The impact of liability for malpractice on the optimal reimbursement schemes for health services

Eberhard Feess and Sonja Ossig*

September 15, 2003

Abstract

We analyze the impact of liability risks for malpractice on the optimal reimbursement schemes for hospitals. In our model, the hospital decides upon two unobservable efforts, a cost reduction effort and a quality improvement effort. We assume that the total effort is positive even without monetary incentives due to some intrinsic motivation, but that motivation is biased towards quality. In our basic model without liability risks, we then find that either a fee-for-service system (FFS) or a fixed-fee prospective payment system (PPS) is optimal, but mixed systems are strictly inferior. With liability risks, mixed systems are in general optimal, and the variable part of costs that should be borne by the hospital is increasing in the degree of the liability risk. This may at least partially explain why countries like Germany where liability risks are low compared to the US have been more reluctant in switching from FFS to PPS.

Key words: principal-agent-theory, multi-task, health care, hospital compensation schemes, liability law

JEL classification: I11, I18, K13, K32, L51

*Both authors: Department of Economics, Aachen University (RWTH), Templergraben 64, 52062 Aachen, Germany. Email: feess@rwth-aachen.de, ossig@rwth-aachen.de.

1 Introduction

1.1 Motivation and main results

In many countries as in the United States and in Germany, the reimbursement systems for hospitals went through tremendous changes in the last two decades. Until the 1980s, the predominant scheme was a fee-for-service (FFS) system which, roughly speaking, means that all "reasonable" costs were covered by the health insurance based on the hospital's cost report.[1] As emphasized in the literature,[2] this traditional cost reimbursement system had two advantages - there was a strong incentive to provide high quality, and there was no incentive to discriminate among patients. However, it has also long been realized that the FFS sets no incentive to keep the cost down since they are covered anyway. In fact, incentives may be perverse when fees per day are calculated on a full cost basis, because the hospital then has an incentive to increase the length of a patient's stay in hospital since the payment is above marginal costs.[3]
In 1984, the United States were the first country to substitute FFS by a fixed-fee prospective payment system (PPS) based on diagnosis related groups (DRG) in the Medicare system.[4] Under a PPS system, the health insurer pays a fixed price per patient, where this lump sum varies from DRG to DRG in order to account for the major diagnoses, additional complications, gender or age. In the following years, PPS started being widely used, and was adopted in the 1990s in Australia as well as in several other countries. In Germany, hospitals are reimbursed on the basis of DRG since 2003.
Whereas it is uncontroversial that PPS is a high-powered incentive scheme leading to lower costs, both the political discussion and the literature has focussed on two problems that may be caused by PPS. First, PPS sets incentives to select low cost-patients whenever DRG do not perfectly take into

[1] For details on country-specific reimbursement schemes see Ellis/McGuire (1986 and 1988) or Chalkley/Malcomson (1998) for the US system and Simon (2000) for the German system.

[2] See e.g. the overview in Newhouse 1996.

[3] For an overview on empiric results see Ellis/McGuire (1988 and 1996) and Meltzer/Chung (2002).

[4] The Medicare System was established to finance medical care for the elderly in the United States and is the largest single purchaser of health care in the US.

account every single characteristic of the patient. Since each patient may in fact be different, a perfect discrimination would require as many DRG's as patients, so that cream-skimming and dumping incentives can never be fully eliminated. Second, even within one perfectly homogenous DRG, the hospital has incentives to reduce quality when the amount of cost reimbursement is given. Hence, the literature has often concluded that combined systems may be superior (see the literature review below).

In our paper, we ignore the selection problem by restricting our attention to homogenous patients, and focus instead on the impact of different reimbursement schemes on the quality- and the cost cutting-incentive. As for this, we develop a simple multi-task principal-agent model where the hospital (the agent) divides its unobservable effort between the two tasks of quality improvement and cost cutting. Whereas costs are observable and verifiable, we assume that the quality is uncontractable, so that the health insurer (the principal) can only set direct incentives for cost-cutting, but not for quality improvement. In our basic model, we find that either FFS or PPS, but no combined system, can be optimal. This result is based on three assumptions that are often emphasized as being important for a doctor's effort choice: First, even without monetary incentives, the *total* effort is positive since doctors are intrinsically motivated. Second, if there are only incentives for cost-cutting (this will be the case with PPS), the quality effort will nevertheless be positive due to the so called Hippocratic Oath that requires doctors to deliver not only the very basic quality of care. And third, if there are no monetary incentives at all (this is the case with FFS), doctors have a bias towards quality.

Besides intrinsic motivation, there are various quality incentives in reality even within a prospective payment system. Patients may get stochastic quality signals that influence their future demand (reputation effects), high quality hospitals may attract better doctors, and the liability risk for malpractice decreases in the quality effort. In our extended model, we focus on the impact of liability, since the liability risk differs widely between countries (see our discussion in section 5) which allows to compare our theoretical findings to reality. We show that, if the liability regime is neither too weak nor too tough in a sense to be specified in the model, then a combined system is superior. However, FFS can be optimal if liability is of minor importance, and PPS can be optimal if liability plays an important role. Although the result is straightforward from a theoretical point of view (the higher is the liability risk, the higher is the quality incentive, and the less are the negative effects when moving from FFS towards PPS), we believe that it is important to take the different liability regimes into account when thinking about the consequences of the reimbursement schemes adopted in different countries.

1.2 Relation to the literature

Following the general idea of yardstick competition (Shleifer 1985), the early literature[5] has argued that PPS can lead to efficient incentives even if the effort costs of cost reduction are private information. If each hospital is reimbursed for the average marginal costs of all other hospitals, the payment is independent of the own behavior, and incentives will thus be efficient.

In departing from his basic idea, however, the preceding literature has identified at least four drawbacks of PPS. First, if there is unobserved heterogeneity among patients, then some hospitals will make negative profits when a simple yardstick reimbursement scheme is applied. If negative profits need to be avoided, hospitals with a favorable portfolio will earn positive rents, thereby causing shadow costs of public funds.[6] Second, incentives for cream-skimming will arise whenever the classification in different DRG's underestimates the costs of treating morbid patients (see e.g. Allen and Gertler (1991)). A third and related drawback of PPS is dumping and has been analyzed in various papers by Ma.[7] Fourth, PPS may lead to underprovision of treatment intensity if the quality is unobservable (see e.g. Ellis (1997) or Chalkley and Malcomson (1998)). If demand depends upon quality, this effect is mitigated, and may even disappear if the relationship becomes close enough (see Allen and Gertler (1991) and Rogerson (1994)).

With respect to our paper, an important point is under which circumstances corner solutions (i.e. either FFS *or* PPS) or mixed systems are optimal. Since FFS gives no incentive for cost-cutting while PPS may lead to underprovision if quality is (partially) unobservable, the literature has often concluded that mixed reimbursement schemes are optimal (see e.g. Ellis and McGuire (1986) and Rogerson (1994)). Conversely, in the multi-task framework considered in Ma (1994), one can always find a fixed fee per patient within a PPS that

[5] See Newhouse (1996) for an overview.

[6] This is closely related to the more general literature comparing cost-plus regulation to price caps (or low-powered to high-powered incentive schemes).

[7] See Ma (1994), Ma (1997), and Ma (1998).

leads to first best incentives for both cost reduction and quality. The reason is as follows: first, the hospital's incentive for cost reduction is identical to the social planner's one since the variable costs are fully borne by the hospital. Second, the incentive for quality improvement comes from the assumption that demand depends on quality. And the higher the fixed fee per patient, the higher the incentive to increase quality in order to attract more patients.[8] Note that the reason why corner solutions are optimal in our basic model are quite different as we consider only the effort incentive for a *given* set of patients. In reality, the impact of quality on demand clearly varies from field to field[9], so that the two approaches may be considered as being complementary.[10]

To the best of our knowledge, Gal-Or (1999) was the first to analyze the impact of liability on the optimal reimbursement scheme. She focuses on the division of liability between an insurer (or a Managed Care Organization) and the hospital. If the insurer can perfectly control the intensity (quality in our paper) of the treatment, then PPS combined with full liability for the hospital leads to a first best. Otherwise, a mixed system arises, and both quality provision and cost reduction are too low. Although related to our findings, our model differs in focusing on the hospital's preference towards quality and on the fact that the cross partial derivative with respect to the two efforts is positive, thereby leading to a "real" multi-task principal-agent problem which Gal-Or does not take into account. At a more theoretical level, our paper borrows from the literature on multi-task principal-agent problems pioneered by Holmström and Milgrom (1991). Finally, the concept of a vague negligence rule adopted in section 4 has been developed by Schwartz (1998) in the context of accountant's liability.

The remainder of the paper is organized as follows: Section 2 introduces the model. Section 3 derives the insurance company's optimal contract under the assumption that there are no liability payments for malpractice. Section 4 captures liability risks by the concept of a vague negligence rule. Section

[8] Ma (1994) also considers cream skimming, but this is less important when comparing to our paper.

[9] For instance, for cosmetic surgeries, the impact may be high, whereas it is almost neglectable for emergency service.

[10] The lower the impact of quality on demand, the higher the fixed fee per person required to induce efficient effort quality incentives in a framework like Ma's one. Then, the shadow costs of public funds become extremely high, thereby limiting the practical scope of the suggestion.

5 compares our findings to reality and concludes.

2 The model

We consider a hospital treating one representative patient. The patient perfectly fits into a specific diagnosis related group (DRG), so that we do not account for selection problems. The hospital exerts total unobservable effort e, where a fraction q is spend on quality improvement, and a fraction c on cost cutting. Hence, $q + c = e$. The hospital's effort cost function is given by $E(e)$, where $E(e) = 0 \; \forall e \leq \overline{e}$, $\frac{dE(e=\overline{e})}{de} = 0$, $\frac{dE}{de} > 0$ and $\frac{d^2E}{de^2} > 0 \; \forall e > \overline{e}$. The underlying idea is that variable effort costs are zero up to some effort level $\overline{e}$ because the medical staff is intrinsically motivated, and hence performs effort $\overline{e}$ voluntarily, i.e. without monetary incentives. Up to $\overline{e}$, there are only fixed (opportunity) costs that we can safely normalize to zero. Our assumptions on $E(e)$ ensure that $\frac{\partial^2 E}{\partial q \partial c} > 0 \; \forall e > \overline{e}$, so that effort costs of providing quality are increasing in the effort spend on cost cutting (and vice versa).
The quality effort q leads to expected treatment benefit of $Q(q)$ where $\frac{dQ}{dq} > 0$ and $\frac{d^2Q}{dq^2} < 0$. Treatment costs C depend on c where $\frac{dC}{dc} < 0$ and $\frac{d^2C}{dc^2} > 0$. Note the difference between E and C: E are the hospital's managerial effort costs, while C are monetary treatment costs that may be covered by a health insurance company. We assume that C is observable and verifiable, while the reimbursement scheme can not be made contingent on E and Q.[11] Assuming that all participants are are risk neutral, social welfare can simply be written as

$$SW = Q(q) - C(c) - E(e). \tag{1}$$

We denote first best levels by superscript "f" and assume that $q^f, c^f > \overline{e}$ to guarantee that the effort chosen without incentives is always too low from a social point of view. Together with our assumptions on $Q(q)$, $C(c)$ and $E(e)$ and recalling that $e = c + q$, this ensures an interior solution for the first best implicitly given by

$$\frac{dQ}{dq} = \frac{\partial E(q, c^f)}{\partial q} \tag{2}$$

[11] In our model, it does not make a difference whether E and Q are unobservable or observable but uncontractible.

and

$$-\frac{dC}{dc} = \frac{\partial E(c, q^f)}{\partial c}. \tag{3}$$

Without monetary incentives, the hospital chooses $\overline{e}$ and faces only fixed effort costs normalized to zero. Then, we assume that the hospital divides $\overline{e}$ on q and c according to $\overline{\gamma} = \frac{\overline{q}}{\overline{e}}$, i.e. $\overline{\gamma}$ is the percentage of costless effort exerted for quality if there are no monetary incentives. Defining $\overline{\gamma}^f$ as the socially optimal division of $\overline{e}$, we assume that $\overline{\gamma} > \overline{\gamma}^f$. This expresses that the hospital has a bias towards quality.[12]
Finally, we assume that there is a lower bound for quality given by $\widetilde{q} < \overline{q}$ even if the insurer sets monetary incentives to spend all effort on cost cutting. This minimum quality effort may arise for different reasons, for instance due to the Hippocratic Oath, according to more generally motivated moral considerations, or coming from liability risks. In section 3, we will simply assume that $\widetilde{q}$ is exogenously given.
We restrict our attention to linear payment schemes where the health insurer covers a fraction α of treatment costs C and pays a fixed amount β. The traditional fee-for-service system (FFS) can then be interpreted as $\alpha = 1$, since the insurer covers costs C regardless of their amount. By contrast, $\alpha = 0$ captures PPS because the hospital receives only a fixed fee β. The lower α, the "higher powered" is the reimbursement scheme with respect to the cost-cutting incentive. Ignoring agency problems between the health insurer and the insure,[13] the hospital's utility function is

$$U = -(1-\alpha)\, C(c) - E(e) + \beta, \tag{4}$$

while the health insurer's expected profit is

$$\Pi = Q(q) - \alpha C - \beta. \tag{5}$$

Finally, the hospital's reservation utility is normalized to 0, and the hospital is protected by limited liability in the sense that up-front payments from the

[12]Note that the quality will nevertheless be below the optimal one since $\overline{q} = \overline{\gamma e} \leq \overline{e} < q^f$.

[13]This implies that the insurer perfectly cares about quality. As common in the literature, the treatment quality being part of the insurer's profit can be justified due to two reasons. First, the relationship between the insurer and the insuree is often longterm, so that the insurer benefits from cured patients. Second, in most European countries health insurers are regulated or even held by the government (e.g. the NHS in Britain).

hospital to the insurer are excluded ($\beta \geq 0$). Our assumptions are clearly quite simple, but they will allow to shed some light on the following two questions: first, why do we often observe either FFS or PPS, but not a mixed system that is usually derived as being optimal in the literature? And second, how does the ranking between the two systems depend on the quality distortion incentive $\overline{\gamma}$, on the minimum quality $\widetilde{q}$ chosen in any case and on the effort cost function?

3 The insurer's optimal contract choice

We now turn to the insurance company's optimal contract defined as $\Gamma^* = (\alpha^*, \beta^*)$[14]. Let us start with

Lemma 1. (i) *In the profit maximizing contract* Γ^**, the hospital gets no rent.* (ii) Γ^* *maximizes social welfare.*

All Proofs are in the Appendix.
The reason for Lemma 1 is that the hospital's utility would always be negative without up-front payment as long as $\alpha \leq 1$, since total effort costs E and a part $(1-\alpha)$ of treatment costs C are then borne by the hospital. Hence, the hospital's limited liability has no impact on the optimal contract since $\beta^* > 0$ anyway, and the preferences of the health insurer and the regulator are aligned.
This given two cases, $\alpha = 1$ and $\alpha < 1$ need to be distinguished to describe the hospital's behavior. The reason why we differentiate between these two cases is that for all $\alpha < 1$, $\widetilde{q}$ is chosen.

Case 1. FFS ($\alpha = 1$). Here, we have $U = -E(e) + \beta$, and the hospital chooses $\overline{q}$ and $\overline{c}$. This leads to

$$SW_{\alpha=1} = \Pi_{\alpha=1} = Q(\overline{q}) - C(\overline{c}). \tag{6}$$

Two inefficiencies arise for $\alpha = 1$: total effort is below the first best one, and the effort division is distorted towards quality.

[14] In the following equilibrium values wear an asterisk.

Case 2. Mixed system or PPS ($\alpha < 1$). For $\alpha < 1$, the hospital's expected utility function is

$$U = -(1-\alpha)\,C(c) - E(e) + \beta. \tag{7}$$

Since there are monetary incentives to reduce treatment costs while there are no incentives to enhance treatment quality Q ($\frac{\partial U}{\partial q} = -\frac{\partial E(\widetilde{q},c)}{\partial q} < 0\ \forall e > \overline{e}$), the hospital chooses the minimum quality effort $\widetilde{q}$ while the equilibrium cost reduction effort c^* is given by

$$-(1-\alpha)\,\frac{dC}{dc} = \frac{\partial E(\widetilde{q},c)}{\partial c}. \tag{8}$$

Since $-(1-\alpha)\,\frac{dC}{dc} > 0\ \forall c$ and $\frac{\partial E(e=\overline{e})}{\partial c} = 0$, there is an interior solution where $c^* + \widetilde{q} > \overline{e}\ \forall \alpha < 1$. In other words, as long as there is no full FFS, the hospital will always choose effort above $\overline{e}$.
The health insurer maximizes

$$SW_{\alpha<1} = Q(\widetilde{q}) - C(c^*) - E(c^* + \widetilde{q}) \tag{9}$$

since β will be adjusted accordingly so that $U(\Gamma^*) = 0$.[15] Since $q^* = \widetilde{q}$ $\forall \alpha < 1$, the second best optimal cost reduction effort c^* is implicitly given by

$$-\frac{dC}{dc} = \frac{\partial E(\widetilde{q},c)}{\partial c}. \tag{10}$$

Comparing Eqn. (10) to the hospital's first order condition given by Eqn. (8) shows that both are identical for $\alpha^* = 0$, i.e. when the hospital is not compensated for the treatment costs at all. Then, there are again two inefficiencies compared to the first best: First, the quality is too low since only $\widetilde{q}$ is chosen. Second, the cost-cutting effort is too high compared to the first best level ($c^*_{\alpha=0} > c^f$) since the effort cost function is convex, and because $q^* < q^f$. Since this (marginal) cost effect does not only hold for the hospital's objective function but also for the (second best) social welfare function, the insurer has no reason to reduce this effect by choosing $\alpha > 0$, thus $SW_{\alpha<1} = SW_{\alpha=0}$.

[15] $\beta^*_{\alpha<1}$ is implicitly given by $U^* = -(1-\alpha)\,C(c^*) - E(c^* + \widetilde{q}) + \beta^* = 0$.

Comparison. Our results demonstrate that, in the basic model, either FFS or PPS but no interior solution can be optimal. Whenever the insurer does not fully cover the treatment costs ($\alpha < 1$), the hospital will only choose $\widetilde{q}$ since there is no benefit from choosing higher quality effort. But then, $\alpha = 0$ is chosen as to fully internalize the treatment costs. The only method to implement a higher quality is to cover all costs (FFS), which leads to $\overline{q} > \widetilde{q}$. To compare the two corner solutions, define $\Delta SW \equiv SW_{\alpha=0} - SW_{\alpha=1}$:

$$\Delta SW = Q(\widetilde{q}) - C(c^*_{\alpha=0}) - E(\widetilde{q}, c^*_{\alpha=0}) - (Q(\overline{q}) - C(\overline{c})) \,. \tag{11}$$

Taking into account that $\widetilde{q} < \overline{q} < q^f$ and that $\overline{\gamma} > \gamma^f$, we can summarize in

Proposition 1. (i) *Either FFS (*$\alpha = 1$*) or PPS (*$\alpha = 0$*) is second best-optimal.* (ii) *FFS leads to inefficiently low total effort, and the division of effort is distorted towards quality.* (iii) *PPS leads to even lower quality effort, and to inefficiently high cost reducing effort.* (iv) *That PPS is superior is the more likely* (a) *the higher* $\widetilde{q}$*,* (b) *the higher* $\overline{\gamma}$*, and* (c) *the less steep the* $E(e)$*-function.*

All results are fairly intuitive. With PPS ($\alpha = 0$), only the minimum quality effort $\widetilde{q}$ is chosen, and the higher this effort, the better. However, this leads to an excessive cost reducing effort above the first best level. By contrast, covering all costs (FFS, $\alpha = 1$) leads to a distortion towards quality expressed by $\overline{\gamma} > \gamma^f$, and to inefficiently low overall effort. The lower $\overline{\gamma}$, the lower the first inefficiency. Finally, the steeper the effort cost function, the less important it is to implement a high overall effort by choosing $\alpha = 0$.

4 Liability for malpractice

The main result of section 3 was that only corner solutions as often observed in reality (FFS *or* PPS) can be optimal if the quality effort $\widetilde{q}$ is the same whenever the cost reduction incentive is positive at all ($\alpha < 1$). In other words, corner solutions are optimal if the minimum quality effort level is exogenously given, for instance according to the Code of Ethics.
In this section, we investigate what happens if $\widetilde{q}$ is an endogenous variable following from the hospital's optimization. This does not mean that we neglect the hospital's intrinsic motivation for quality, but rather expresses the natural view that the hospital nevertheless reacts to shifts in opportunity

costs. Then, the quality effort q will depend on the incentive to spend effort on cost reduction c (and thus on α) because, the higher c, the higher are the costs of q due to the convexity of the cost function $E(e)$. As for this, we now assume that the hospital's risk of being held liable for malpractice decreases in q.

We model the liability risk in a simple way. In all countries, liability for malpractice is negligence based, but the negligence standard adopted by the court is obviously imperfect in the sense that it cannot be perfectly anticipated ex ante. Thus, the concept of a *vague negligence rule* initially developed in the literature on liability for accounting firms[16] seems to be appropriate. Let us define $L(q)$ as the *expected* liability payments as a function of q. To make our points as easy as possible, we assume that $L(q) = L^{\max} - \tau \cdot Q(q)$, so that there is a simple linear relationship between the expected quality and the variable part of the expected liability payment. $\tau < (>)1$ means that the impact of the quality effort q on the expected liability payment is lower (higher) than the impact on the expected quality itself, so that we speak of a weak (a tough) vague negligence rule if $\tau < 1 (\tau > 1)$.[17] Clearly, $\tau < 1$ seems to be more realistic.

Under these circumstances, the hospital's expected utility function becomes

$$U = -(1-\alpha)\,C(c) - L^{\max} + \tau Q(q) - E(e) + \beta. \tag{12}$$

Since β will again be adjusted such that $U = 0$, the hospital accepts the equilibrium contract and gets no rent. Let us first consider FFS, i.e. $\alpha = 1$. Then, $U = -L^{\max} + \tau Q(q) - E(e) + \beta$, and instead of choosing $\overline{q}$ and $\overline{c}$ as without liability, the hospital increases its quality effort up to

$$\tau \frac{dQ}{dq} - \frac{\partial E(c^*, q)}{\partial q} = 0 \tag{13}$$

where our assumption that $\overline{e} < q^f$ ensures that $c^* = 0$. Hence, for $\alpha = 1$, introducing vague negligence leads to a further decline of the cost reducing effort since $c^* = 0$ is chosen instead of $\overline{c}$. This leads to

[16] See Schwartz (1998).

[17] The advantage of this simple specification is that $\frac{dL}{dq} = -\tau \frac{dQ}{dq}$ which simplifies the proofs for the comparative statics. More generally, one could write $L(q) = l(q)Q(q)$ and hence $\frac{dL}{dq} = \frac{dl}{dq}Q(q) + l(q)\frac{dQ}{dq}$ which is more tedious for the comparative statics without leading to additional insights. Of course, our specification could also be interpreted as a partial strict liability rule since, for $\tau = 1$, we simply have ordinary strict liability.

Lemma 2. *Suppose FFS applies* ($\alpha = 1$). (i) *If the liability rule is weakly tough* ($\tau \geq 1$), *then* $q^* > q^f$. (ii) *If the liability rule is weak* ($\tau < 1$), *then* $q^* \lesseqgtr q^f$.

The intuition for Lemma 2 is straightforward. For $\tau = 1$, the quality risk is fully internalized. However, as $c^* = 0$, and since marginal costs for q are increasing in c^*, the quality effort will be above the first best level. This incentive to deliver too high quality is aggravated if $\tau > 1$. For $\tau < 1$, however, there are countervailing effects leading to $q^* \lesseqgtr q^f$. On the one hand, there is an incentive to underinvest in quality as the liability risk is below the quality risk. On the other hand, q is ceteris paribus higher than q^f as $c^* = 0 < c^f$.

Let us know consider the case of $\alpha < 1$ to check if an interior solution may now be optimal. For $\alpha < 1$, the first order conditions for the hospital's effort choice are

$$\tau \frac{dQ}{dq} - \frac{\partial E(c^*, q)}{\partial q} = 0 \tag{14}$$

$$-(1-\alpha)\frac{dC}{dc} - \frac{\partial E(q^*, c)}{\partial c} = 0. \tag{15}$$

Note that for $\tau = 1$ and $\alpha = 0$, we have $c^* = c^f$ and $q^* = q^f$ since both the quality risk and the cost effect are perfectly internalized. However, the vague negligence will generally be imperfect in the sense that $\tau \neq 1$. Let us start with

Lemma 3. (i) $\frac{dq^*}{d\alpha} > 0$, (ii) $\frac{dc^*}{d\alpha} < 0$.

Lemma 3 simply expresses that the incentives for quality (for cost-cutting) are increasing (decreasing) if the insurer covers a higher percentage of the costs. This leads to

Proposition 2. (i) $\frac{d\alpha^*}{d\tau} \leq 0$. ($ii$) *If the vague negligence rule is weak* ($\tau < 1$), *then* $0 < \alpha^* \leq 1$. (iii) *If the vague negligence rule is weakly tough* ($\tau \geq 1$) *then* $\alpha^* < 1$.

The higher τ, the higher the quality incentive, and this is balanced by choosing a low α (see Lemma 2). This explains part (i) of the Proposition. Part (ii) says that, in the realistic case where the liability rule does not internalize the total expected quality risk, the insurer covers a fraction $\alpha > 0$ of the

treatment costs. Hence, *PPS cannot be optimal for* $\tau < 1$. The reason is as follows: For $\alpha = 0$ and $\tau < 1$, we know that, if $c^* = c^f$ were chosen, we would have $q^* < q^f$. But the lower q, the higher c due to the convex effort cost function $E(e)$. Hence, for $\alpha = 0$, the hospital would choose too high a cost reduction effort ($c^* > c^f$) and too low a quality effort ($q^* < q^f$). To mitigate this effect, the insurer sets $\alpha > 0$ which reduces c^* and increases q^* via the marginal cost effect. Note the difference to the former section where PPS could be optimal, since we did not adopt the (opportunity)cost principle when deriving $\widetilde{q}$. But whenever this is taken into account and $\tau < 1$, the insurer should cover part of the hospital's cost.
To see why $\alpha = 1$ can be optimal for $\tau < 1$, note that $\alpha = 1$ leads to the maximum reachable quality effort which is again due to the fact that $c = 0$ for $\alpha = 1$ combined with the convexity of the cost function. If the liability rule is "very weak" ($\tau \to 0$), then $\alpha = 1$ may be optimal as we are actually back in the no-liability situation of section 3.[18] However, the health insurer cannot implement a first best by adjusting α appropriately, since for $\tau < 1$ *and* $\alpha > 0$, the effort division could be optimally adjusted, but the overall effort will definitely be too low as only part of the quality risk and part of costs are internalized.
For a tough vague negligence rule ($\tau > 1$), we get a similar result (part (*iii*) of the Proposition). If the insurer covered all costs, then we would have $q^* > q^f$ due to $\tau > 1$. Furthermore, we would have $c^* = 0$, and these inefficiencies can be mitigated by choosing $\alpha < 1$. Hence, FFS can never be optimal if the liability rule is tough, and PPS can only be optimal if the marginal impact of the quality effort on the liability risk is higher than on quality itself.
Summing up, by contrast to the case without liability, interior solutions for the cost reimbursement share α^* arise if the liability rule is neither "very" weak nor "very" tough. The tougher the liability rule, the smaller the part of the treatment costs the insurer should cover.

5 Discussion

In a simple model about the optimal cost sharing between a health insurer and a hospital, we have found the following: With full cost reimbursement

[18]To avoid misunderstandings, let us emphasize that the definition of a "weak" rule is *not* that the overall liability risk is small ($L^{\max}$ may be high), but that the impact of quality effort q on the liability exposure is small.

(FFS), both the cost-reducing effort and the quality effort are too low, but quality is relatively high (although below the first best one) due to the doctors' preferences. With PPS, costs are reduced, but the quality is even lower than with FFS. Mixed systems are strictly inferior to PPS, because the hospital will only provide the minimum quality as soon as there is cost sharing, and because the incentive for cost-cutting is reduced. Hence, our model offers a potential rationale for the fact that many countries have radically switched from FFS to PPS.

In our extended model with liability risks, one of our results is that FFS can be superior with low liability risks and PPS with high liability risks. Hence, one should expect that countries with high liability risks (thereby providing high quality incentives) have a higher incentive to switch from FFS to PPS. This seems to be the case when comparing the reimbursement schemes in Germany and the US, where PPS plays an increasingly important role since the mid 1980s, whereas PPS was implemented in Germany only very recently. Although liability is negligence based in both countries, there are at least two factors leading to the result that the liability risk for malpractice is in fact much higher in the US.

First, although the burden of proof is allocated to the patient in both countries, the evidence required to trigger liability is considerably higher in Germany.[19] In Germany, the so called "prima facie"-rule requires that the mistreatment was definitely apt to have caused the damage in some way. In the USA, the "res-ipsa-loquitur" rule means that the potential injurer is exonerated if he proves that a different result is at least as likely as the one presented by the plaintiff. From an economic point of view, the different rules can be interpreted in the sense that the probability threshold that the damage has been caused by malpractice is higher in Germany.

The second difference relates to the amounts awarded to plaintiffs. In the USA, the average amount was approximately \$271.000 in 2001,[20] whereas the scale of damages resulting from successful tort litigation in Germany seems to be much lower. Although there is no official source that collects and analyzes malpractice data in Germany, one of the major insurers for German doctors' liability risks, the DBV-Winterthur, estimated the average liability payment to be about \$25.000 in 1999. Furthermore, even though the civil legislation in Germany does not assign a limit for the amount of damage

[19] See e.g. Faure/Koziol (2001) and Kennedy/Grubb (2000).

[20] See National Practitioner Data Bank Annual Report 2001.

to be paid, the maximum non-pecuniary payments for pain and suffering are hardly ever higher than \$250.000).[21] Conversely, as a reaction to the so called mega awards up to \$100.000.000 that have lead to physicians leaving the market, damage caps of \$250.000 or \$350.000 have been introduced in some US-states.
Summing up, these factors indicate that the liability system for malpractice and its use in the USA indeed provide strong incentives for quality effort, so that switching from FFS to PPS is likely to be more efficient compared to systems as the one in Germany. For mixed systems that we have found to be optimal if the liability risk is neither too very strong nor very weak, our model suggests that the cost sharing parameter should be higher in countries like Germany compared to countries like the USA, and that the liability risk should be taken into account when designing the reimbursement scheme.

Appendix

Proof of Lemma 1. Part (i). First, $\alpha^* \leq 1$ since $\alpha > 1$ would lead to $c < \overline{c}$, thereby reducing Π at no benefit. Second,

$$U = -(1-\alpha)\, C(c) - E(e) + \beta < \beta \ \forall \alpha < 1 \tag{16}$$

which requires $\beta > 0$ to fulfill the hospital's participation constraint $U \geq 0$. And since β has no incentive effect at all, the insurer sets β such that $U(\Gamma^*) = 0$. Part (ii). Since $\Pi(\Gamma^*) = \Pi(\Gamma^*) + U(\Gamma^*) = SW(\Gamma^*)$, the insurance company maximizes social welfare. ■

Proof of Proposition 1. Parts (i) to (iii) are obvious from above. Part (iv). Part (a). $\frac{d(\Delta SW)}{d\widetilde{q}} = \frac{dQ}{d\widetilde{q}} > 0$ since $\frac{\partial C}{\partial c^*} = \frac{\partial E^*}{\partial c^*}$ $\forall \widetilde{q}$ due to (second-best) optimality. (b) $\frac{d(\Delta SW)}{d\overline{\gamma}} = -\frac{d(Q(\overline{q})-C(\overline{c}))}{d\overline{\gamma}} < 0$ as $\frac{d(Q(\overline{q})-C(\overline{c}))}{d\overline{\gamma}} > 0$ and $\overline{\gamma} > \gamma^f$ by definition of γ^f, and by the concavity of $Q(q)$ and $C(c)$.(c) Define a set of functions $\widetilde{E}(e) = \phi_i E(e_i)$ with $\phi_i > 1$. Consider any two functions $\widetilde{E}(e_1) = \phi_1 E(e_1)$ and $\widetilde{E}(e_2) = \phi_2 E(e_2)$ where $\phi_1 < \phi_2$ and $E(e_1) = E(e_2)$ $\forall e_1 = e_2$. Define e_2^* as the optimal e_2, and $SW_2^*(e_2^*)$ as maximum welfare. Define e_1^* and $SW_1{}^*(e_1^*)$ analogously. Since $-\frac{dC(c^*_{\alpha=0})}{dc} = \phi_i \frac{\partial E(\widetilde{q}, c^*_{\alpha=0})}{\partial e_i}$, we have $e_1^* > e_2^*$. Assume that, for ϕ_1, $e_1 = e_2^*$ is chosen instead of e_1^*, which is clearly not optimal. But even then,

$$SW_1(e_1 = e_2^*) = \int_0^{e_2^*} \left(-\frac{dC(c^*_{\alpha=0})}{dc} - \phi_1 \frac{\partial E(\widetilde{q}, c^*_{\alpha=0})}{\partial e_1} \right) de_1 \tag{17}$$

[21] See e.g. Faure/Koziol (2001) p.52.

$$> SW_2^*(e_2^*) = \int_0^{e_2^*} \left(-\frac{dC(c^*_{\alpha=0})}{dc} - \phi_2 \frac{\partial E(\widetilde{q}, c^*_{\alpha=0})}{\partial e_2} \right) de_2$$

since $\phi_1 < \phi_2$. By definition of optimality, $SW_1^*(e_1^*) > SW_1(e_1 = e_2^*) > SW_2^*(e_2^*)$, and hence $SW_1^*(e_1^*) > SW_2^*(e_2^*)$. ■

Proof of Lemma 2. Implicit differentiation of Eqn. (13) yields $\frac{dq^*}{d\tau} = -\left(\frac{\frac{dQ}{dq}}{\tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c^*,q)}{\partial q^2}} \right)$. Since $\frac{dQ}{dq} > 0$ and since $\tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c^*,q)}{\partial q^2} < 0$ due to the fact that Eqn. (13) expresses a maximum, we have $\frac{dq^*}{d\tau} > 0$. Furthermore, $\frac{dq^*}{dc} = -\left(\frac{-\frac{\partial^2 E(c^*,q)}{\partial q \partial c}}{\tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c^*,q)}{\partial q^2}} \right)$. From $\frac{\partial^2 E(c^*,q)}{\partial q \partial c} > 0$, it follows $\frac{dq^*}{dc} < 0$. Since $q^*(e^f, \tau = 1) = q^f$, part (i) of the Lemma follows. For part (ii), we have $q^* < (>) q^f$ if the c-effect (the τ-effect) dominates. ■

Proof of Lemma 3.

Applying the implicit function theorem to

$$\tau \frac{dQ}{dq} - \frac{\partial E(c,q)}{\partial q} = 0 \tag{18}$$

and

$$-(1-\alpha)\frac{dC}{dc} - \frac{\partial E(q,c)}{\partial c} = 0 \tag{19}$$

yields

$$\frac{dq^*}{d\alpha} = -\frac{\begin{vmatrix} 0 & -\frac{\partial^2 E(c,q)}{\partial q \partial c} \\ \frac{dC}{dc} & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}}{\begin{vmatrix} \tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & -\frac{\partial^2 E(c,q)}{\partial q \partial c} \\ -\frac{\partial^2 E(q,c)}{\partial c \partial q} & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}} \tag{20}$$

and

$$\frac{dc^*}{d\alpha} = -\frac{\begin{vmatrix} \tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & 0 \\ -\frac{\partial^2 E(q,c)}{\partial c \partial q} & \frac{dC}{dc} \end{vmatrix}}{\begin{vmatrix} \tau \frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & -\frac{\partial^2 E(c,q)}{\partial q \partial c} \\ -\frac{\partial^2 E(q,c)}{\partial c \partial q} & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}} \tag{21}$$

Assuming an interior solution, the determinate of the denominator is positive, and we restrict our attention to the numerator. For Eqn. (20) we get

$$\frac{dC}{dc}\frac{\partial^2 E(c,q)}{\partial q \partial c} < 0 \tag{22}$$

since $\frac{dC}{dc} < 0$ and $\frac{\partial^2 E(c,q)}{\partial q \partial c} > 0$. Thus $\frac{dq^*}{d\alpha} > 0$. The numerator of (21) yields

$$\frac{dC}{dc}\left[\tau\frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2}\right] > 0 \tag{23}$$

since $\frac{dC}{dc} < 0$ and $\left[\tau\frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2}\right] < 0$ for Eqn. (18) expressing a maximum. ■

Proof of Proposition 2.
We proceed in two steps:
(1) We show that $\frac{dq^*}{d\tau} > 0$ and $\frac{dc^*}{d\tau} < 0$ $\forall \alpha < 1$. (2) We show that $\frac{dq^*}{d\alpha} > 0$ and $\frac{dc^*}{d\alpha} < 0$ $\forall \tau$. Note that steps (1) and (2) together are sufficient to prove part (*i*) of the Proposition: the higher τ, the higher is q^* and the lower is c^*, and since $\frac{dq^*}{d\alpha} > 0$ and $\frac{dc^*}{d\alpha} < 0$, it follows that high q^* and low c^* are countervailed by choosing low α. To prove part (*ii*) of the Proposition, one needs additionally take into account that $q^* = q^f$ and that $c^* = c^f$ if $\alpha = 0$ *and* $\tau = 1$. Thus, it follows from steps (1) and (2) that $q^* < q^f$ and $c^* > c^f$ for $\alpha = 0$ if $\tau < 1$. By definition of optimality, social welfare is then always higher if $\alpha > 0$ as this leads to lower c^* and to higher q^*. Part (*iii*) follows analogously by considering the case $\tau \geq 1$. ■
Step (1). Applying the implicit function theorem to

$$\tau\frac{dQ}{dq} - \frac{\partial E(c,q)}{\partial q} = 0 \tag{24}$$

and

$$-(1-\alpha)\frac{dC}{dc} - \frac{\partial E(q,c)}{\partial c} = 0 \tag{25}$$

yields

$$\frac{dq^*}{d\tau} = -\frac{\begin{vmatrix} \frac{dQ}{dq} & -\frac{\partial^2 E(c,q)}{\partial q \partial c} \\ 0 & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}}{\begin{vmatrix} \tau\frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & -\frac{\partial^2 E(c,q)}{\partial q \partial c} \\ -\frac{\partial^2 E(q,c)}{\partial c \partial q} & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}} \tag{26}$$

and

$$\frac{dc^*}{\partial\tau} = -\frac{\begin{vmatrix} \tau\frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & \frac{dQ}{dq} \\ -\frac{\partial^2 E(q,c)}{\partial c\partial q} & 0 \end{vmatrix}}{\begin{vmatrix} \tau\frac{d^2Q}{dq^2} - \frac{\partial^2 E(c,q)}{\partial q^2} & -\frac{\partial^2 E(c,q)}{\partial q\partial c} \\ -\frac{\partial^2 E(q,c)}{\partial c\partial q} & -(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c} \end{vmatrix}} \tag{27}$$

Assuming an interior solution, the determinate of the denominator are positive, and we restrict our attention to the determinants of the numerators. For the numerator of equation (26) we get

$$\frac{dQ}{dq}\left[-(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c}\right] < 0 \tag{28}$$

since $\frac{dQ}{dq} > 0$ and $\left[-(1-\alpha)\frac{d^2C}{dc^2} - \frac{\partial^2 E(q,c)}{\partial^2 c}\right] < 0$ which follows from $\tau\frac{dQ}{dq} - \frac{\partial E(c,q)}{\partial q} = 0$ indicating a maximum. Hence $\frac{dq^*}{d\tau} > 0$. The numerator of (27) is

$$\frac{dQ}{dq}\frac{\partial^2 E(q,c)}{\partial c\partial q} > 0 \tag{29}$$

since $\frac{dQ}{dq} > 0$ and $\frac{\partial^2 E(q,c)}{\partial c\partial q} > 0$. Thus, $\frac{dc^*}{d\tau} < 0$.

Step (2). see the Proof of Lemma 3.■

References

[1] Allen R. and P. Gertler (1991): Regulation and the Provision of Quality to Heterogenous Consumers: The Case of Prospective Pricing of Medical Services, *Journal of Regulatory Economics*, 3, 361-375.

[2] Chalkley, M. and J. M. Malcomson (1998): Government Purchasing of Health Services, *Discussion Paper in Economics and Econometrics*, University of Southampton.

[3] Ellis, R. (1997), Creaming, skimping and dumping: Provider competition on the intensive and extensive margins, *Journal of Health Economics*, 17(5): 537-555.

[4] Ellis, R. and T. G. McGuire (1986): Provider Behavior under Prospective Reimbursement: Cost Sharing and Supply, *Journal of Health Economics*, 5, 129-151.

[5] Ellis, R. and T. G. McGuire (1988), Insurance Principles and the Design of Prospective Payment Systems, *Journal of Health Economics*, 7, 215-237.

[6] Ellis, R. and T. G. McGuire (1996): Hospital Response to Prospective Payment: Moral Hazard, Selection and Practice-style effects, *Journal of Health Economics*, 15, 257-277.

[7] Faure, M. and H. Koziol (2001): Cases on Medical Malpractice in a Comparative Perspective, Tort and Insurance Law, Vol. 1, Springer, Berlin.

[8] Gal-Or, E. (1999): Optimal Reimbursement and Malpractice Sharing Rules in Health Care Markets, *Journal of Regulatory Economics*, 16, 237-265.

[9] Holmström, B. and P. Milgrom (1991): Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics, and Organization*, 7, 24-52.

[10] Kennedy, I. and A. Grubb (2000): Medical Law, Butterworths Law.

[11] Ma, C.-T. A. (1994): Health Care Payment Systems: Cost and Quality Incentives, *Journal of Economics and Management Strategy*, 3(1), 93-112.

[12] Ma, C.-T. A. (1997): Cost and Quality Incentives in Health Care: Altruistic Providers, Boston University, Department of Economics.

[13] Ma, C.-T. A. (1998): Health Care Payment Systems: Cost and Quality Incentives: A reply, *Journal of Economics and Management Strategy, 7, 139-142*.

[14] Meltzer, D. and J. Chung (2002): Effect of Competition under Prospective Payment on Hospital Costs among High- and Low-Cost Admissions: Evidence from California, 1983 and 1993, *Frontiers in Health Policy Research*, 5, 53-102.

[15] Newhouse, J. P. (1996): Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection, *Journal of Economic Literature*, 1236-1263.

[16] Rogerson, W. P. (1994): Choice of Treatment Intensity by a Nonprofit Hospital under Prospective Pricing, *Journal of Economics and Management Strategy*, 3(1), 7-51.

[17] Schwartz, R. (1998): Auditors' Liability, Vague Due Care, and Auditing Standards, Review of Quantitative Finance and Accounting, 11 (2), 183-207.

[18] Shleifer, A. (1985): A Theory of Yardstick Competition, *RAND Journal of Economics*, 16 (3), 319-327.

[19] Simon, M. (2000): Neue Krankenhausfinanzierung - Experiment mit ungewissem Ausgang: Zur geplanten Umstellung auf ein DRG-basiertes Fallpauschalensystem, *Publication series of the research unit of Public Health*, Wissenschaftszentrum Berlin für Sozialforschung.