Prof. Regina T. Riphahn, Ph.D.
Friedrich-Alexander-University Erlangen-Nuremberg
© Udo Schneider

Double Moral Hazard and Outcome-based Remuneration of Physicians

Udo Schneider^a

^a *University of Bayreuth*

Correspondence address:

University of Bayreuth
Department of Law and Economics,
Economics and Public Finance
D-95440 Bayreuth, Germany

Phone: +49-921-552929;
Fax: +49-921-555821
E-mail: Udo.Schneider@uni-bayreuth.de

Abstract

The discussion about health care systems focuses on the dynamics of expenditures and on the weak growth of the revenue base. In this discussion it is widely overseen that medical expenditures and supply of medical services crucially depend on the compensation of physician services. The paper analyses the implementation of an outcome-based payment system in the presence of asymmetric information. Two cases are studied in detail: first, the common situation of physician's moral hazard and second, a double moral hazard model. The choice of insurance and payment contracts then depends on the characteristics of asymmetric information.

Keywords: outcome-based remuneration, double moral hazard, health policy

JEL-classification: I 11, I 12, D 82

1 Introduction

The political discussion about reforming the health care systems centres on the dynamics of expenditures and on missing growth of the revenue base. Especially the structure of financing, payment, and regulation in the health care system aggravates the problems in this field of economics. But the discussion ignores one characteristic of health care markets, by name the unequal distribution of knowledge about health status and treatment options. Asymmetric information, i. e. the lack of incentive compatibility, is one of the prevalent topics in the health care market. The economic consequences of asymmetric information between two contract parties consist in the fact that one party might take advantage of the resulting scope. Moreover, in case of diverging individual goals it is possible that individual interests of both the physician and the patient are conflicting so that the optimal treatment quality is not always in the main interest.

Although the physician-patient relationship is the key relationship in health care, the starting point of health care reforms is the revenue and expenditure situation. Especially in Germany it is obvious that health policy changed during the last decade. In the first half of the nineties, the discussion was focused on the personal responsibility of the patient. Nowadays, the core of the discussion is about the future of financing health care. Thus, personal responsibility of the patient and patient orientation has taken a back seat in the debate about the future design of the health care system as well as different payment systems for physician services. In contrast to the background of this situation, one has to analyse in how far an outcome-based payment system can be seen as a step towards efficiency and higher effectiveness in health care.¹ The main interest of this paper is to show the consequences of asymmetric information for the formulation of insurance and payment contracts when an outcome-based remuneration system is implemented. Thereby, in addition to the moral hazard of the physician the patient's health related behaviour is also analysed. In this situation of double moral hazard, the insurer has to consider the incentives of the contract parameters on medical services and the patient's compliance.

¹ In other parts of the economy performance-based compensation systems appear on the agenda more often. From a theoretical point of view especially the vast literature about principal-agent problems can be seen as a starting point for performance- or outcome-based remuneration systems (see e. g. Holmström (1979) and Holmström (1982)).

The paper is organized as follows. Chapter two discusses basic principles of outcome-based compensation and deals with resulting implementation problems. Moreover, a survey about the literature concerning theoretical models of physician payment systems in general and outcome-based systems in particular is presented. In section three, the basic model is developed, followed by the derivation of optimal insurance and remuneration contracts: first, in the situation of physician moral hazard, and second if double moral hazard is present. The fourth chapter deals with the resulting implications for health policy. The paper ends with a conclusion.

2 Outcome-based payment systems

2.1 Basic principles

The behaviour of health care providers and the supply of physician services depend heavily on the underlying remuneration system (cf. Advisory Council for the Concerted Action in Health Care (2003)). In the case of a fee-for-service remuneration, the physician has an incentive to expand the quantity of supplied services. The use of a lump-sum payment instead leads to selection effects: flat rate payments encourage the physician to refuse patients with difficult diseases and capitation fees have disadvantages if the patient's treatment is cost-intensive. In other words, we find a selection with respect to health risks. If the health care provider earns a fixed salary it is possible that this has undesired effects concerning his productivity.

The idea behind outcome-based compensation systems is the separation of provision of medical services from remuneration. The major targets are a higher effectiveness and an efficient health care provision (cf. Advisory Council for the Concerted Action in Health Care (1998) and Stade and Stahlecker (2001)). The aim of remunerating the physician according to health outcomes is to archive a more appropriate provision of health care services. Therefore, the treatment process needs to focus more on the health production process. Therefore, it is necessary to verify the medical services supplied. On the one hand, medical procedures have to be analysed with respect to their effectiveness and on the other hand, medical facilities should be audited regarding their productivity. If someone takes a look at the health care system today, a payment system that focuses entirely on health outcomes is not

implemented.² One of the main issues for this finding might be that a patient has no ability to assess health care quality and professional skills of the physician. By the implementation of an outcome-based payment system the patient's opinion would gain weight in the decision process in contrast to a fee-for-service system.

Generally, what is needed for an outcome-based payment system is a definition of outcomes of the health care process. Therefore, three conditions of health outcomes have to be considered. One possibility is a full recovery as a starting point for the remuneration of physician services. However, situations in which a full recovery is not possible or not achievable with maintainable inputs are not covered in this environment. Especially in cases of multi-morbidity or chronic diseases, the physician often has to balance between treatment success and adverse effects of medical care. Here, an improvement in the health status may be an adequate criterion for the remuneration. Moreover, there might be the case in which even an improvement is nearly impossible to achieve and stabilising the health status can be viewed as a success from a medical perspective. In the end, the implementation of a payment system that is only based on the improvement of the health status neglects these medical aspects and leads to an unjustified penalization of the provider of medical services. Thus it is necessary to find valid indicators to measure the effect of a medical treatment independent of the exact definition of health outcome.

An important factor concerning the implementation of outcome-based payment systems is the transparency of the results of the health production process (cf. Advisory Council for the Concerted Action in Health Care (1998), p. 65ff.). It is important that the medical providers evaluate the medical treatment. In addition to informing the third-party payer it is necessary to involve the patient directly in the treatment process. First, the individual comprehension of health outcome is necessary to encourage the patient to exhibit the best possible health related effort. Second, the patient needs detailed information about the quality of health care. Beyond that, implementation of a payment system based on outcomes is only feasible in accordance to medical providers and medical associations. They must have a virtual interest for outcome-based remuneration.³ Moreover, the position of a gatekeeper for medical services is central

² For a survey of outcome-based payment systems see Lu and Donaldson (2000).

³ One advantage of outcome-based payment systems is that medical providers have an interest for training on the job in order to improve the service quality.

for the physician's future position in his relation to patient and insurer.⁴ Especially the communication between physician and patient is of major importance as well as the mutual information transmission is (cf. Schneider (2004)). Here, it is worth mentioning that the physician-patient relationship should be settled on a basis of confidence.

The implementation of such a remuneration system is fairly problematic. One issue is the causal connection of outcome and medical services. Generally, a causal interrelationship cannot be found because besides medical services numerous other factors like environmental conditions, patient's self-healing capabilities and of course the present symptoms influence the treatment outcome. Therefore, a deterministic relation between inputs and outputs is unascertainable. In practice, we observe several attempts to implement outcome-based or performance-oriented remuneration systems of physician services. There exist medical guidelines or modular concepts for payment systems. For the first alternative, the physician is supposed to treat the patient in accordance with existing guidelines to give him some kind of support about decisions concerning therapy and compliance.⁵ A modular payment system as a second option consists of several components (cf. Advisory Council for the Concerted Action in Health Care (1998), p. 69ff.). These are e. g. a flat rate payment for covering the fixed costs of treatment, a charge for frequent and standardized services as well as a compensation for special services such as home visits or referrals to specialists. Moreover, the provider gets a results-orientated bonus in case of a successful treatment which may also be on the basis of a 'group result'.⁶

2.2 Theoretical models of outcome-based compensation

In health economics literature, only few papers explicitly deal with the question of compensation in accordance to patient-related health outcomes. The vast majority captures only the administrative effects of the implementation of such a payment system (cf. Rochaix (1998)). First of all, there is not a clear distinction between outcome-based or performance-based remuneration. The latter incorporates the idea of guidelines where the input in the

⁴ This relation between physician, patient and insurer is known as health care triangle (cf. Massialos and Dixon (2002), p. 2).

⁵ Guidelines are systematically developed, evidence-based statements decided by consensus. They aim at giving decision and orientation guide to physicians and patients for medical treatment in the case of certain characteristic conditions (cf. Advisory Council for the Concerted Action in Health Care (1998), p. 72).

⁶ Even these proposals of outcome-based payment systems face the problems mentioned above.

treatment process is important for its output. In contrast to this, outcome-based payment systems focus on the success of the treatment process.

In the theoretical literature about compensation of ambulatory physician services, most of the work concentrates on the contract design for physicians with regard to incentive compatibility and efficiency.⁷ Performance- or outcome-based payments play only an underpart. Amongst these works, Zweifel (1994) was one of the first to develop a model about contracts between physician and patient that incorporates elements of outcome and effort. Using a principal-agent framework with physician's moral hazard, the optimal contract is one that gives the physician a share of the treatment outcome in monetary terms. This solution combines elements of outcome-based remuneration and effort-related payment systems. As long as a higher physician effort leads to a higher probability of a good health outcome, the remuneration increases. On the controversy, if the physician only provides a low effort level, the probability of a good outcome decreases and therefore, the remuneration decreases too. This solution has two main disadvantages. First, the patient's signing of the contract is subject to some kind of urgency if he suffers a severe illness. Second, the patient is not in a position to put the physician's fee into action. It follows that the patient needs help from so called complementary agents whose primary tasks are information brokering, negotiation and conclusion of contracts (cf. Zweifel, Lehmann and Steinmann (2002)).

The starting point in the work of Stade and Stahlecker (2001) are the rising health care expenditures. They aim at analyzing the influence of different remuneration regimes on physician services and therefore on the development of expenditures. Hence, they model the relation between a sickness fund and a physician in the context of a principal-agent model using a combination of an outcome-based and a service-based remuneration. The physician (agent) maximizes his expected profits by treating a patient. The sickness fund acts as a principal that maximizes his expected profits with respect to the remuneration parameters. The main result is that a combination of outcome- and service-based compensation dominates a solely service-based payment system.

Leonard and Zivin (2003) assume that the outcome of the health production process is influenced by both the physician and the patient. This situation is called dual hidden effort. It

⁷ Examples of research in this field are the models of Ellis and McGuire (1990), Selden (1990), Ma (1994), Chalkley and Malcomson (1996), Kwon (1997), Ma and McGuire (1997), Chalkley and Malcomson (1998), De Fraja (2000) or Jelovac (2001).

describes the case where on the one hand the physician is unable to verify the patient's effort and on the other hand the patient has no exact knowledge about the medical effort of the physician. In this model a fee-for-service compensation is compared to an outcome-based remuneration system. The implementation of the negotiated contract lies in the responsibility of a third party, namely the employer or regulator. The most remarkable result is that an outcome-based payment dominates a fee-for-service payment in the case of a complementary relation between physician's services and patient's effort. In all other cases a service-based remuneration is preferred.

A model of performance-based payment is presented in Lu, Ma and Yuan (2003). The main focus is on the incentives on patient selection and matching if patient heterogeneity and asymmetric information are present. Performance-based contracts include a basic compensation and an opportunity for additional compensation adapted from measures such as quality of care and treatment outcomes. It is assumed that treatment performance depends on the severeness of illness and the intensity of treatment. By the use of a performance-based compensation system it is possible to reduce the information mismatch between physician and patient. Moreover, this form of contracts leads to a better match between illness severity and treatment intensity and to more referrals.

To sum up, in the presented models the analysis concentrates on allocation purposes and neglects the interaction between physician and patient. Only in the paper by Leonard and Zivin 2003 this subject is discussed. Nevertheless, all of these approaches omit the implementation problems associated with performance-based and outcome-based remuneration systems. Instead, the paper at hand presents a model of double moral hazard between physician and patient concentrating on health policy implications and implementation aspects.

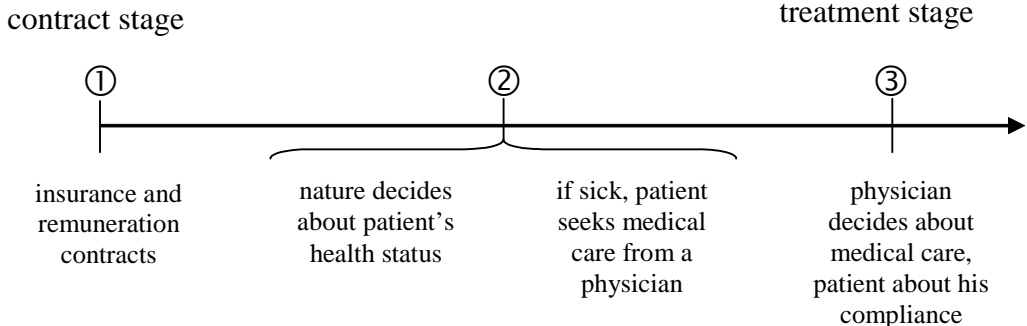
3 Asymmetric information and optimal insurance-provider contracts

3.1 Structure of the model

The model describes an agency relationship between physician and patient. In addition, an insurance company is incorporated. The model consists of three stages (see figure 1). At the first stage, the insurance company decides about the insurance and remuneration contracts. Here, the contract parameters are determined. At stage two, nature chooses the patient's state of health. If he is sick, he will visit the physician and demand medical treatment. In other

cases, the game ends. At stage three, the physician and the patient decide simultaneously about the medical services provided and the compliance (health related behaviour). To solve this model, we use the method of backward induction, i. e. it is necessary to look at decisions about the treatment of physician and patient first and then to incorporate these results in the contract stage.

Figure 1: Stages of the game



For the patient’s utility function U , it is assumed that he owns an initial wealth W and is risk averse in disposable income (U is concave). The probability of getting sick p is exogenously given, with $p \in [0, 1]$. This means that the patient has ex ante no influence on his state of health. In the case of illness, the patient suffers a health shock in monetary units L . This shock depends negatively on the amount of consumed medical services m and on his compliance a , respectively. L is a convex function of both variables. This formulation implies diminishing marginal benefits of both inputs.

The patient can purchase insurance for which he has to pay a premium σ . Moreover, in the case of treatment, he has to pay a fixed share β of the expenditures for medical treatment (coinsurance: $0 < \beta < 1$). These expenditures depend on the two elements of the physician remuneration system: First, the outcome of the treatment h and second, a lump sum payment Φ . The outcome-indicator h is a random variable that depends on medical services and compliance and is restricted to the interval $[0, 1]$. It is a twice differentiable and increasing function of the amount of medical services and compliance $h(m, a)$. Moreover, it is assumed that h is concave in m and a . If it takes the value 1, the outcome of the medical process can be interpreted as a full recovery. In the other extreme case where $h=0$, medical treatment and compliance take no effect on the state of health and so the patient suffers the full loss. Total

expenditures for medical services, which are equal to the physician's remuneration are the product of outcome and lump sum payment $h(m)\Phi$.

The patient's compliance in the case of illness is associated with disutility $D(a)$ which is a convex function of the compliance a so that in this case utility is additive-separable in income and disutility. The expected utility is then:

$$EU = p \left[U \left(W - \sigma - \beta h(m, a) \Phi - L(m, a) \right) - D(a) \right] + (1 - p) U(W - \sigma) \quad (3.1)$$

In the subsequent analysis the following notation is used as a simplification to express the patient's utility:

$$U_1 = U \left(W - \sigma - \beta h(m, a) \Phi - L(m, a) - D(a) \right)$$

$$U_2 = U(W - \sigma)$$

If the patient is sick his income-related utility is U_1 , if he is healthy, he only has to pay the insurance premium and his utility is denoted U_2 , with $U_1 < U_2$.

The expected utility of the physician is additive-separable in remuneration and effort. For the medication he gets a lump-sum payment Φ that is directly related to health outcome h . In detail, this means that if the patient recovers fully ($h = 1$) the physician will receive the full payment Φ . If the health status of the patient remains poor ($h = 0$) the physician gets no payment at all. For any values of h between 0 and 1, he only gets a fraction of Φ . Overall, it is assumed that the physician is risk neutral because he is able to spread the income risk over all patients. The effort of supplying medical services is $C(m)$ which is a convex function. The expected utility is:

$$EV = p \left[h(m, a) \Phi - C(m) \right] \quad (3.2)$$

The insurance company is risk neutral and finances the health care expenditures in the case of an illness (p) against a premium σ . The insurance pays for all treatment costs except the co-payment share β . The insurance supplies this service at actuarial fair premiums on a competitive insurance market:

$$\sigma = p(1 - \beta) h(m, a) \Phi \quad (3.3)$$

Considering the information structure, the following two model alternatives are regarded. First, in a restricted model it is assumed that the patient reveals no compliance ($a = 0$). In this case, health outcome only depends on the physician's actions. In the second scenario, patient's compliance influences health outcomes and health shock and the insurer takes patient's compliance into account.

3.2 *Insurance contracts with asymmetric information and no compliance*

To develop some sort of benchmark regime, suppose that the insurance company does not account for the patient's compliance and that the health outcomes as well as the health shock depend only on medical services. This corresponds to the situation in which the patient shows no health-relevant, treatment-accompanying activities at all. In this case, the patient exhibits no compliance and there is no disutility associated with his behaviour. We can then write equations (3.1) to (3.3) as:

$$EU = p U (W - \sigma - \beta h (m) \Phi - L (m)) + (1 - p) U (W - \sigma) \quad (3.4)$$

with

$$U_1 = U (W - \sigma - \beta h (m) \Phi - L (m))$$

$$U_2 = U (W - \sigma)$$

$$EV = p [h (m) \Phi - C (m)] \quad (3.5)$$

$$\sigma = p (1 - \beta) h (m) \Phi \quad (3.6)$$

The insurer is unable to observe the medical services provided by the physician. The latter maximises his expected utility of treatment on stage three ($p = 1$) by choosing the adequate level of medical services.⁸ In this case, the maximisation of equation (3.2) can be written as:

$$\max_m h (m) \Phi - C (m) \quad (3.7)$$

and the resulting first-order condition is

$$h_m \Phi = C' \quad (3.8)$$

⁸ The patient's decision about his compliance is inapplicable in this situation.

Here, the marginal medical effort on the right-hand side equals the expected marginal benefits on the left. The latter depends on the marginal health outcome and on the lump-sum payment. For questions about the contract design it is interesting how an increase in the lump-sum payment affects the provision of medical services. Therefore, by applying the implicit-function theorem it follows from equation (3.7):

$$\frac{dm}{d\Phi} = - \frac{h_m}{h_{mm} \Phi - C''} > 0 \Rightarrow m(\Phi) \quad (3.9)$$

It is obvious that a higher payment has a positive effect on the medical services provided by the physician. The numerator is positive and equation (3.9) requires that the second order derivative of (3.7) in the denominator also is negative. Hence, in a world of asymmetric information about the quality of the provided medical services it is important to note how a physician reacts if the lump-sum payment changes. The resulting condition is further referred to as the incentive compatibility of the physician.

The implications for the contractual arrangements are analysed at the contract stage. In analogy to Stewart (1994) it is assumed that the insurance company maximises patient's expected utility with respect to the coinsurance parameter and the lump-sum payment. The insurer considers the following restrictions. First, the physician's expected utility has to be at least as great as his reservation utility \bar{V} (participation constraint). Second, the insurance premium equals expected expenditures and third, the lump-sum payment has a positive effect on medical services (incentive compatibility constraint). The optimisation problem is further denoted as problem 1:

$$\begin{aligned} \max_{\beta, \Phi} & p \left[U(W - \sigma - \beta h(m) \Phi - L(m)) \right] + (1 - p) U(W - \sigma) \\ \text{s.t.} & p \left[h(m) \Phi - C(m) \right] \geq \bar{V} \\ & \sigma = p(1 - \beta) h(m) \Phi \\ & m = m(\Phi) \end{aligned} \quad (3.10)$$

To simplify the analysis, the zero profit condition of the insurer as well as the physician's incentive condition are included in the patient's expected utility and the physician's participation constraint.

For the patient's coinsurance parameter β the resulting first order condition after rearranging is:⁹

$$U_1' = U_2' \quad (3.11)$$

It is clear from equation (3.11) that the marginal utility in case of an illness equals marginal utility if no treatment is necessary. Because of the concavity of the utility U this is only possible if the arguments of the utility function equal each other. For this to be true, two conditions must be fulfilled. First, the health shock L has to be absorbed, i.e. m has to be large enough so that the patient faces no loss. Second, the patient must not have any out-of-pocket expenditure for the medical care. In this case, the coinsurance parameter β is zero and the patient is fully insured.¹⁰

The first-order condition with respect to the lump-sum payment Φ after using some algebra is given by the following equation:

$$\begin{aligned} \lambda p h(m) - p L_m m_\Phi U_1' &= p \beta h(m) U_1' \\ &+ p (1 - \beta) h(m) (p U_1' + (1 - p) U_2') \\ &+ p \beta h_m m_\Phi \Phi U_1' \\ &+ p (1 - \beta) h_m m_\Phi \Phi (p U_1' + (1 - p) U_2') \end{aligned} \quad (3.12)$$

The expected marginal benefits of a higher lump-sum payment on the left-hand side equal the expected marginal costs of an increase on the right-hand side.¹¹ To begin with the benefits on the left, they consist of two parts. First, the direct effect of an increase is the expected utility gain of the physician because his income is expected to rise. The second effect is indirect and results from a better health status of the patient. The higher lump-sum payment raises the supply and use of medical services. Furthermore, the monetary health shock decreases. Considering the cost of a higher lump-sum payment on the right-hand side one has to distinguish between two direct and two indirect effects. The first term denotes the direct

⁹ A detailed derivation of the results is presented in the appendix.

¹⁰ Note that if the monetary health shock cannot be reduced to zero by the use of medical care the coinsurance parameter is negative which means that the patient gets a payment to compensate the loss. This situation is ruled out by the assumption that the coinsurance parameter β has to be strictly larger than zero.

¹¹ The Lagrangian parameter λ can be interpreted as a fixed weight to the physician's expected utility (cf. Rees (1985), p. 21). In this case, a Pareto-optimal situation is regarded under consideration of incentive compatibility conditions.

coinsurance effect. Given a higher lump-sum payment the patient c. p. has higher out of pocket expenditures. The second term is called the direct premium effect. The increase in the lump-sum payment c. p. raises medical expenditures. In order that the zero profit condition of the insurance company (equation (3.6)) remains valid, the premium has to increase for both possible states of health, sick and healthy. The last terms are the indirect coinsurance and premium effects. Both appear due to an increase in medical services which results from the incentive compatibility condition (3.9). As before, the indirect coinsurance effect is only relevant if the patient seeks medical treatment by a physician and the expected rise in the insurance premium contemporaneously affects the healthy state of nature.

Consider that some of the above effects disappear if no coinsurance is incorporated in the insurance contract. Together with the result obtained from (3.11) that the coinsurance parameter β is zero and that the marginal utility is the same in all conditions, equation (3.12) simplifies to:

$$\lambda p h(m) - p L_m m_\Phi U'_1 = p h(m) U'_1 + p h_m m_\Phi \Phi U'_1. \quad (3.13)$$

Here, the direct and indirect coinsurance effects in the case of illness vanish and the new first-order condition only depends on the marginal utility given a physician visit. The left-hand side is unchanged and on the right hand side, only the expected premium effects remain in the equation. This means a higher maximum lump-sum payment raises the insurance premium in two ways. First, the same health outcome leads c. p. to a higher physician reimbursement. Second, the higher lump-sum payment induces a higher level of medical services, which enhances the probability of a good health outcome. Moreover, the negative premium effects are no more suspended through the coinsurance which means that any rise in expected health care expenditures leads to an increase in the insurance premium depending on the probability of an illness (p).

All in all, it is obvious that in a situation with no compliance the patient chooses full insurance. The reason is that there is no moral hazard on the patient's side. In contrast to this, the moral hazard of the physician remains a problem. The physician neglects the impact of his behaviour on the patient when maximizing his expected utility. For that reason, outcome-based payments introduced set to balance the sum of expected marginal benefits of the physician and the patient and marginal costs of the patient.

3.3 Double moral hazard and remuneration contracts

In contrast to the no-compliance model, we analyse the more general model described in chapter 3.1 in the section at hand. Here, the patient's compliance influences the measure of health outcome h positively but at a decreasing rate. Moreover, the monetary health shock L is reduced, with $L_a < 0$ and $L_{aa} > 0$. Besides these positive effects compliance is associated with disutility which is a convex function. The latter reduces patient's expected utility in the case of a treatment. As before, we analyse different stages of the game. First, both, the physician and the patient, maximise their utilities in the case of a medical treatment non-cooperatively (stage three). At this point, we face the situation of double moral hazard where neither the patient nor the physician can observe or monitor the other one's action. In the next stage, nature decides about patient's health status and at the final stage, the insurer sets the insurance and payment parameters with respect to the results of the treatment stage.

The patient maximises his expected utility with respect to his compliance a :

$$\max_a U \left(W - \sigma - \beta h(m, a) \Phi - L(m, a) \right) - D(a) \quad (3.14)$$

And the resulting first order condition is given by:

$$\left(-L_a - \beta h_a \Phi \right) U_1' = D' \quad (3.15)$$

The optimal choice of compliance equals the expected marginal utility and the marginal disutility. The marginal benefits are on the left-hand side of equation (3.15). It is clear that the loss-reducing effect of a higher compliance ($-L_a$) must exceed the negative effect of a higher coinsurance ($-\beta h_a \Phi$). Applying the implicit-function theorem to equation (3.15) one obtains the effect of a higher coinsurance rate on the level of compliance:

$$\frac{da}{d\beta} = \frac{h_a \Phi U_1' + h(m, a) \Phi \left(-L_a - \beta h_a \Phi \right) U_1''}{\left(-L_{aa} - \beta h_{aa} \Phi \right) U_1' + \left(-L_a - \beta h_a \Phi \right)^2 U_1'' - D''} \quad (3.16)$$

The denominator of equation (3.16) is the sufficient condition of a utility maximum. Its sign depends on the first term in parenthesis. If it is negative then the whole denominator is negative, too. The numerator displays the effect of a higher coinsurance on expected marginal

benefits of compliance. Unfortunately, its sign remains unclear.¹² The first term is positive and the second term is negative as long as the term in parenthesis is positive as it is assumed for equation (3.15). A higher coinsurance rate leads to an increase in patient's compliance if, in absolute terms, the first term in the nominator is smaller than the second term. In fact, this result depends on patient's influence on the health outcome h and thereby on the physician remuneration and on out-of-pocket expenses.

The effect of a higher lump-sum payment on the level of compliance is:

$$\frac{da}{d\Phi} = \frac{\beta h_a U_1' + \beta h(m,a) (-L_a - \beta h_a \Phi) U_1''}{(-L_{aa} - \beta h_{aa} \Phi) U_1' + (-L_a - \beta h_a \Phi)^2 U_1'' - D''} \quad (3.17)$$

Like in equation (3.16) the overall effect of Φ on a is not clear. The denominator corresponds to the sufficient condition for a utility maximum and is negative if the first term in parenthesis is negative. The numerator shows the effect of a higher lump-sum payment on the marginal benefits of compliance. The first term is positive and the second one is negative. As a consequence of these results, the effects of coinsurance parameter and lump-sum payment on patient's compliance are uncertain and it follows for the patient's incentive constraint that:

$$a(\beta, \Phi) \quad \text{with} \quad \frac{\partial a}{\partial \beta} \begin{matrix} > \\ \equiv \\ < \end{matrix} 0 \quad \text{and} \quad \frac{\partial a}{\partial \Phi} \begin{matrix} > \\ \equiv \\ < \end{matrix} 0 \quad (3.18)$$

Simultaneously, the physician maximises his expected utility with respect to the medical services at the non-cooperative stage. The maximisation problem and the resulting first order condition are given by:

$$\max_m \quad h(m, a) \Phi - C(m) \quad (3.19)$$

and

$$h_m \Phi = C' \quad (3.20)$$

In this non-cooperative environment the physician chooses that level of medical services which equals the marginal revenue and the marginal effort of providing health care. The

¹² This is in contrast to the effect in Schneider (2004) who analyses a double moral hazard problem with conventional insurance and a fee-for-service remuneration system.

former is determined by the higher probability of recovery because of a higher level of medical services and the level of the lump-sum payment. Now, one can ask how changes in the remuneration system affect the amount of medical services supplied. Here, the lump-sum payment Φ influences the choice of medical services because it determines the physician's income together with health outcome h . Total differentiating of (3.20) leads to the following expression:

$$\frac{d m}{d \Phi} = - \frac{h_m}{h_{mm} \Phi - C''} > 0 \Rightarrow m(\Phi) \quad (3.21)$$

The numerator is the partial derivation of the outcome indicator with respect to medical services and is positive. The denominator is the sufficient condition for a maximum and therefore negative. In total, an increase in the outcome-based payment has a positive effect on the medical services supplied and is referred to as the physician's incentive constraint.

The results above ((3.18) and (3.21)) are essential for the derivation of the optimal contract parameters in the case of double sided asymmetric information between physician and patient. For the contract stage, it is assumed that the insurer cannot observe the individual actions of the physician and the patient. Therefore, effort contingent contracts are not possible. Instead, the insurance company will maximise the patient's expected utility (equation (3.1)) subject to the physician's participation constraint (equation (3.2)), the zero profit-condition of the insurance company (equation (3.3)) and the incentive constraints of the patient and the physician (equations (3.18) and (3.21)). Following, this situation is denoted as problem 2:

$$\begin{aligned} \max_{\beta, \Phi} & p \left[U(W - \sigma - \beta h(m, a) \Phi - L(m, a)) - D(a) \right] + (1 - p) U(W - \sigma) \\ \text{s.t.} & p \left[h(m, a) \Phi - C(m) \right] \geq \bar{V} \\ & \sigma = p(1 - \beta) h(m, a) \Phi \\ & m = m(\Phi) \\ & a = a(\beta, \Phi) \end{aligned} \quad (3.22)$$

The maximisation variables are the patient's coinsurance parameter and the lump-sum payment of the physician. To simplify the analysis the zero profit condition as well as the incentive conditions are inserted in the patient's expected utility and the physician's participation constraint.

After rearranging, for the patient's coinsurance parameter β the resulting first-order condition is:¹³

$$\begin{aligned}
& \lambda p h_a a_\beta \Phi + p h(m, a) \Phi \left(p U_1' + (1-p) U_2' \right) \\
& = p h(m, a) \Phi U_1' + p h_a a_\beta \Phi U_1' \\
& + p h_a a_\beta \Phi (1-\beta) \left(p U_1' + (1-p) U_2' \right).
\end{aligned} \tag{3.23}$$

To interpret the condition above it is necessary to take a closer look at the effects of an increase in the co-payment parameter. If the compliance increases as a result of a higher β ($a_\beta > 0$) the left-hand side in equation (3.23) shows the expected marginal utility or revenue gains of the patient and the physician. The first term is the marginal rise in physician's income due to a better health outcome. The second term is the direct premium effect of a higher coinsurance rate on the patient's utility, since it is possible for the insurer to lower premiums in case of a higher co-payment. In addition, the right-hand side contains the marginal utility loss of the patient due to an increase in β . The first two terms on the right are the direct and indirect coinsurance effect. Here, for the direct effect, a higher coinsurance rate leads to higher out-of pocket expenses and to a loss in utility. Moreover, the higher coinsurance rate causes a higher compliance and therefore a better health outcome (indirect effect). The third term is the indirect premium effect that leads to an increase in the insurance premium because better outcomes increase the expenditures for physician's services.

In other cases, if the compliance remains unchanged or if it decreases the indirect effects of a change in the coinsurance parameter vanish or change their sign. In detail, the results mainly depend on the physician's remuneration regime and the patient's influence on the remuneration (see equation (3.17)). In case of a fee-for-service remuneration, the patient's compliance has no influence on the physician's payment because this depends only on the medical services supplied, independently of the patient's actions.¹⁴

It is clear that the result in (3.23) is different to equation (3.11). In section 3.2, the patient is fully insured because the marginal utilities if sick and healthy are equal. Instead, in the double

¹³ Again, the reader will find a detailed derivation of the results is presented in the appendix.

¹⁴ It is clear that ex ante moral hazard may lead to an increase in medical services to achieve full recovery. In this case, the patient's behaviour influences the remuneration of the physician and the effect of an increase in the co-payment parameter may be unclear.

moral hazard situation the patient faces a positive coinsurance. The first term on the left is the expected marginal income of the physician (indirect effect). The second term shows the expected premium reduction for the patient if he is healthy (direct effect). In equilibrium the sum of these two effects equals the expected marginal costs of a higher fraction of medical expenses the patient has to bear. The first term on the right is the direct expected net premium effect if the patient is ill and visits the physician. Here, two partial effects have to be distinguished. On the one hand, a higher expenditure share leads to a lower insurance premium and on the other hand, it causes a higher out of pocket expense. It is necessary for the implementation of a coinsurance that the latter effect dominates the premium effect, i.e. that the patient suffers from a financial loss in case of an illness even if he is insured against the risk of medical expenditures. This can be described as the incentive effect of a coinsurance. The second term on the right-hand side is the expected indirect effect on the insurance premium because of a higher coinsurance parameter. If the compliance increases through an increase in the coinsurance parameter this leads to a better health outcome h and therefore to a higher income of the physician. This again will raise the insurance premium the patient has to pay whether he uses medical services or not. An optimal coinsurance has to balance these positive effects for both the physician and the patient and the negative effects for the patient.

If one takes a look at the derivation of the maximisation problem with respect to the lump-sum payment Φ it is evident that the first order condition differs again from the situation in which the patient shows no compliance. The simplified condition is:

$$\begin{aligned}
& \lambda p h(m, a) + \lambda p h_a a_\Phi \Phi - p L_m m_\Phi U_1' \\
& = p \beta h(m, a) U_1' \\
& \quad + p (1 - \beta) h(m, a) \left(p U_1' + (1 - p) U_2' \right) \\
& \quad + p \beta \Phi \left(h_m m_\Phi + h_a a_\Phi \right) U_1' \\
& \quad + p (1 - \beta) \Phi \left(h_m m_\Phi + h_a a_\Phi \right) \left(p U_1' + (1 - p) U_2' \right)
\end{aligned} \tag{3.24}$$

As before, the results depend on the sign of a_Φ . If a positive effect of an increasing lump-sum payment is assumed ($a_\Phi > 0$) then the left-hand side contains the expected marginal benefits of the patient and the physician and the right-hand side marks the expected marginal losses in patient's utility. In detail, the first two terms on the left are the direct and the indirect effect of a higher lump-sum payment on the physician's expected income. The direct effect follows

from the increase in the payment parameter Φ . The indirect effect depends on the influence of this parameter on the patient's compliance and thus on health outcome. The third term is the patient's expected marginal utility of a higher level of medical services due to a higher lump-sum payment. Because of the higher payment the physician supplies a higher level of medical services, which reduces the monetary health shock L .

The expected marginal losses of the patient can be decomposed as follows. The first term on the right is the direct coinsurance effect of a higher lump-sum payment in case of an illness. The higher compensation increases out-of-pocket expenditures and reduces the patient's net income. It's the same case for the second term, the direct premium effect that matters in all states of health. The higher reimbursement has to be balanced by increasing the insurance premium to guarantee for zero profits. The last two terms are the indirect coinsurance and premium effects. Their magnitude depends on the influence of the lump-sum payment on medical services and compliance and thus on health outcome h .

It is obvious that the incorporation of patient's compliance (problem 2) leads to more complex first-order conditions. A major difference to problem 1 is the fact that patient's reactions to changes in the insurance and remuneration parameters are uncertain. In the presented model we have assumed that patient's compliance depends positively on a higher coinsurance rate and a higher lump-sum payment for the physician. Instead, assuming a negative dependence or independency changes the results presented in this section. In case that the coinsurance parameter and the lump-sum payment influence patient's compliance positively, a full insurance of the patient is not possible any more. This is due to moral hazard of the patient. The difference between equations (3.12) and (3.24), the first-order conditions for the lump-sum payment, is that in the former case the indirect effects of the lump-sum payment on the patient's choice of actions are neglected. Therefore, given the situation without compliance (problem 1) the equilibrium is second best and the situation above that considers the compliance (problem 2) can be described as third best.

4 Implications for health policy

The key question for future payment systems for physician services is: In which way are we able to finance and grant good medicine? Considering the implementation of an outcome-based compensation for physician services there arise several problems. First, the consequences of the insurance and payment parameters β and Φ on patient's compliance are

ambiguous. This is crucial if one takes considers the results presented in section 3.3 because interpretation of the first-order conditions depends on the sign of the partial derivatives of a . Furthermore, first, it is not only difficult for the insurance company to estimate the effect of insurance and payment parameters on patient's compliance and the supply of medical services. Second, it is also complex to draw conclusions about the magnitude of these effects. As a consequence, the sign of indirect effects in equations (3.23) and (3.24) and therefore the interpretation of the first-order conditions are uncertain.

Second, a related problem is that health outcomes cannot always be identified by the insurer. Whether the reported health status is an objective or subjective indicator is not verifiable. In fact, one has to ask what the relevant health outcome h for the compensation is. In the theoretical model, h serves as an indicator that shows the resulting health status after treatment and can be interpreted as the degree of recovery. It does not incorporate that the probability of a recovery also depends on the severance of illness, the illness history and other more or less random factors. So, medical services and compliance are only two factors amongst others that the insurer has to carry in mind when considering outcome-based payment systems. For practical issues the indicator h ought to be chosen illness specific. This requires much more information about diagnosis, treatment, and patient behaviour which raises the costs of implementation.

At third, the insurer lacks knowledge about the interdependence of medical services and compliance in addition to the problem of measuring health outcomes. Two questions are of interest: First, how does the choice of medical services influence the patient's compliance (and vice versa)? Second, is it possible for the insurer to assign health outcomes to the physician or to the patient? The first question is relevant with respect to the implementation of incentives for both the physician and the patient by the insurance and payment system in order to obtain the optimal treatment quality. If the implementation of a coinsurance results in higher compliance but in lower medical services, health outcome might be negatively affected independently of the patient's contact decision. In such a situation, insurance companies or governmental regulators face a trade-off: On the one hand, setting the appropriate incentives for patients enforces a better compliance and health outcome increases. On the other hand, if compliance and medical effort are substitutes, this decreases the probability of a full recovery (cf. Schneider (2004)). With respect to the second question, it is clear that the insurer cannot draw conclusions about the fraction the patient has contributed to the resulting health status. If

policy makers do not take this problem into account, it is possible that the physician's compensation depends on a high level on the patient's behaviour.

Therefore, information about the productivity of medical services and the patient's health-related behaviour is needed for a successful implementation of an outcome-based compensation system. Moreover, the reaction of provider and patient to changes in insurance and payment parameters and information about health outcomes are important to archive an incentive compatible system. Hence, an improved communication between the actors of the health-care triangle (patient, physician and insurer) is one method to incorporate the aims of patient and physician. In a system of health care provision and financing, the patient plays a central part. On the one hand, he has an interest in a high quality medicine and on the other hand, he seeks insurance with an entire financing of treatment costs. This trade-off makes it difficult for the insurer to set the appropriate incentives. A system of full insurance supplies the patient with incentives to consume medical services at no costs but might not give him encouragement to provide the required compliance. A high coinsurance improves his compliance in some cases but may lead to a lower demand for medical services. Then, the physician has no opportunity to diagnose or treat the patient's complaints. The crux about the patient's compliance is that it is necessary for the success of medical treatment but that setting the right incentives seems nearly impossible. Therefore, in addition to medical guidelines, a system of guidelines for patients seems necessary to ensure that patient's health related behaviour is in accordance to medical treatment. Here, the physician has the position of a gatekeeper and is responsible for informing the patient about medical standards and his compliance.

5 Conclusion

Incentive problems are widespread in health economics. The discussion about the remuneration of physician services and its influence on the supply of health care services plays an important role in the discussion about the future perspectives of health care systems. Outcome-based compensation systems for physician services are one attempt to include the quality of health care provision into the remuneration. For ambulatory services, these payment systems are barely established yet. Among several reasons especially the problems of asymmetric information are an obstacle for the implementation. In detail, these refer to problems of mutual asymmetric information in the health care triangle physician-patient-insurer, the determinants of health outcomes and the incentive effect of insurance and

payment parameters on individual behaviour. Moreover, there is uncertainty about the measurement of health outcomes.

Using a theoretical model of double-sided asymmetric information between physician and patient, it is possible to show how the optimal solutions for coinsurance and remuneration parameters depend on the asymmetric information. All in all, the problems of asymmetric information discussed lead to the conclusion that the patient plays a central part in the health care triangle and that his behaviour determines health outcomes and the effects of remuneration systems.

Appendix

The first-order conditions derived in sections 3.2 and 3.3 are simplified conditions resulting from the differentiation with respect to the insurance and remuneration parameters in conjunction with the first-order conditions of the maximisation problems at stage three. Concerning problem 1 the Lagrangian can be written as:

$$\begin{aligned} \Gamma = & p U \left[W - p (1 - \beta) h(m(\Phi)) \Phi - \beta h(m(\Phi)) \Phi - L(m(\Phi)) \right] \\ & + (1 - p) U \left[W - p (1 - \beta) h(m(\Phi)) \Phi \right] \\ & + \lambda \left[h(m(\Phi)) \Phi - C(m(\Phi)) \right]. \end{aligned} \quad (\text{A.1})$$

The resulting first-order condition for the coinsurance parameter is:

$$\begin{aligned} \frac{\partial \Gamma}{\partial \beta} : & p^2 h(m(\Phi)) \Phi U_1' - p h(m(\Phi)) \Phi U_1' \\ & + p (1 - p) h(m(\Phi)) \Phi U_2' = 0 \end{aligned} \quad (\text{A.2})$$

Rearranging leads to equation (3.11). The derivation of the Lagrangian with respect to the lump-sum payment is more complex. The resulting first-order condition is given by:

$$\begin{aligned} \frac{\partial \Gamma}{\partial \Phi} : & - p^2 (1 - \beta) h(m(\Phi)) U_1' - p \beta h(m(\Phi)) U_1' \\ & - p (1 - p) (1 - \beta) h(m(\Phi)) U_2' + \lambda p h(m(\Phi)) \\ & - p^2 (1 - \beta) h_m m_\Phi \Phi U_1' - p \beta h_m m_\Phi \Phi U_1' - p L_m m_\Phi U_1' \\ & - p (1 - p) \beta h_m m_\Phi \Phi U_2' + \lambda p h_m m_\Phi \Phi - \lambda C_m m_\Phi = 0 \end{aligned} \quad (\text{A.3})$$

Together with the first-order condition of the physician's maximisation problem (equation (3.8)), one can rewrite the condition. This is presented as equation (3.12) in section 3.2.

If patient's compliance is incorporated into the model the Lagrangian of problem 2 is given by the following expression:

$$\begin{aligned}
\Gamma = & p U \left[W - p (1 - \beta) h (m (\Phi), a (\beta, \Phi)) \Phi - \beta h (m (\Phi), a (\beta, \Phi)) \Phi \right. \\
& \left. - L (m (\Phi), a (\beta, \Phi)) \right] - p D (a (\beta, \Phi)) \\
& + (1 - p) U \left[W - p (1 - \beta) h (m (\Phi), a (\beta, \Phi)) \Phi \right] \\
& + \lambda \left[h (m (\Phi), a (\beta, \Phi)) \Phi - C (m (\Phi)) \right].
\end{aligned} \tag{A.4}$$

Differentiation with respect to β gives the first-order condition of the coinsurance parameter.

$$\begin{aligned}
\frac{\partial \Gamma}{\partial \beta} : & p^2 h (m, a) \Phi U_1' - p h (m, a) \Phi U_1' + p (1 - p) h (m, a) \Phi U_2' \\
& - p^2 (1 - \beta) h_a a_\beta \Phi U_1' - p \beta h_a a_\beta \Phi U_1' - p L_a a_\beta U_1' \\
& - p (1 - p) (1 - \beta) h_a a_\beta \Phi U_2' - p D_a a_\beta + \lambda p h_a a_\beta \Phi = 0.
\end{aligned} \tag{A.5}$$

Together with equation (3.15), the patient's condition for a utility maximising compliance, the above equation simplifies to (3.23). The first-order condition for the lump-sum payment Φ is:

$$\begin{aligned}
\frac{\partial \Gamma}{\partial \Phi} : & - p^2 (1 - \beta) h (m, a) U_1' - p \beta h (m, a) U_1' - p (1 - p) (1 - \beta) h (m, a) U_2' \\
& + \lambda p h (m, a) - p^2 (1 - \beta) h_m m_\Phi \Phi U_1' - p \beta h_m m_\Phi \Phi U_1' \\
& - p L_m m_\Phi U_1' - p (1 - p) \beta h_m m_\Phi \Phi U_2' + \lambda p h_m m_\Phi \Phi - \lambda p C_m m_\Phi \\
& - p^2 (1 - \beta) h_a a_\Phi \Phi U_1' - p \beta h_a a_\Phi \Phi U_1' - p L_a a_\Phi U_1' \\
& - p (1 - p) \beta h_a a_\Phi \Phi U_2' + \lambda p h_a a_\Phi \Phi - p D_a a_\Phi = 0
\end{aligned} \tag{A.6}$$

Using the first-order conditions of the patient's and physician's maximization problems ((3.15) and (3.20)) condition (A.6) can be written as (3.24).

References

- Advisory Council for the Concerted Action in Health Care. Health Care Finance, User Orientation and Quality, Volume I: Health Care Finance and User Orientation, Report 2003 n/a, Bonn; 2003.
- Advisory Council for the Concerted Action in Health Care. The Health Care System in Germany Cost Factor and Branch of the Future, Volume II: Progress and Growth Markets, Finance and Remuneration, Special Report 1997, Summary n/a, Bonn; 1998.
- Chalkley, M., and J.M. Malcomson. Contracting for Health Services with Unmonitored Quality. *Economic Journal* 1998;108; 1093-1110.
- Chalkley, M., and J.M. Malcomson. Contracts for the National Health Service. *Economic Journal* 1996;106; 1691-1701.
- De Fraja, G. Contracts for Health Care and Asymmetric Information. *Journal of Health Economics* 2000;19; 663-677.
- Ellis, R.P., and T.G. McGuire. Optimal Payment Systems for Health Services. *Journal of Health Economics* 1990;9; 375-396.
- Holmström, B. Moral Hazard and Observability. *Bell Journal of Economics* 1979;10; 74-91.
- Holmström, B. Moral Hazard in Teams. *Bell Journal of Economics* 1982;13; 324-340.
- Jelovac, I. Physicians' Payment Contracts, Treatment Decisions and Diagnosis Accuracy. *Health Economics* 2001;10; 9-25.
- Kwon, S. Payment Systems for Providers in Health Insurance Market. *Journal of Risk and Insurance* 1997;64; 155-173.
- Leonard, K., and J.G. Zivin. Outcome versus Service Based Payments in Health Care: Lessons from African Traditional Healers. Working Paper 9797. National Bureau of Economic Research: Cambridge, MA. 2003.
- Lu, M., Ma, C.A., and L. Yuan. Risk selection and matching in performance-based contracting. *Health Economics* 2003;12; 339-345.
- Lu, M., and C. Donaldson. Performance-Based Contracts and Provider Efficiency. *The State of the Art. Disease Management & Health Outcomes* 2000;7; 127-137.
- Ma, C.A. Health Care Payment Systems: Cost and Quality Incentives. *Journal of Economics*

- and Management Strategy 1994;3; 93-112.
- Ma, C.A., and T.G. McGuire. Optimal Health Insurance and Provider Payment. *American Economic Review* 1997;87; 685-704.
- Massialos, E., and A. Dixon. 2002. Funding Health Care: An Introduction. In: E. Mossialos, A. Dixon, J. Figueras, and Kutzin, J. (Eds), *Funding Health Care: Options for Europe*. Open University Press, Buckingham et al.2002.
- Rees, R. The Theory of Principal and Agent: Part 1. *Bulletin of Economic Research* 1985;37; 3-26.
- Rochaix, L. 1998. Performance-tied payment systems for physicians. In: R. B. Saltman, J. Figueras, and Sakellarides, C. (Eds), *Critical Challenges for Health Care Reform in Europe*. Open University Press, Buckinham et al.1998.
- Schneider, U. Asymmetric Information and the Demand for Health Care - the Case of Double Moral Hazard. *Journal of Applied Social Science Studies* 2004;124; 233-256.
- Selden, T.M. A Model of Capitation. *Journal of Health Economics* 1990;9; 397-409.
- Stade, U., and P. Stahlecker. Erfolgsorientierte ärztliche Entlohnung?: Ein Modell zur Problematik einer aktuellen gesundheitspolitischen Diskussion. Beiträge aus dem Institut für Statistik und Ökonometrie 58. Fachbereich Wirtschaftswissenschaften, Institut für Statistik und Ökonometrie: Universität Hamburg. 2001.
- Stewart, J. The Welfare Implications of Moral Hazard and Adverse Selection in Competitive Insurance Markets. *Economic Inquiry* 1994;32; 193-208.
- Zweifel, P. 1994. Eine Charakterisierung von Gesundheitssystemen: Welche sind von Vorteil bei welchen Herausforderungen? In: P. Oberender (Eds), *Probleme der Transformation im Gesundheitswesen*. Nomos Verlagsgesellschaft, Baden-Baden.1994.
- Zweifel, P., Lehmann, H., and L. Steinmann. 2002. Patching up the Physician-Patient Relationship: Insurers vs. Governments as Complementary Agents. In: B. Lindgren (Eds), *Individual Decisions for Health*,. Routledge, London et al.2002.