ANNALS OF ECONOMICS AND STATISTICS
NUMBER 101/102, JANUARY/JUNE 2011

Optimal Health Care Contract under Physician Agency

Philippe CHONÉ
CREST-LEI and CNRS UMR 2773

Ching-to Albert MA
Boston University

We model asymmetric information arising from physician agency and its effect on the design of payment and health care quantity. The physician aims to maximize a combination of physician profit and patient benefit. The degree of substitution between profit and patient benefit in the physician agency is the physician's private information, as is the patient's intrinsic valuation of treatment quantity. The equilibrium mechanism depends only on the physician agency parameter, and exhibits extensive pooling, with prescribed quantity and payment being insensitive to the agency characteristic or patient's actual benefit. The optimal mechanism is interpreted as managed care where strict approval protocols are placed on treatments.*

I. Introduction

The economics of the health market is concerned with the interaction between insurers, consumers and providers. In this trilateral relationship, the nexus between a patient and a physician is arguably the most fundamental and complicated. The term "Physician Agency" has been used in the literature to refer to a range of issues arising from the influence of physicians on health care use, (MCGUIRE [2000]). Yet researchers have not reached a consensus on the formal model of physician agency. The reason perhaps originates from our suspicion of a pure profit maximization paradigm to model physician agency. We tend to believe that physician-patient interactions are influenced by factors such as power, motivation, medical training and current practice, ethics, and altruism.

Once economists depart from a pure profit-maximization approach, it is unclear what is the most compelling alternative. The literature, both theoretical and empirical, has somehow gyrated toward a pragmatic but natural assumption: physician-patient interaction leads to physician objectives including both physician profits and patient benefits. "We assume that the physician maximizes the sum of his income and patient benefit," is used frequently. In fact, the terms "perfect agency" and "imperfect agency" are often used to mean the extent to which the patient's benefit counts towards physician preferences. Furthermore, in empirical studies, researchers often include provider fixed-effects to control for practice styles, presumably because they believe that financial incentives alone cannot capture the full extent of provider behaviors.[1] In this paper we study the effect of physician agency on health care contracts.

1. Here is a sample of papers using some form of this assumption: CHALKLEY and MALCOMSON [1998], DRANOVE and SPIER [2003], DUSHEIKO *et. al.* [2006], ELLIS and MCGUIRE [1986;1990], MA [1998], MA and RIORDAN [2002], NEWHOUSE [1970], ROCHAIX [1989], ROGERSON [1994].

*JEL: D82, I1, I10, L15 / KEY WORDS: Physician Agency, Altruism, Optimal Payment, Health Care Quantity, Managed Care

Our model consists of a consumer whose health benefits from a treatment may vary due to her medical conditions or preferences. When she seeks treatment from a physician, an agency relationship begins. Physician agency is defined by physician preferences which weigh his profit and the patient's health benefit. Physician agency is complex, and influenced by physicians' medical training, degree of altruism, concern for patients, past experiences, personalities, cultural backgrounds, and local area practice guidelines. We model physician agency complexity by allowing the preference weight on patient benefit to vary. For example, a physician who is more altruistic towards his patients will have a higher agency weight on patient benefit.

Our model incorporates two dimensions of uncertainty: the consumer's health benefit, and the physician agency weight on benefit. Patients have different health benefits, and physicians have diverse preferences. Our approach relaxes the usual assumption that physician agency puts a fixed weight on patient benefit. Furthermore, we let the physician possess private information on both the consumer's health benefit and the physician agency weight.

Our model belongs to the class of multi-dimensional adverse selection problems (see ARMSTRONG and ROCHET [1999], and ROCHET and CHONÉ [1998]). One suspects that a physician cares more about a more severely sick patient who benefits more from treatment. Accordingly, one may regard the two dimensions as positively correlated. Our setup is consistent with this view, but it is sufficiently general to include the extreme cases of perfect and zero correlations between a patient's benefit and agency weight. Our model is most relevant to intermediate cases of imperfect correlation in which one dimension cannot be inferred from the other.

An insurer or managed care company designs a payment and quantity contract for the physician. We will assume that the consumer has full insurance. A general mechanism will be studied. The physician, who behaves according to the preferences in the physician agency, picks from a menu of quantity-payment pairs, indexed by the consumer's intrinsic benefit and the physician agency's weight on consumer benefit. According to the revelation principle, these schedules are equivalent to a direct mechanism where the physician reports his private information, and where it is an equilibrium for him to do so honestly. Also, the physician must not suffer unlimited financial loss from treating the patient. We assume that the physician's profit is bounded below. We study the incentive-compatible mechanism which maximizes the consumer's expected benefit less the payment to the physician. We also use the most general mechanism, without any restriction on payment schemes being linear.

Our first result asserts that the extraction of information about the consumer's intrinsic benefit information is impossible. The optimal schedule only depends on the physician agency weight on consumer benefit, not the consumer's true benefit. The insurer will have to infer the distribution of the consumer's benefit from the information of the physician agency. The design of payments and quantities is to provide incentives for the physician agency to reveal the weight truthfully.

The second key result is that the program for the optimal quantity-payment schedule actually translates to a choice of a pooling region, in which the quantity is insensitive to the physician agency parameter. This is an unusual step, and is seldom found among solutions for optimal mechanisms.[2] The physician's profit level turns out to be decreasing in the agency weight, while

2. We assume no countervailing incentives, and adopt the usual hazard rate conditions for monotonicity.

the quantity must be increasing. Profits, however, can only be positively related to quantity. The tension caused by incentive compatibility between quantities and nonnegative profits leads to the choice of pooling.

The optimal mechanism must have pooling, and pooling can even be complete. So even information about the physician agency weights on patient benefits will never be completely extracted, and may not be extracted at all. The latter situation is likely when the physician agency weight on patient benefit is very large compared to the intrinsic patient benefit. In other words, when the discrepancy between the physician agency's weight on patient benefit and the intrinsic valuation is sufficiently large, the insurer will *not* attempt to extract the intrinsic information through agency.

Our main result on the optimal mechanism can be interpreted as a form of quantity restriction. Managed care aims to control agency by limiting physicians' discretion over health care quantities, and we derive this result from our model. In earlier work, this is usually taken as an assumption (see for example BAUMGARDNER [1991]. Other attempts to consider managed care propose allocation rules, which are often left as exogenous (FRANK, GLAZER and MCGUIRE [2000]; KEELER, CARTER and NEWHOUSE [1998]).

We compare the optimal quantities with the first best. The expected quantities are the same across the two regimes, but on average there is a reduction in the range of quantities under asymmetric information. Compared to the first best, on average the optimal mechanism assigns more quantities to patients with low valuations, and the opposite is true for those with high valuations.

We next compare the optimal mechanism with the second best, where the physician agency weight is assumed to be known. Even when the physician agency preferences are known, incentives to misreport a patient's intrinsic valuation information persist. These incentives must be removed. In contrast to the third best (where information of both patient valuation and physician agency is unavailable to the insurer), the second-best mechanism can tie quantities to intrinsic patient information. Surprisingly, there is a strong symmetry between the second best and the third best where the agency preference information is unknown to the insurer. In the second best, there is always pooling, and it may be complete. There is also compression of quantities in the same fashion.

ARROW [1963] pointed out the market failure due to the missing information about health status. The subsequent literature has highlighted other sources of market failures such as risk selection (GLAZER and MCGUIRE [2000]), cost and quality effort (MA [1994]), and creaming and dumping (ELLIS [1998]), etc. The literature has concentrated on problems of "hidden information" and "hidden action" of the provider. Our model follows the same line of investigation: the consumer's health status is unknown. Our model of the asymmetric information arising from the physician agency is novel.

While we simply hypothesize that physician agency is represented by preferences weighing physician profit and patient benefits, MA and MCGUIRE [1997] explicitly model collusion between a patient and a provider, but the physician is only interested in maximizing profit. DRANOVE [1988] examines bilateral asymmetric information between the physician and the patient. While this interaction is studied explicitly in Dranove's model, the design of optimal payment is not considered.

Recently, there has been some interest in incentive theories when agents are partly motivated by monetary rewards and partly by work activities; see BESLEY and GHATAK [2005], and DIXIT [2005], and the references there. These papers argue that public and private firms may adopt goals other than pure profit maximization to attract workers who are more motivated by work activities. Our model can be interpreted as one where physicians are altruistic towards their patients, so physicians preferences include both monetary rewards and their patients' benefits, which correspond to their activities. Our model adopts a mechanism design approach. The intensity of utility due to work activities is private information here, and the appropriate incentive constraints are considered.

JACK [2005] considers incentives for cost and quality choices by health care providers with unknown altruism. Jack assumes that the provider derives utility from supplying qualities, but this utility is unknown to the regulator. The provider's choices of quality and cost effort are also unobservable, so the model contains elements of hidden information and hidden action. Jack, however, assumes away any patient heterogeneity, but this is a critical element in our paper. Furthermore, JACK [2005] adopts a reservation utility constraint, and derives the optimal menu of cost-sharing schemes. Our model does not consider hidden action. A key modeling element here is that we impose a minimum profit constraint for the physician, rather than a reservation utility constraint. While the optimal mechanism in Jack yields full separation, ours must exhibit some pooling.[3] In Appendix B, we solve our model using reservation utility instead of minimum profit constraints, to contrast our results with Jack's.

SECTION II presents the model. The following section contains the characterization of incentive compatible payment and quantity schedules. SECTION IV derives the main results, compares the optimal mechanism with the first best, and provides some examples. We also derive the optimal mechanism when the physician agency parameter is public information while the intrinsic patient information remains unknown. The results and comparisons are in SECTION V. The last section contains some concluding remarks. Proofs of results are given in Appendix A, while results of a model using reservation utility are presented in Appendix B.

II. The Model

We now describe a general model of physician agency and quantity-payment design. An insurance or managed care company writes an insurance contract with a consumer and a payment contract with a physician. This company may also be a public agency. If the consumer becomes sick, she seeks medical treatment from the doctor. For simplicity, and as in most managed care plans, we assume that the insurance coverage is complete and thus the patient does not bear any monetary expense when she seeks medical care.

Due to illness conditions, a consumer's severities vary, and so do her benefits from treatment. Upon diagnosis the doctor learns the consumer's conditions or her benefits from treatment. This information becomes the physician's private information, and is unknown to the insurer.

After interacting with the patient, the doctor prescribes a treatment quantity for her. The real-valued variable $q \geq 0$ denotes the health care quantity. We use a real-valued parameter

3. In our paper, quantity distortions are caused by the patients' unobserved heterogeneity. Without such heterogeneity, the insurer can impose the first-best quantity while satisfying the provider's minimum profit constraint, even under unknown altruism. This contrasts with JACK [2005]'s hidden action model.

$\alpha > 0$ to characterize patient severity or potential benefit; this parameter varies according to a distribution. For a consumer with parameter α, her benefit from quantity q is $\alpha V(q)$, where V is a strictly increasing and strictly concave function. The function V is assumed to be common knowledge while the value of α is the physician's private information.[4]

The physician bears a cost $C(q)$ when he prescribes a quantity q. The function C is strictly increasing, and strictly convex. The cost function is common knowledge, and we assume that the cost of treatment is verifiable information. This also means that the treatment quantity is verifiable. If the physician is paid an amount R after he provides quantity q to the patient, his profit is $R - C(q)$.

Being unable to order treatment quantity or bargain with the insurer directly, the patient must interact with the physician to obtain health care. This interaction is the physician agency that we now describe. After their interaction, the joint action of the physician and the patient is based on the following utility: $R - C(q) + \beta V(q)$, where $\beta > 0$ captures the physician agency weight on the patient's benefit. That is, the physician, representing the patient, aims to maximize $R - C(q) + \beta V(q)$.

The physician agency utility may be regarded as a form of altruism. The patient delegates her treatment decision to the altruistic physician, who derives utility from treatment quantities. The degree of altruism is the physician's private information. The physician agency may involve complex agreements and understanding unknown to the insurer. This complexity is modeled by the physician's private information on both the patient's and the physician's valuation of benefits.

The important assumption is that physician agency is complex: the parameter β is unknown to the insurance company, and is the physician's private information. Physician agency is imperfect. The parameter β need not be the same as α. We can also interpret β as a physician's practice type. We adopt a general assumption that (α, β) follows a joint distribution function. Again, both α and β are the physician's private information.

We assume that the joint distribution of (α, β) is common knowledge. Nevertheless, we will mostly work with the marginal distribution of β, $G(\beta)$. We let the marginal distribution of β have a strictly positive and continuous density function g on the support $[\underline{\beta}, \overline{\beta}]$. An assumption on the expectation of α conditional on β will be made later.

The managed care company designs an optimal mechanism respecting the physician's private information about α and β. We use a general mechanism. Since costs and quantities are verifiable, they can be explicitly specified by the mechanism. According to the Revelation Principle, the optimal mechanism must be one in which the physician reports α and β truthfully.

A mechanism is defined by the following pair of functions: $(q(\alpha, \beta), R(\alpha, \beta))$, where $q(\alpha, \beta)$ is the quantity the physician provides and $R(\alpha, \beta)$ his payment when he reports α and β.[5] We call $R - C(q)$ the physician's profit.

4. Alternatively, we can regard $\alpha V(q)$ as the valuation of a payer or regulator which provides health care to some insured population.

5. We will let the mechanism be deterministic. We believe that stochastic mechanisms (which let quantities and payments be determined by lotteries) are suboptimal. The convexity of C implies that the total expected cost is higher than the cost of the expected quantity if q is determined via a lottery. Likewise, the concavity of V implies that the expected utility is lower than the utility of the expected quantity. A stochastic mechanism either raises costs or reduces benefits, but has no impact on incentives since the physician's preferences exhibit no risk aversion.

An altruistic physician may be willing to forgo some financial gains, but should not be expected to incur very large losses. Accordingly, we impose a minimum profit condition: $R(\alpha,\beta) - C(q(\alpha,\beta)) \geq L$ for each (α,β), where L is a finite and possibly negative number. We will normalize L to 0, but, again, it can be negative. Alternatively, one may consider an *ex ante* minimum profit constraint where the physician must not incur very large *expected* losses. A physician then will have to commit to treat a patient even when *ex post* the financial loss turns out to be more than L. Indeed, BARDEY and LESUR [2006] use an *ex ante* participation constraint, but in their model all physicians have the same degree of altruism.[6]

A mechanism is said to be *incentive compatible* if

$$R(\alpha,\beta) - C(q(\alpha,\beta)) + \beta V(q(\alpha,\beta)) \geq R(\alpha',\beta') - C(q(\alpha',\beta')) + \beta V(q(\alpha',\beta')), \quad (1)$$

for all $\alpha, \alpha', \beta, \beta'$. Obviously, an incentive compatible mechanism induces truth telling. A mechanism is said to satisfy *minimum profit* if

$$R(\alpha,\beta) - C(q(\alpha,\beta)) \geq 0, \quad (2)$$

for all α, β. We say that a mechanism is *admissible* if it is incentive compatible and satisfies minimum profit. The mechanism presumes that when the physician chooses a treatment quantity, he is guided by an altruistic preference ordering, but that he is willing to participate when his treatment decisions yield profits of at least L. One can also view the physician-agency as being represented by a lexicographical preference ordering: as long as profits are at least L, the physician-agency preferences apply. From now on, we set L to 0.

Before we proceed to derive the optimal mechanism, we write down a perfect information benchmark. To allow for the possibility that the insurer is the government, we use a general welfare index. Let $S_c \equiv \alpha V(q) - R(q)$ be consumer surplus net of the payment to the physician. Let welfare W_λ be a weighted sum of net consumer surplus and agency utility:

$$\begin{aligned} W_\lambda &= \lambda S_c + (1-\lambda)[R - C(q) + \beta V(q)] \\ &= [\lambda\alpha + (1-\lambda)\beta]V(q) - (1-\lambda)C(q) - (2\lambda - 1)R, \end{aligned}$$

where λ is the weight on consumer surplus. The patient's benefit enters the welfare index in two ways: first through the patient's utility, and second through the physician agency utility.

Transferring money from consumers to pay physicians is costly if and only if $\lambda > 1/2$; otherwise, there would not be any concern for efficiency. Therefore, we let λ be between $1/2$ and 1. The first-best quantity together with transfer maximize W_λ under the constraint $R - C(q) \geq 0$. When $\lambda \geq 1/2$, the insurer minimizes payment to the physician, so that the minimum profit constraint binds, $R - C(q) = 0$, and the first-best quantity $q^\star$ is given by

$$\frac{C'(q^\star)}{V'(q^\star)} = \alpha + \frac{1-\lambda}{\lambda}\beta.$$

6. The results for our model would have been similar if we had used an *ex ante* minimum income constraint. The model would have to be expanded to include an additional dimension besides the physician agency and patient severity dimensions. These results are available from the authors.

In the limit case $\lambda = 1/2$ (consumer surplus and agency utility being given equal weights), the first best is given by $C'(q^\star)/V'(q^\star) = \alpha + \beta$. On the other hand, when $\lambda = 1$ (welfare consisting solely of consumer surplus), the first best satisfies $C'(q^\star)/V'(q^\star) = \alpha$.

For simpler notation, we let $\lambda = 1$, and concentrate hereafter on the maximization of net consumer surplus. This is consistent with competition in the insurance or managed care markets: a firm that fails to pick a mechanism to maximize consumer surplus will be driven out of the market. However, the qualitative results hold as long as the physician agency rent is costly ($\lambda > 1/2$), as we show at the end of SECTION IV.1.

III. Characterization of Admissible Mechanisms

The direct revelation mechanism consists of the quantity and payment functions that depend on both α and β, and it must be an equilibrium for the physician to report α and β truthfully. Nevertheless, the physician's preferences only depend on β, and it appears that the physician's private information about α cannot be extracted directly. This turns out to be valid, but some arguments are necessary.

For a given mechanism (q,R), we define a physician's maximum or indirect utility by

$$U(\beta) \equiv \max_{\alpha',\beta'} \{R(\alpha',\beta') - C(q(\alpha',\beta')) + \beta V(q(\alpha',\beta'))\}, \tag{3}$$

which is independent of α. Clearly, there cannot be any strict incentive for the physician to report any particular value of α.

Nevertheless, suppose that, for some β, the physician is indifferent between *all* quantity-payment pairs as α changes. That is,

$$R(\alpha,\beta) - C(q(\alpha,\beta)) + \beta V(q(\alpha,\beta)) = R(\alpha',\beta) - C(q(\alpha',\beta)) + \beta V(q(\alpha',\beta)),$$

for all α, α', and all these are equal to $U(\beta)$. Then it is an optimal response for him to report α truthfully. However, making quantities and payments contingent on α appears to rely on a delicate balance. This sort of knife-edge construction can become problematic when the incentives for truthful revelation of β are to be considered simultaneously. In other words, although for a given β it may be possible to construct R and q as functions of α to satisfy the indifference requirement, this must interfere with the incentive constraint for nearby values of β.

The proof of the following, as well as other results in the paper, can be found in Appendix A:

Lemma 1. An incentive compatible mechanism $(q(\alpha,\beta), R(\alpha,\beta))$ must have both quantity q and payment R independent of α for almost every β.

The basic idea for Lemma 1 is as follows. As we have already shown, the indirect utility function U is independent of α. Now because U is the maximum of affine linear functions of β, it must be convex in β, and hence differentiable almost everywhere. Incentive compatibility then implies that the function V must be constant in α; otherwise, the differentiability of U over a dense set of β will be violated. So for almost every value of β, q must not be a function of α; likewise for R.

Lemma 1 is a significant result. Extracting information on α directly is impossible. When the optimal mechanism is only based on β, it is potentially possible that two patients with identical characteristics will receive very different health care, depending on the way each interacts with her physician. The optimal mechanism can only base health care quantities on the physician agency parameter β, and it must consider the conditional distribution of α given β, as we will see later.

We assume that the distribution of β admits a density. It follows that the set of β where the schedule could depend on α (which has zero Lebesgue measure according to Lemma 1) has no impact on the objective function. We therefore have

Corollary 1. Without loss of generality, an incentive compatible mechanism can be written as $(q(\beta), R(\beta))$.

The following Lemma characterizes incentive compatible mechanisms in terms of the quantity q and the indirect utility U.

Lemma 2. A mechanism $(R(\beta), q(\beta))$ is incentive compatible if and only if the indirect utility $U(\beta) = R(\beta) - C(q(\beta)) + \beta V(q(\beta))$ is a convex function of β and satisfies

$$U'(\beta) = V(q(\beta)),$$

for all $\beta \in [\underline{\beta}, \overline{\beta}]$.

This result is a straightforward consequence of the Envelope Theorem. The convexity of the indirect utility is equivalent to the monotonicity of the quantity: $U''(\beta) = V'(q(\beta))q'(\beta) \geq 0$, so q must be nondecreasing. The function U, however, may not be monotone, its derivative depending on the sign of V. The benefit function V is an ordinal measure, its sign being irrelevant. We only insist on V being increasing and concave. The indirect utility must be continuous, since it is convex. However for any incentive compatible mechanism, the quantity q is not necessarily continuous. An upward jump in q corresponds to a kink in U.

Now we make use of the indirect utility function to simplify the set of minimum profit constraints. We do this by establishing a monotonicity result. The profit for a physician with parameter β is $\pi(\beta) = R(\beta) - C(q(\beta))$, and we use the definition of U to rewrite it as:

$$\pi(\beta) = U(\beta) - \beta V(q(\beta)) = U(\beta) - \beta U'(\beta). \tag{4}$$

There is a geometric representation of profit. Consider the graph of $U(\beta)$ on the (β, U) plane. The value of profit at any β is the intersection of the tangent of U at β with the U-axis.

Differentiating the right-hand side of (4), we show that the physician's profit is nonincreasing in β:

$$\pi'(\beta) = \frac{\mathrm{d}}{\mathrm{d}\beta}[U(\beta) - \beta U'(\beta)] = -\beta U''(\beta) = -\beta V'(q(\beta))q'(\beta) \leq 0. \tag{5}$$

The inequality follows from the convexity of U. In other words, incentive compatibility implies that the physician's profit must be nonincreasing in β: the more the physician cares about patient benefit, the lower his profit level. The monotonicity of physician profit gives the following:

Lemma 3. For any incentive compatible mechanism, the minimum profit constraint is satisfied for all β if and only if it is satisfied for $\beta = \overline{\beta}$. Moreover, if the minimum profit constraint binds at $\tilde{\beta}$, then the minimum profit constraints for any $\beta > \tilde{\beta}$ must also bind. In other words, binding minimum profit constraints can only occur on an interval $[\hat{\beta}, \overline{\beta}]$. Finally, whenever minimum profit constraints bind, quantities must become constant with respect to β, resulting in pooling: $q(\beta) = \hat{q}$.

Lemma 3 is the key to the analysis. As functions of β, quantities must be nondecreasing while profits nonincreasing in an incentive compatible mechanism. From (5), $\pi'(\beta) = 0$ if and only if $q'(\beta) = 0$. Setting quantities constant for a range of β implies that the corresponding profits are zero. Pooling quantities for a range of β means binding minimum profit constraints, which save on information rent. The optimal mechanism must consider this pooling interval. In the next section, we prove that pooling must exist in an optimal mechanism. In fact, complete pooling over the entire range of β may be optimal.

IV. The Optimal Payment and Quantity

We let the objective function of the insurer be the expected consumer benefit less the expected payment to the physician. This is consistent with competition in the insurance or managed care markets. A firm that fails to pick a mechanism to maximize consumer benefit will be driven out of the market. At the end of SECTION IV.1, we present results with a weighted average of consumer surplus and physician agency utility as the objective function.

For a given mechanism (q,R), the insurer's objective function is

$$W = \iint [\alpha V(q(\beta)) - R(\beta)]\, h(\alpha,\beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta,$$

where h is the joint density of α and β. Integrating out α, we write W as

$$W = \int_{\underline{\beta}}^{\overline{\beta}} [\alpha_{\mathrm{m}}(\beta) V(q(\beta)) - R(\beta)]\, g(\beta)\, \mathrm{d}\beta, \tag{6}$$

where $\alpha_{\mathrm{m}}(\beta) \equiv E(\alpha|\beta) = \int \alpha h(\alpha,\beta)/g(\beta)\mathrm{d}\alpha$ is the conditional mean of α given β. In other words, $\alpha_{\mathrm{m}}(\beta)$ is the insurer's assessment of a consumer's average valuation given the altruism or physician agency parameter β.

IV.1. *Deriving the Optimal Mechanism*

An optimal mechanism is an admissible mechanism (q,R) that maximizes (6). We use the definition of U, Lemma 3, and integration by parts to eliminate R, and then find the optimal quantity schedule. From Lemma 3, we can write the integral (6) as the sum of two integrals, one over $[\underline{\beta}, \hat{\beta}]$ and the other $[\hat{\beta}, \overline{\beta}]$, where $\hat{\beta}$ is the lower limit of the pooling interval. Replacing the payment $R(\beta)$ by $U(\beta) + C(q(\beta)) - \beta V(q(\beta))$, we get

$$\int_{\underline{\beta}}^{\hat{\beta}} [\alpha_{\mathrm{m}}(\beta) V(q(\beta)) - R(\beta)]\, g(\beta)\, \mathrm{d}\beta = \int_{\underline{\beta}}^{\hat{\beta}} [(\alpha_{\mathrm{m}}(\beta) + \beta) V(q(\beta)) - C(q(\beta)) - U(\beta)] g(\beta)\, \mathrm{d}\beta.$$

Now we integrate the utility term by parts. Using $U' = V(q)$ (from Lemma 2) and $U(\hat{\beta}) = \pi(\hat{\beta}) + \hat{\beta}V(q(\hat{\beta})) = \hat{\beta}V(q(\hat{\beta}))$, we get

$$\int_{\underline{\beta}}^{\hat{\beta}} [\alpha_{\mathrm{m}}(\beta)V(q(\beta)) - R(\beta)]\, g(\beta)\, \mathrm{d}\beta =$$
$$\int_{\underline{\beta}}^{\hat{\beta}} \left\{ \left[\alpha_{\mathrm{m}}(\beta) + \beta + \frac{G(\beta)}{g(\beta)}\right] V(q(\beta)) - C(q(\beta)) \right\} g(\beta)\, \mathrm{d}\beta$$
$$- G(\hat{\beta})\hat{\beta}V(q(\hat{\beta})).$$

For the integral on the pooling interval $[\hat{\beta}, \overline{\beta}]$, we let the quantity be a constant $\hat{q}$. Because profit is zero over this interval, the optimal payment is $C(\hat{q})$. The objective function is

$$W = \int_{\underline{\beta}}^{\hat{\beta}} \left\{ \left[\alpha_{\mathrm{m}}(\beta) + \beta + \frac{G(\beta)}{g(\beta)}\right] V(q(\beta)) - C(q(\beta)) \right\} g(\beta)\mathrm{d}\beta \tag{7}$$

$$- G(\hat{\beta})\hat{\beta}V(\hat{q}) + \int_{\hat{\beta}}^{\overline{\beta}} [\alpha_{\mathrm{m}}(\beta)V(\hat{q}) - C(\hat{q})]g(\beta)\mathrm{d}\beta. \tag{8}$$

The problem of finding the optimal mechanism can now be reformulated as follows:[7] choose a nondecreasing function $q(\beta)$, $\underline{\beta} \leq \beta \leq \hat{\beta}$, and numbers $\hat{q}$ and $\hat{\beta}$, with $\underline{\beta} \leq \hat{\beta} \leq \overline{\beta}$, to maximize (8).

We now introduce our main assumption, which we adopt for the rest of the paper:

Assumption A $\quad \alpha_{\mathrm{m}}(\beta) + \beta + \dfrac{G(\beta)}{g(\beta)}$ is continuous and nondecreasing in β.

Assumption A is true under the following two assumptions:

Assumption A1 $\quad (G(\beta))/g(\beta)$ is continuous and nondecreasing in β.

Assumption A2 $\quad \alpha_{\mathrm{m}}(\beta)$ is continuous and nondecreasing in β.

Assumption A1 is the familiar monotone hazard rate condition, satisfied for many classes of distributions (uniform, normal, exponential, etc.) Assumption A2 says that the conditional expectation of α is increasing in the physician's altruism parameter β. This seems a natural assumption to make. If the physician's concern for the patient takes into account the patient's valuation, a higher patient average valuation is associated with a physician who exhibits a higher degree of altruism. In any case, Assumption A is weaker than Assumptions A1 and A2.

7. Another form of the problem is also tractable. The objective function can be expressed in terms of the indirect utility:

$$W = \int_{\underline{\beta}}^{\overline{\beta}} [(\alpha_{\mathrm{m}}(\beta) + \beta)U' - C(V^{-1}(U')) - U]g(\beta)\mathrm{d}\beta,$$

which is linear in U and strictly concave in U'. The incentive (U convex) and the minimum profit ($U - \beta U' \geq 0$) constraints define a convex set. The existence and the uniqueness of the optimum follow from standard arguments. In this formulation, calculus of variation can be used to characterize the optimal mechanism. We present here a more intuitive method. The solution via variation is available from the authors.

The form of the objective function in (8) actually reveals the various aspects of the problem. The integral from β to $\hat{\beta}$ refers to the regime of positive profits for the physician. As we have noted at the end of SECTION II, when the minimum profit constraint does not bind, the total social benefit $[\alpha+\beta]V(q)$ becomes relevant. Because the information of α cannot be extracted directly, it is replaced by the conditional expectation $\alpha_{\mathrm{m}}(\beta)$. The hazard rate G/g is the familiar adjustment for the rent due to asymmetric information: the "virtual" social benefit is $\alpha_{\mathrm{m}}(\beta)+\beta+G(\beta)/g(\beta)$. The term $\hat{\beta}V(q(\hat{\beta}))$ is a measure of the indirect utility at the beginning of the pooling region. For any $\beta > \hat{\beta}$, the quantity becomes fixed, the physician's profit zero, and the indirect utility βV. So the pooling quantity $\hat{q}$ determines the indirect utility level $\hat{\beta}V(\hat{q})$, which is the base indirect utility level for all those physicians with β smaller than $\hat{\beta}$, hence the factor $G(\hat{\beta})$. Finally, the choices of $\hat{q}$ and $\hat{\beta}$ completely determine the pooling regime, which is the last term in (8).

The optimization program is separable with respect to $\hat{q}$ and $q(\beta)$, β in $[\underline{\beta},\hat{\beta}]$. We apply pointwise optimization to obtain the first-order condition for $q(\beta)$, β in $[\underline{\beta},\hat{\beta}]$:

$$\left[\alpha_{\mathrm{m}}(\beta)+\beta+\frac{G(\beta)}{g(\beta)}\right]V'(q(\beta)) = C'(q(\beta)). \tag{9}$$

Under Assumption A, the quantity defined by (9) is nondecreasing (recall that C'/V' is increasing).

Next we show that the optimal quantity is continuous at $\hat{q}$. The intuition is this: if the optimal quantity jumps upward at $\hat{q}$, then the pooling interval can be reduced. The value of $\hat{\beta}$ can be increased while the (higher) quantity at $\hat{q}$ can be kept constant. In other words, if there is a jump, in quantity at $\hat{q}$, the minimum profit continues to hold while the pooling interval can be made smaller. Less pooling means that more information about α is revealed.

Lemma 4. Suppose that $\underline{\beta} < \hat{\beta} < \overline{\beta}$. The optimal quantity is continuous at $\hat{\beta}$. The quantity $\hat{q}$ in the pooling region $[\hat{\beta},\overline{\beta}]$ must satisfy

$$\left[\alpha_{\mathrm{m}}(\hat{\beta})+\hat{\beta}+\frac{G(\hat{\beta})}{g(\hat{\beta})}\right]V'(\hat{q}) = C'(\hat{q}). \tag{10}$$

We continue with the characterization of the optimal $\hat{q}$. Using the first-order condition with respect to $\hat{q}$, we have the following

Lemma 5. The pooling interval satisfies the condition

$$\int_{\hat{\beta}}^{\overline{\beta}} \left\{\alpha_{\mathrm{m}}(\beta)V'(\hat{q}) - C'(\hat{q})\right\} g(\beta)\mathrm{d}\beta = \hat{\beta}G(\hat{\beta})V'(\hat{q}). \tag{11}$$

The left-hand side of (5) measures the usual (expected) marginal benefit and cost of the quantity $\hat{q}$. The term on the right-hand side measures the change in the base indirect utility level. Raising $\hat{q}$ has a negative effect on the objective function since it gives more profit to all physicians with β less than $\hat{\beta}$.

Equation (11) implies that there is no solution with an empty pooling interval; that is, $\hat{\beta}=\overline{\beta}$ cannot satisfy (11). If $\hat{\beta}$ was set at $\overline{\beta}$, then reducing $\hat{\beta}$ must improve the objective function. This would only lead to a second-order loss in the efficiency of q since (9) was satisfied at $\overline{\beta}$, but this would result in a first-order gain since profits for all physicians would be reduced. So Lemma 5 implies that in the optimal mechanism there must be some pooling in β (although an incentive compatible mechanism that is fully separating in β is feasible).

Finally, we characterize the pooling region $\hat{\beta}$. Equations (10) and (11) together determine $\hat{\beta}$ and $\hat{q}$ if in fact they yield an interior solution $\hat{\beta} > \underline{\beta}$. It is, however, possible that these two equations yield a solution of $\hat{\beta}$ below $\underline{\beta}$, in which case, the quantity will be constant, and equation (11) with $\hat{\beta}$ set at $\underline{\beta}$ will determine $\hat{q}$. With α_μ denoting the unconditional mean of α, we present the condition for an interior solution:

Lemma 6. The pooling interval is in the interior of the support of β, $\hat{\beta} > \underline{\beta}$, if and only if

$$\alpha_\mu > \alpha_{\mathrm{m}}(\underline{\beta}) + \underline{\beta}. \tag{12}$$

When can separation be optimal? We know that the best prospect is for β near $\underline{\beta}$. So now suppose that there is complete pooling, say, the quantity is fixed at $\hat{q}$ for all β. A complete-pooling quantity must be based on the unconditional mean α_μ. What can be gained by some separation at $\underline{\beta}$? Recall that in a separating region, the social benefit $[\alpha_{\mathrm{m}}(\beta)+\beta]V(q)$ is relevant due to strictly positive profits. When $\alpha_\mu > \alpha_{\mathrm{m}}(\underline{\beta}) + \underline{\beta}$, the complete pooling quantity $\hat{q}$ is too high for $\underline{\beta}$. Lowering the quantity from $\hat{q}$ for $\underline{\beta}$ and then subsequently increasing it for higher β (due to Assumption A) will reduce the inefficiency due to the excessive quantity $\hat{q}$. Although this will entail some profits for the physician, it is worthwhile since inequality (12) is strict.

The condition for some separation (12) will fail to hold if the support of β is much larger than that of α, or when the variation of $\alpha_{\mathrm{m}}(\beta)$ is small. When (12) is violated, extracting information of α via β must lead to high profits to the physician due to quantities always being higher than under complete pooling. In this case, the optimal quantity is constant on the whole interval $[\underline{\beta}, \overline{\beta}]$ and given by

$$\alpha_\mu V'(\hat{q}) = C'(\hat{q}). \tag{13}$$

We summarize our results by the following:

Proposition 1. Under Assumption A, the optimal mechanism is defined as follows:

1. If $\alpha_\mu \leq \alpha_{\mathrm{m}}(\underline{\beta}) + \underline{\beta}$, the optimal quantity for each value of β is given by equation (13). The physician earns zero profit.

2. If $\alpha_\mu > \alpha_{\mathrm{m}}(\underline{\beta}) + \underline{\beta}$, there exists $\hat{\beta}$, with $\underline{\beta} < \hat{\beta} < \overline{\beta}$, and the following are properties for the optimal quantities:

   (a) For $\underline{\beta} \leq \beta \leq \hat{\beta}$, the optimal quantity $q(\beta)$ is strictly increasing and satisfies (9).

   (b) For $\hat{\beta} \leq \beta \leq \overline{\beta}$, the optimal quantity is constant and equal to $\hat{q}$, where $\hat{q}$ and $\hat{\beta}$ satisfy equations (10) and (11).

   (c) The physician earns strictly positive profit if and only if his value of β is less than $\hat{\beta}$.

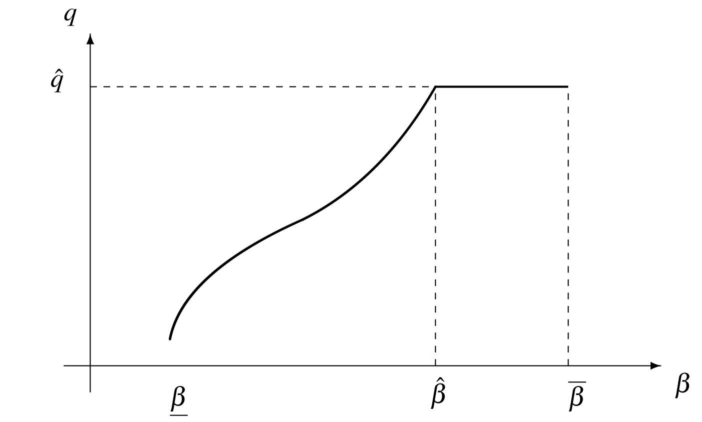

FIGURE 1. – OPTIMAL QUANTITY

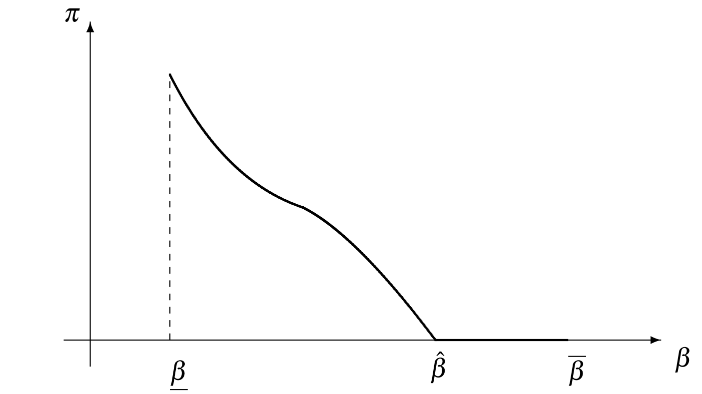

FIGURE 2. – PHYSICIAN PROFIT IN OPTIMAL MECHANISM

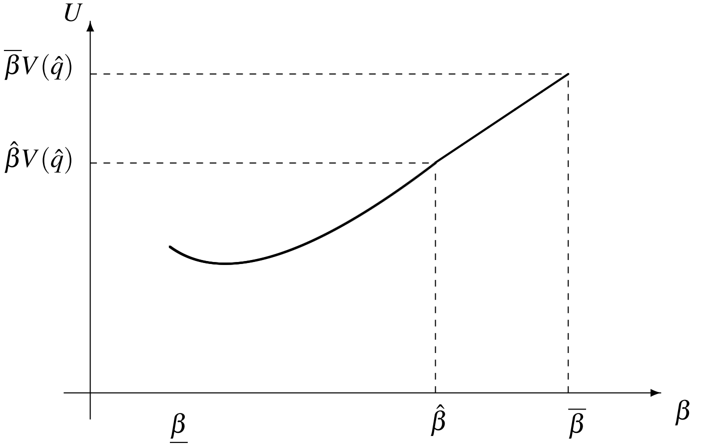

FIGURE 3. – INDIRECT UTILITY IN OPTIMAL MECHANISM

In FIGURES 1, 2 and 3, we display the typical shapes of the optimal quantity, profit, and indirect utility. The indirect utility is convex. It is strictly convex up to $\hat{\beta}$, and then becomes linear. Indeed, for $\beta \geq \hat{\beta}$, $U(\beta) = \beta V(\hat{q})$. Accordingly, the physician's profit, $\pi = U - \beta U'$, is strictly decreasing until $\hat{\beta}$, and then becomes zero.[8]

Properties of the pooling result hold when the welfare is a combination $W_\lambda = \lambda S_c + (1-\lambda)U$, with $\lambda > 1/2$. In that case, the optimal quantity in the non-pooling region is given by

$$\lambda \frac{C'(q)}{V'(q)} = \lambda(\alpha+\beta) + (2\lambda-1)\frac{G(\beta)}{g(\beta)}. \tag{14}$$

The pooling threshold and quantity $\hat{\beta}$ and $\hat{q} = q(\hat{\beta})$ are characterized by equation (14) evaluated at $\beta = \hat{\beta}$, together with

$$\int_{\hat{\beta}}^{\overline{\beta}} \{[\lambda\alpha + (1-\lambda)\hat{\beta}]V'(\hat{q}) - \lambda C'(\hat{q})\} g(\beta)\mathrm{d}\beta = (2\lambda-1)\hat{\beta}G(\hat{\beta})V'(\hat{q}). \tag{15}$$

The pooling interval $[\hat{\beta}, \overline{\beta}]$ degenerates towards the singleton $\{\overline{\beta}\}$ as λ goes to $1/2$. The pooling result follows from the tradeoff between rent extraction and efficiency, together with the minimum profit constraint.[9]

Pooling in the optimal schedule can be interpreted as quantity limits in managed care. In the pooling interval, the quantity is insensitive to β (and α). Moreover, the limit applies to higher values of β, which correspond to higher expected severities or benefits. What is more, the quantity restriction may be extensive, in which case the managed care plan offers the same quantity as that which is based on the average severity of the entire population. Where the optimal quantity is increasing with β, it is based on the expectation of α conditional on β. We next investigate how the optimal quantity is related to the first best.

IV.2. *Comparing the Optimal Mechanism with the First Best*

The comparison between the first best and optimal quantity in Proposition 1 is not quite straightforward, because the first best only depends on α while the optimal quantity depends on β. For a comparison, we calculate the expected first-best quantities conditional on β. For each value of β, we consider the conditional distribution of α, and the corresponding first-best quantities in this distribution. For ease of exposition, we compare $C'(q)/V'(q)$ (which is increasing by assumption) rather than the quantity q itself across the asymmetric information and first-best regimes.

Since $C'(q^\star)/V'(q^\star) = \alpha$ at the first best, we have

$$\mathrm{E}\left\{ \frac{C'(q^\star(\alpha))}{V'(q^\star(\alpha))} \middle| \beta \right\} \equiv \alpha_{\mathrm{m}}(\beta). \tag{16}$$

In the optimal mechanism, $C'(q)/V'(q)$ is a function of β given by Proposition 1. The comparison between the first best and optimal quantity functions is given by the following proposition.

8. Proposition 1 includes the first best as a special case. If there is no uncertainty concerning α, $\alpha_\mu = \alpha_{\mathrm{m}}(\beta)$, all β. The first part of Proposition 1 applies, and equation (13) becomes exactly the one that defines the first-best quantity $q^*(\alpha)$ at $\lambda = 1$.

9. The derivations of these results are available from the authors.

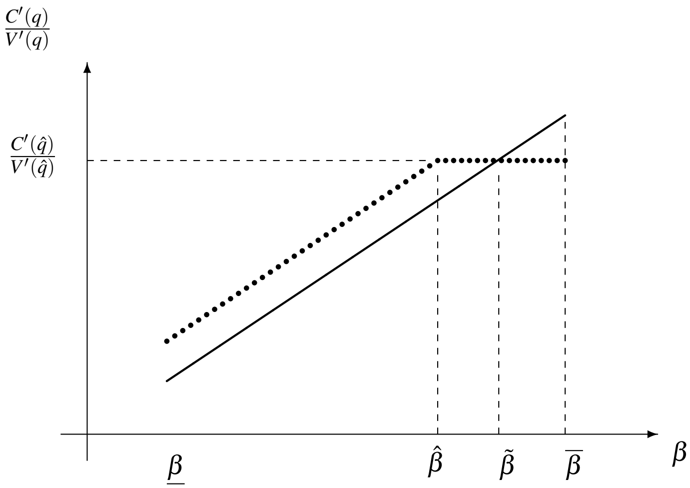

FIGURE 4. – COMPARISON BETWEEN FIRST BEST (SOLID LINE) AND THIRD BEST (DOTTED LINE)

Proposition 2. At the optimum, we have

$$\int_{\underline{\beta}}^{\overline{\beta}} \frac{C'(q(\beta))}{V'(q(\beta))} g(\beta) \mathrm{d}\beta = \alpha_\mu. \tag{17}$$

If $\alpha_{\mathrm{m}}(\beta)$ is nondecreasing (Assumption A2), there exists $\tilde{\beta}$ with $\hat{\beta} < \tilde{\beta} < \overline{\beta}$ such that

$$\begin{cases} \dfrac{C'(q(\beta))}{V'(q(\beta))} \geq \alpha_{\mathrm{m}}(\beta) & \text{for } \beta \leq \tilde{\beta} \\ \dfrac{C'(q(\beta))}{V'(q(\beta))} = \dfrac{C'(\hat{q})}{V'(\hat{q})} \leq \alpha_{\mathrm{m}}(\beta) & \text{for } \beta \geq \tilde{\beta}. \end{cases}$$

From Proposition 1, we compute the expected value of C'/V' with respect to β. Equation (17) says that the unconditional expectation of $C'(q)/V'(q)$ is exactly α_μ, the unconditional mean of α or the consumer's average valuation. If we assume that $\alpha_{\mathrm{m}}(\beta)$ is nondecreasing, then on average there is overprovision of quantities for low values of β, and underprovision for high values of β. In the separating region (if this region exists), the optimal quantity is always distorted upwards due to information rent and the social surplus consideration—see equation (9). Because the unconditional expectation of C'/V' must be the same across the two regimes, there must be downward distortion in the pooling regime. Managed care, on average, is associated with a compression of service variations. FIGURE 4 illustrates Proposition 2. The graph shows two plots of $C'(q)/V'(q)$ against β; the solid line is for the first best, and the other for the optimal mechanism.

IV.3. *Some Examples*

A few examples illustrate the scope of our results in Proposition 1. In Example 1, α and β are independent. This can be regarded as a benchmark. By contrast, in each of Examples 2 and 3, there is a deterministic relationship between and α and β. Here, the physician routinely expresses the consumer's preferences in a simple, affine-linear fashion. In Examples 4 and 5, α and β are both unbounded and correlated. In these two examples, the physician agency exhibits a lot of complexity. (Although our results in Proposition 1 have been written for distributions with finite support, they remain applicable to these examples.)

Example 1. *Independent α and β.* Suppose that α and β are independent. Then $\alpha_{\mathrm{m}}(\underline{\beta}) = \alpha_\mu$. The first part of Proposition 1 applies: there is complete pooling. Learning about α from any report of β is impossible.

Example 2. *Multiplicative α and β.* Suppose that $\beta = \alpha\theta$ where $\theta > 0$ is fixed and known. We have: $\underline{\beta} = \theta\underline{\alpha}$ and $\alpha_{\mathrm{m}}(\underline{\beta}) = \underline{\beta}/\theta = \underline{\alpha}$. There is complete pooling if and only if

$$\alpha_\mu - \underline{\alpha} \leq \theta\underline{\alpha} \qquad \text{or} \qquad \theta \geq \alpha_\mu/\underline{\alpha} - 1.$$

That is, complete pooling is optimal when altruism is high compared to a measure of the range of α. In the separating region (if there is one), the optimal quantity is given by $C'/V' = \beta(1+1/\theta) + G/g$.

Example 3. *Additive α and β.* Suppose that $\beta = \alpha + \theta$, where again θ is fixed and known. Then $\underline{\beta} = \underline{\alpha} + \theta$ and $\alpha_{\mathrm{m}}(\underline{\beta}) = \underline{\beta} - \theta = \underline{\alpha}$. There is complete pooling if and only if

$$\alpha_\mu - \underline{\alpha} \leq \underline{\alpha} + \theta \qquad \text{or} \qquad \alpha_\mu - 2\underline{\alpha} \leq \theta.$$

In the separating regime region (if there is one), the optimal quantity is given by $C'/V' = 2\beta - \theta + G/g$.

For Examples 2 and 3, it is easy to check the following comparative statics results: the higher the value of θ, the bigger the pooling range (that is, the pooling threshold $\hat{\beta}$ decreases with θ). The expected consumer surplus $W(\theta)$ decreases with θ. A higher value of the altruism parameter increases the informational asymmetry between the physician and the insurer and exacerbates the conflict between them. The insurer has to put more restraint on the physician's behavior, and this results in a lower expected surplus for the patient.

Example 4. *Lognormal Distributions.* Suppose that $\beta = \alpha\theta$, and that α and θ are log-normally distributed. That is, $\ln\alpha$ and $\ln\theta$ are normal. Let a_μ and t_μ be the expectations of $\ln\alpha$ and $\ln\theta$, respectively, and σ_a^2 and σ_t^2 their variances. Finally, let ρ denote the correlation between $\ln\alpha$ and $\ln\theta$. When $\rho = 0$, the random variables α and θ are independent, but parameters α and β remain correlated.

The distributions of α, θ and β have a common support $[0, +\infty[$. In other words $\underline{\alpha} = \underline{\beta} = 0$ and $\overline{\alpha} = \overline{\beta} = +\infty$. The expectation of α is $\exp(a_\mu + \sigma_a^2/2) \equiv \alpha_\mu$. The random variable $\ln\beta = \ln\alpha + \ln\theta$ is normal, with expectation $b_\mu = a_\mu + t_\mu$ and variance $\sigma_b^2 = \sigma_a^2 + \sigma_t^2 + 2\rho\sigma_a\sigma_t$. The distribution of $\ln\alpha$ conditionally on $\ln\beta$ is normal, with expectation

$$l_\mu(\beta) = \mathrm{E}\,(\ln\alpha|\ln\beta) = a_\mu + \frac{\sigma_a^2 + \rho\sigma_a\sigma_t}{\sigma_b^2}(\ln\beta - b_\mu)$$

and some variance σ_l^2 that is strictly smaller than σ_a^2.

We can compute the conditional mean of α given β:

$$\alpha_{\mathrm{m}}(\beta) = \mathrm{E}\ (\alpha|\beta) = \exp\ [l_{\mu}(\beta) + \sigma_l^2/2].$$

Assumption A2 is satisfied if and only if $\sigma_a^2 + \rho\sigma_a\sigma_t \geq 0$. Therefore, Assumption A2 in this example is equivalent to $\ln\alpha$ and $\ln\beta$ being nonnegatively correlated. Assumption A1 is satisfied since the log-normal distribution has a monotone hazard rate.[10]

Finally, we have $\underline{\beta} + \alpha_{\mathrm{m}}(\underline{\beta}) = \alpha_{\mathrm{m}}(0) = 0 < \alpha_{\mu}$. According to Proposition 1, pooling must not be complete. There exists $\hat{\beta} > 0$ such that the optimal quantity is constant for $\beta \geq \hat{\beta}$ and is given by $C'(q)/V'(q) = \alpha_{\mathrm{m}} + \beta + G/g$ for $\beta \leq \hat{\beta}$.

Example 5. *Independent Exponential Distributions.* Suppose that $\beta = \alpha + \theta$, and that α and θ are independently and exponentially distributed, each with density $\exp(-x)$, on $[0, +\infty)$. Then we have $\underline{\alpha} = \underline{\beta} = 0$ and $\overline{\alpha} = \overline{\beta} = +\infty$. An exponential distribution is a special case of a gamma distribution. More precisely, α and θ each is a gamma distribution with parameters 1 and 1. The sum of two independent gamma distributions with an identical second parameter is again a gamma distribution (DEGROOT [1986, pp. 288-290]). So β follows a gamma distribution with parameters 2 and 1. Accordingly, the density of β is $g(\beta) = \beta\exp(-\beta)$ on $[0, +\infty)$ and the distribution function is $G(\beta) = 1 - (1+\beta)\exp(-\beta)$. The hazard rate is

$$\frac{G(\beta)}{g(\beta)} = \frac{\exp(\beta) - 1 - \beta}{\beta},$$

and increasing in β on $[0, +\infty)$. The distribution of α conditional on β is the uniform distribution on $[0, \beta]$. Therefore, we have $\alpha_{\mathrm{m}}(\beta) = \beta/2$, and Assumption A is satisfied. Because $\alpha_{\mu} = 1$ and $\underline{\beta} = \alpha_{\mathrm{m}}(\underline{\beta}) = 0$, Proposition 1 says that there exists $\hat{\beta} > 0$ such that the optimal quantity is constant for $\beta \geq \hat{\beta}$, and for $\beta \leq \hat{\beta}$ the quantity $q(\beta)$ is given by

$$\frac{C'(q)}{V'(q)} = \frac{3}{2}\,\beta + \frac{\exp(\beta) - 1 - \beta}{\beta}.$$

V. Second Best Physician Agency

In the previous section, information about α and β is only known to the physician. This may be regarded as a third best. If the value of α were known to the managed care company, the physician's concern for the patient's benefit would be irrelevant and the first-best quantity that maximized $\alpha V(q) - C(q)$ could be implemented. In this section, we consider a second best, where the value of β is known to the managed care company, but the information on α remains the physician's private information.

When $\beta > 0$ is known, Corollary 1 does not apply. So we must consider mechanisms in which the physician is asked to report α. Now we assume that the distribution of α admits a

10. Let g and G be the density and distribution functions of the log-normal distribution, and ϕ and Φ the density and distribution functions of the standard normal distribution. Then $G(x) = \Phi(\ln x)$ and $g(x) = \phi(\ln x)/x$; so $G(x)/g(x) = x\Phi(\ln x)/\phi(\ln x)$. Because Φ/ϕ is increasing, G/g is increasing.

strictly positive density f on the interval $[\underline{\alpha},\overline{\alpha}]$, with $\underline{\alpha} \geq 0$. A mechanism –a pair of functions $(q(\alpha),R(\alpha))$– is said to be incentive compatible if for all α and α'

$$R(\alpha)-C(q(\alpha))+\beta V(q(\alpha)) \geq R(\alpha')-C(q(\alpha'))+\beta V(q(\alpha')). \tag{18}$$

The physician's preferences do not depend on α, so (18) can be written as

$$R(\alpha)-C(q(\alpha))+\beta V(q(\alpha)) = U, \tag{19}$$

for all α, and some constant U. Instead of working with $(q(\alpha),R(\alpha))$, we shall use the quantity function $q(.)$ and the level of utility U (a scalar) as instruments. Given a quantity schedule $q(\alpha)$ and a constant U, we can use (19) to recover the payment $R(\alpha)$. Besides the incentive constraints, a mechanism must ensure that the physician makes a nonnegative profit: $R(\alpha)-C(q(\alpha)) \geq 0$, for all α.

Although a mechanism satisfying (19) removes all incentives for the physician to misreport α, there cannot be any strict incentive for the truthful revelation of this information. The physician's preferences do not directly depend on α. Here, we make the usual assumption that a physician truthfully reveals the information of α if there is no incentive to do otherwise. In effect, we select one equilibrium among a large set of equilibria induced by $(q(\alpha),U)$. These other equilibria are supported by other physician reporting strategies, for example, the physician always reporting the highest (or lowest) value of α whenever he is indifferent between reports.

Given the truthful revelation of the information of α, the objective function of the insurer is

$$W = \int_{\underline{\alpha}}^{\overline{\alpha}} \{\alpha V(q)-R(\alpha)\} f(\alpha)\mathrm{d}\alpha = \int_{\underline{\alpha}}^{\overline{\alpha}} \{(\alpha+\beta)V(q(\alpha))-C(q(\alpha))\} f(\alpha)\mathrm{d}\alpha - U.$$

The optimal mechanism maximizes W subject to the minimum profit constraints: $\beta V(q(\alpha)) \leq U$ for all α.

The solution is easy to describe. Obviously, pointwise optimization can be applied where the minimum profit constraint does not bind. This yields a first-order condition: $(\alpha+\beta)V'(q) = C'(q)$. When the physician earns positive profits, the social benefit $(\alpha+\beta)V(q)$ should be considered, and so the first-order condition describes the appropriate marginal benefit and cost calculations. This also yields an optimal quantity schedule $q(\alpha)$ that is increasing in α. So for a given U, the minimum profit constraints will bind for all values of α above a certain threshold, say, $\hat{\alpha}$; once $\alpha > \hat{\alpha}$, the optimal quantity becomes constant.

The optimal choice of U is never too high, so that some of the minimum profit constraints must bind. Again, this can be explained by the envelope argument. If the threshold $\hat{\alpha}$ was originally at the upper support, then lowering U would reduce profits for all physicians, a first-order gain. This would only lead to a second-order loss since the marginal conditions originally were satisfied. In other words, there must be some pooling. If the value of β is very high, however, the minimum profit constraint may become binding even at the lower support $\underline{\alpha}$; that is, $\hat{\alpha} = \underline{\alpha}$. In this case, the optimal quantity becomes constant for all values of α, and given by $\alpha_\mu V'(q) = C'(q)$. Again, there may be complete pooling.

Proposition 3. When the value of β is public information, the optimal mechanism is defined as follows.

1. If $\alpha_\mu - \underline{\alpha} \leq \beta$, the optimal quantity is constant and given by

$$\alpha_\mu V'(q) = C'(q). \tag{20}$$

2. If $\alpha_\mu - \underline{\alpha} > \beta$, then there exists $\hat{\alpha}$, with $\underline{\alpha} < \hat{\alpha} < \overline{\alpha}$, such that the optimal quantities have the following properties:

   (a) For $\underline{\alpha} < \alpha < \hat{\alpha}$, the optimal quantity is strictly increasing and satisfies

$$(\alpha + \beta)V'(q) = C'(q). \tag{21}$$

   (b) For $\hat{\alpha} \leq \alpha \leq \overline{\alpha}$, the optimal quantity is a constant $\bar{q}$, and given by

$$\int_{\hat{\alpha}}^{\overline{\alpha}} \left\{\alpha V'(\bar{q}) - C'(\bar{q})\right\} f(\alpha)\mathrm{d}\alpha = \beta V'(\bar{q})F(\hat{\alpha}). \tag{22}$$

   The optimal quantity is continuous at $\hat{\alpha}$ so that the equation

$$(\hat{\alpha} + \beta)V'(\bar{q}) = C'(\bar{q}) \tag{23}$$

   together with (22) determine $\hat{\alpha}$ and $\bar{q}$.

   (c) The physician earns a strictly positive profit if and only if α is less than $\hat{\alpha}$.

The symmetry in Propositions 1 and 3 is striking, although the incentive constraints in the second best and third best are quite different.[11] The characteristics of the optimal quantity and physician profits in the second best can easily be illustrated by FIGURES 1 and 2—the necessary modification being a change in the label of the horizontal axis from β to α. The quantitative differences between the second best mechanism and that in SECTION IV can be quite large.

The symmetry does not end here. The comparison between the second best and the first best actually parallels that between the third best and the first best. From equations (21) and (22),

$$\int_{\underline{\alpha}}^{\overline{\alpha}} \frac{C'(q(\alpha))}{V'(q(\alpha))} f(\alpha)\mathrm{d}\alpha = \alpha_\mu,$$

which is symmetric to equation (17) in Proposition 2. From this, we can easily compare the second best with the first best. Again, given the value of β, in the second best, there is always overprovision of quantities for lower values of α, and underprovision for high α. FIGURE 4 illustrates this comparison if the label of the horizontal axis is changed to α.

These surprising comparisons indicate that the design of optimal payment and quantity depends critically on the existence of physician agency. Asymmetric information concerning physician agency adds one more dimension to the problem, but the basic issue is the missing information about the consumer's valuation of health care quantities. Physician agency is a relationship through which an insurer must attempt to extract this missing information.

In this paper, we have maintained the assumption of minimum profits for the physician. Under this assumption, optimal payment and quantity depend only on the derivative of V. In

11. There is no information rent term G/g for the physician agency in Proposition 3.

Appendix B, we derive the (third-best) optimal mechanism when the minimum profit constraints are replaced by reservation utility constraints. The results there indicate that function V itself would determine the optimal mechanism. If in our model we add or subtract a constant from the utility function V, all results remain unchanged. In Appendix B, we show that this robustness does not apply under reservation utility constraints. The appendix also draws some connection between the model here and the literature on countervailing incentives (LEWIS and SAPPINGTON [1989], JULLIEN [2000]).

VI. Conclusion

We hypothesize that physicians interact with patients in complex ways, and have proposed a model of asymmetric information for the complexity in physician agency. How physician agency weighs physician profit and patient benefit is unknown to the insurer. This, however, is only one piece of missing information. The insurer does not know the patient's valuation for health care either. We study the optimal mechanism when these two pieces of information are unknown to the insurer.

We view the design of an optimal mechanism as an attempt to base payment and quantity on a patient's valuation of health care benefit. The insurer recognizes that this valuation information is unavailable. What is more, an attempt to extract this information must face the difficult task of resolving the complexity of physician agency. It is only through the physician agency that an insurer can get to this information.

The optimal mechanism exhibits properties commonly found in managed care. For example, the insurer imposes a fixed quantity for an extensive range of patient characteristics. This is due to the information about patient valuation being too costly to extract. More important, any variation of quantities is related to physician agency characteristics. The optimal mechanism cannot tie quantities directly to intrinsic patient valuation. Two patients with identical health care problems may receive different health care quantities depending on the particular relationship each has with her physician.

The complexity of physician-patient interactions affects many aspects of the study of the health market. The usual program of inducing cost efficiency and service quality necessarily assumes some provider objective. Usually, the objective is assumed to be known. Obviously, when uncertainty of provider objectives is present, the optimal mechanism will have to consider tradeoff differently. Moreover, problems such as dumping and skimping have to be reconsidered if the physician agency shows some preferences toward patient welfare or benefits, and if these preferences are private information.

We have assumed that the physician agency relationship is stable, and the consumer delegates decisions to the physician. Consumers would often like to search for the "right" doctors. The matching process between physicians and patients, and competition among physicians for patients are interesting and important research questions.

ACKNOWLEDGEMENTS

We thank Richard Frank, Yuk-fai Fong, Billy Jack, Tom McGuire, Joe Newhouse, and Bernard Salanié for their suggestions. We are grateful to many seminar participants for their comments. We also thank editor Alain Trannoy and two referees for their helpful advice. The second author acknowledges support from the National Institute of Mental Health through grant RO1-MH-60894, and the Agency for Healthcare Research and Quality through grant PO1-HS-10803. Opinions here belong to the authors.

Correspondence:
Philippe Choné (corresponding author)
CREST-LEI and CNRS UMR 2773 15 boulevard Gabriel Péri, 92245 Malakoff, France
chone@ensae.fr

Ching-to Albert Ma
Boston University Department of Economics, Boston University, 270 Bay State Road, Boston, Massachusetts 02215, USA
ma@bu.edu

Appendix

A. Proof of Lemmas and Propositions

Proof of Lemma 1. Consider an incentive compatible mechanism $(q(\alpha,\beta),R(\alpha,\beta))$. Recall that the indirect utility is defined by

$$U(\beta)=\max_{\alpha',\beta'}\{R(\alpha',\beta')-C(q(\alpha',\beta'))+\beta V(q(\alpha',\beta'))\}.$$

Since U is the upper bound of affine functions of β, it is convex in β (ROCKAFELLAR [1972, Theorem 5.5]). Therefore U is differentiable for almost every β in $[\underline{\beta},\overline{\beta}]$ (ROCKAFELLAR [1972, Theorem 25.5]).

For $\beta<\beta'$ and for all α,α', the incentive constraints imply

$$V(q(\alpha,\beta))\leq\frac{U(\beta')-U(\beta)}{\beta'-\beta}\leq V(q(\alpha',\beta')). \tag{A.1}$$

Therefore, for $\beta'<\beta<\beta''$ and all α'

$$\frac{U(\beta)-U(\beta')}{\beta-\beta'}\leq\min_{\alpha'}V(q(\alpha',\beta))\leq\max_{\alpha'}V(q(\alpha',\beta))\leq\frac{U(\beta'')-U(\beta)}{\beta''-\beta}.$$

As $\beta'\to\beta^-$ and $\beta''\to\beta^+$, the left and right derivatives of U at β satisfy

$$\left(\frac{\mathrm{d}U}{\mathrm{d}\beta}\right)_-\leq\min_{\alpha'}V(q(\alpha',\beta))\leq\max_{\alpha'}V(q(\alpha',\beta))\leq\left(\frac{\mathrm{d}U}{\mathrm{d}\beta}\right)_+. \tag{A.2}$$

If $\min_{\alpha'} V(q(\alpha',\beta)) < \max_{\alpha'} V(q(\alpha',\beta))$, U would *not* be differentiable at β. But U is differentiable for almost every β. So we must have $\min_{\alpha'} V(q(\alpha',\beta)) = \max_{\alpha'} V(q(\alpha',\beta))$, or $q(\alpha,\beta) = q(\alpha',\beta)$ for almost every β. In turn, this implies $R(\alpha,\beta) = R(\alpha',\beta)$ for almost every β (since U does not depend on α). □

Proof of Corollary 1. On the set (of zero Lebesgue measure) where U is not differentiable, we change the schedule as follows. For any β in this set, we pick any α_0 and replace, for all α', $(q(\alpha',\beta), R(\alpha',\beta))$ by $(q(\alpha_0,\beta), R(\alpha_0,\beta))$. In other words, we select an arbitrary value in the subgradient of U at β, under the single constraint that this value is the *same* for all α. This selection does not change the value of the objective function, since it only affects a set of zero measure. The resulting schedule depends only on β, leads to the same value of the objective function as before, and is incentive compatible. It follows that without loss of generality we can consider only schedules depending only on β. For almost every β, the indirect utility function is differentiable at β and its derivative is $V(q(\beta))$ (see equation (A.2) or simply apply the Envelope Theorem). □

Proof of Lemma 2. The first part of Lemma 2 follows from the proof of Lemma 1. We now prove the second part. So consider a pair (q,U), with U convex and $U' = V(q)$. Define the payment R by $R(\beta) = U(\beta) + C(q(\beta)) - \beta V(q(\beta))$. The incentive compatibility constraint is

$$R(\beta) - C(q(\beta)) + \beta V(q(\beta)) \geq R(\beta') - C(q(\beta')) + \beta V(q(\beta')).$$

After substituting R by $U(\beta) + C(q(\beta)) - \beta V(q(\beta))$, the incentive constraint becomes

$$U(\beta) \geq U(\beta') + V(q(\beta'))(\beta - \beta').$$

The inequality in the above is valid because $U'(\beta') = V(q(\beta'))$ and U is convex. So (q,R) is incentive compatible. □

Proof of Lemma 3. Since the profit π is nonincreasing, $\pi(\beta) \geq 0$ for all β is equivalent to $\pi(\overline{\beta}) \geq 0$. Moreover, if there exists $\hat{\beta} < \overline{\beta}$ such that $\pi(\hat{\beta}) = 0$, then for all $\beta \geq \hat{\beta}$, we must have $\pi(\beta) = 0$. On that interval, the profit is identically 0, so its derivative $\pi'(\beta) = \beta V'(q(\beta))q'(\beta)$ must be zero as well. This implies that q is constant on that interval. □

Proof of Lemma 4. Suppose that q jumps upward at $\hat{\beta}$; that is, $\hat{q} > q(\hat{\beta}_-)$, where $q(\hat{\beta}_-)$ is given by (9) at $\beta = \hat{\beta}$. Then we could slightly increase $\hat{\beta}$, while keeping $\hat{q}$ constant. This change would respect the monotonicity requirement. The impact on the objective function is given by

$$\left(\frac{\partial W}{\partial \hat{\beta}}\right)_{\hat{q}\text{ fixed}} = \left[\alpha_{\mathrm{m}}(\hat{\beta}) + \hat{\beta} + \frac{G(\hat{\beta})}{g(\hat{\beta})}\right][V(q(\hat{\beta}_-)) - V(\hat{q})] - [C(q(\hat{\beta}_-)) - C(\hat{q})]. \tag{A.3}$$

Using $\alpha_{\mathrm{m}}(\hat{\beta}) + \hat{\beta} + G/g(\hat{\beta}) = C'(q(\hat{\beta}_-))/V'(q(\hat{\beta}_-))$ and the assumption that $\hat{q} > q(\hat{\beta}_-)$, we know that the above derivative is strictly positive (recall that the cost function C is strictly convex and the benefit V is strictly concave). So increasing $\hat{\beta}$ would increase the objective function. We conclude that q is continuous at $\hat{\beta}$. □

Proof of Lemma 5. We consider a small variation in $\hat{q}$. In principle, we have to change $\hat{\beta}$ accordingly to respect the monotonicity requirement at $\hat{\beta}$. Nevertheless, due to the continuity of q (see equation (A.3) with $\hat{q} = q(\hat{\beta})$), the direct impact of the induced variation in $\hat{\beta}$ on the objective function is zero:

$$\left(\frac{\partial W}{\partial \hat{\beta}}\right)_{\hat{q}\text{ fixed}} = 0.$$

So the total impact of the change is

$$\frac{\partial W}{\partial \hat{q}} = \int_{\hat{\beta}}^{\overline{\beta}} \left\{\alpha_{\mathrm{m}}(\beta)V'(\hat{q}) - C'(\hat{q})\right\} g(\beta)\mathrm{d}\beta - \hat{\beta}G(\hat{\beta})V'(\hat{q}),$$

which gives equation (11) and achieves the proof of the Lemma. □

Proof of Lemma 6. Use equations (10) and (11) to eliminate $\hat{q}$. After simplifying and applying integration by parts, we obtain the following equation for $\hat{\beta}$

$$\int_{\hat{\beta}}^{\overline{\beta}} \left\{\left[\alpha_{\mathrm{m}}(\beta) + \beta + \frac{G(\beta)}{g(\beta)}\right] - \left[\alpha_{\mathrm{m}}(\hat{\beta}) + \hat{\beta} + \frac{G(\beta)}{g(\hat{\beta})}\right]\right\} g(\beta)\mathrm{d}\beta - \overline{\beta} = 0. \tag{A.4}$$

By Assumption A, the left-hand side of (A.4) is continuous and nonincreasing in $\hat{\beta}$; it is equal to $-\overline{\beta}$ at $\hat{\beta} = \overline{\beta}$. Also, it is equal to $\alpha_{\mu} - \alpha_{\mathrm{m}}(\underline{\beta}) - \underline{\beta}$ at $\hat{\beta} = \underline{\beta}$, where α_{μ} is the unconditional mean of α. So as $\hat{\beta}$ varies between $\underline{\beta}$ and $\overline{\beta}$, the left-hand side of (A.4) varies between $\alpha_{\mu} - \alpha_{\mathrm{m}}(\underline{\beta}) - \underline{\beta}$ and $-\overline{\beta}$.

If $\alpha_{\mu} - \alpha_{\mathrm{m}}(\underline{\beta}) - \underline{\beta} > 0$, there exists a unique solution $\hat{\beta}$ to (A.4) satisfying $\underline{\beta} < \hat{\beta} < \overline{\beta}$. Otherwise, if $\alpha_{\mu} - \alpha_{\mathrm{m}}(\underline{\beta}) - \underline{\beta} \leq 0$, there exists no value of $\hat{\beta}$ between $\underline{\beta}$ and $\overline{\beta}$ to fulfill (A.4), and we have a corner solution $\hat{\beta} = \underline{\beta}$. □

Proof of Proposition 2. Equation (17) follows from (9) and (11):

$$\begin{aligned}\int_{\underline{\beta}}^{\overline{\beta}} \frac{C'(q(\beta))}{V'(q(\beta))} g(\beta)\mathrm{d}\beta &= \int_{\underline{\beta}}^{\hat{\beta}} \left[\alpha_{\mathrm{m}}(\beta) + \beta + \frac{G(\beta)}{g(\beta)}\right] g(\beta)\mathrm{d}\beta + \int_{\hat{\beta}}^{\overline{\beta}} \alpha_{\mathrm{m}}(\beta)g(\beta)\mathrm{d}\beta - \hat{\beta}G(\hat{\beta})\\ &= \alpha_{\mu}.\end{aligned}$$

From equation (9), $C'(q)/V'(q) \geq \alpha_{\mathrm{m}}(\beta)$ for $\beta \leq \hat{\beta}$. Now the existence of $\tilde{\beta} > \hat{\beta}$ where $C'(\hat{q})/V'(\hat{q}) \leq \alpha_{\mathrm{m}}(\beta)$ on $(\tilde{\beta}, \overline{\beta})$ follows from equation (17). □

Proof of Proposition 3. For a given level of utility U, we choose each $q(\alpha)$ to maximize the objective function subject to the minimum profit constraints. Pointwise maximization leads to equation (21), and we must check that it satisfies the minimum profit constraints. Let us define $\bar{q}(U)$ by $\beta V(\bar{q}(U)) = U$; for later use, note that

$$\bar{q}'(U) = \frac{1}{\beta V'(\bar{q}(U))}.$$

Then any $q(\alpha)$ satisfying (21) is an optimal quantity if and only if $q(\alpha) \leq \bar{q}(U)$. It is easy to verify that any $q(\alpha)$ satisfying (21) is increasing in α. Accordingly, we can define $\hat{\alpha}(U)$ such that for $\alpha \leq \hat{\alpha}(U)$, (21) holds, and the value of $\hat{\alpha}(U)$ is given by

$$\frac{C'(\bar{q}(U))}{V'(\bar{q}(U))} = \hat{\alpha}(U) + \beta. \tag{A.5}$$

Now the objective function can be written as a function of U alone:

$$W = \int_{\underline{\alpha}}^{\hat{\alpha}(U)} \{(\alpha+\beta)V(q(\alpha)) - C(q(\alpha))\} f(\alpha)\mathrm{d}\alpha \\ + \int_{\hat{\alpha}(U)}^{\overline{\alpha}} \{(\alpha+\beta)V(\bar{q}(U)) - C(\bar{q}(U))\} f(\alpha)\mathrm{d}\alpha - U,$$

where $q(\alpha)$ satisfies (21) on $[\underline{\alpha}, \hat{\alpha}(U)]$. Differentiating with respect to U yields

$$\begin{aligned} W'(U) &= \bar{q}'(U) \int_{\hat{\alpha}(U)}^{\overline{\alpha}} \{(\alpha+\beta)V'(\bar{q}(U)) - C'(\bar{q}(U))\} f(\alpha)\mathrm{d}\alpha - 1 \\ &= \frac{1}{\beta} \int_{\hat{\alpha}(U)}^{\overline{\alpha}} \left\{(\alpha+\beta) - \frac{C'(\bar{q}(U))}{V'(\bar{q}(U))}\right\} f(\alpha)\mathrm{d}\alpha - 1, \end{aligned}$$

where the second equality follows from the definition of $\bar{q}'(U)$.

The functions $\bar{q}(U)$ and $\hat{\alpha}(U)$ are nondecreasing with respect to U. It follows that W' is nonincreasing and, therefore, W is a concave function of U. The optimal level of U is given by $W'(U) = 0$, if such a U exists. This yields equation (22).[12]

To prove the first part of the Proposition, recall that $\hat{\alpha}(U)$ is nondecreasing in U: the higher the value of U, the larger the pooling interval. There is complete pooling if and only if $W'(U) \leq 0$ for a U where $\hat{\alpha}(U) = \underline{\alpha}$. For such a value of U, we have, from (A.5),

$$W'(U) = \frac{1}{\beta}\left[\alpha_\mu + \beta - \frac{C'(\bar{q}(U))}{V'(\bar{q}(U))}\right] - 1 = \frac{1}{\beta}[\alpha_\mu - \hat{\alpha}(U)] - 1 = \frac{1}{\beta}[\alpha_\mu - \underline{\alpha}] - 1.$$

So when $W'(U) \leq 0$, or $\alpha_\mu - \underline{\alpha} \leq \beta$ as in the first part of the Proposition, it is a corner solution. The physician earns zero profit, while the optimal quantity satisfies (20). □

B. Minimum Profit versus Reservation Utility Constraints

To assess the role of minimum profit, we solve a version of the model with a reservation utility constraint for the physician (or the physician agency). We assume that the reservation utility does not depend on the agency parameter β; without loss of generality, we let the reservation utility be 0, and a mechanism must guarantee a nonnegative indirect utility. That is, $U \equiv \beta V(q(\beta)) + R(\beta) - C(q(\beta)) \geq 0$. Results turn out to be rather different under this constraint. First, the magnitude of V matters, while it does not when a minimum profit constraint

12. If a physician can incur a loss $L < 0$, the minimum profit constraint is given by $\beta V(q(\alpha)) \leq U + L$. In the proof of Proposition 3, we must replace U by $U + L$. Since U is endogenous, the optimal value of α, which determines the pooling interval, is unchanged. We simply replace the payment R by $R - L$.

is used instead (results in the propositions above depend only on the derivative of V, or the marginal utility). Although the function V can plausibly be interpreted as valuation and expressed in monetary terms, formally V may just be an ordinal measure of the patient's benefit. So adding or subtracting a constant from V should not have a bearing on economic principles, but this is not true.[13] In other words, we argue here that a minimum profit constraint is more appealing.

Suppose first the function V is everywhere positive. Then the indirect utility is increasing. It follows that the reservation utility constraint binds at $\underline{\beta}$. Integrating by parts the utility term ($\int Ug = \int U'(1-G)$) yields

$$W = \int_{\underline{\beta}}^{\overline{\beta}} \left\{ \alpha_{\mathrm{m}}(\beta)U' - C(V^{-1}(U')) + \beta U' - \frac{1-G}{g}\, U' \right\} g(\beta)\mathrm{d}\beta.$$

Maximizing pointwise with respect to U' leads to

$$\frac{C'(q)}{V'(q)} = \beta + \alpha_{\mathrm{m}}(\beta) - \frac{1-G(\beta)}{g(\beta)}. \tag{B.1}$$

The right-hand side of (B.1) is increasing in β under Assumption A2 and $(1-G)/g$ nonincreasing. So the quantity schedule satisfying (B.1) is incentive compatible, and therefore optimal. The optimal quantity schedule exhibits no pooling.

Suppose now that V is everywhere negative. Then U is decreasing and the reservation utility constraint binds at $\overline{\beta}$. Integrating by parts the utility term ($\int Ug = -\int U'G$) and maximizing pointwise yields

$$\frac{C'(q)}{V'(q)} = \beta + \alpha_{\mathrm{m}}(\beta) + \frac{G(\beta)}{g(\beta)}, \tag{B.2}$$

which is increasing under Assumption A. This is therefore the solution.

Finally, suppose that $V(q)$ is negative for small q and positive for large q. Let q_1 be defined by $V(q_1) = 0$. Since $U' = V(q)$, the indirect utility U first decreases and then increases. So suppose that U is decreasing on $[\underline{\beta}, \beta_1]$, constant on $[\beta_1, \beta_2]$, and increasing on $[\beta_2, \overline{\beta}]$, $\beta_1 \leq \beta_2$. Because the reservation utility constraint must bind, $U(\beta) = 0$ for all $\beta \in [\beta_1, \beta_2]$.

We now show that $\beta_1 < \beta_2$. Suppose to the contrary that $\beta_1 = \beta_2$. Integrating by parts on the two intervals and maximizing with respect to q leads to the following: the quantity is given by (B.2) on $[\overline{\beta}, \beta_1]$ and by (B.1) on $[\beta_1, \overline{\beta}]$. This leads to a downward discontinuity of U at β_1, violating the monotonicity of q. It follows that $\beta_1 < \beta_2$.

Since $U(\beta) = 0$ for all $\beta \in [\beta_1, \beta_2]$, we have $U' = V(q) = 0$ on that interval and $q(\beta) = q_1$. By the same computations (integration by parts and pointwise maximization), we conclude that the optimal quantity is given by

$$\frac{C'(q)}{V'(q)} = \begin{cases} \beta + \alpha_{\mathrm{m}}(\beta) + \frac{G(\beta)}{g(\beta)} & \text{if } \beta \leq \beta_1 \\[2ex] \dfrac{C'(q_1)}{V'(q_1)} & \text{if } \beta_1 \leq \beta \leq \beta_2 \\[2ex] \beta + \alpha_{\mathrm{m}}(\beta) - \frac{1-G(\beta)}{g(\beta)} & \text{if } \beta \geq \beta_2, \end{cases}$$

13. For example, results in this paper remain unchanged if we replace the function V by $V - 10,000$. This no longer holds true if a reservation utility constraint replaces our minimum profit constraint.

where β_1 and β_2 are given by

$$\frac{C'(q_1)}{V'(q_1)} = \beta_1 + \alpha_{\mathrm{m}}(\beta_1) + \frac{G(\beta_1)}{g(\beta_1)} = \beta_2 + \alpha_{\mathrm{m}}(\beta_2) - \frac{1-G(\beta_2)}{g(\beta_2)}.$$

When the reservation utility constraint $U \geq 0$ replaces the minimum profit constraint $\pi \geq 0$, pooling may result; any pooling must occur in the strict interior of the support of β. Nevertheless, the reason for pooling is very different. Because of the change of the sign of V, there are countervailing incentives, as in LEWIS and SAPPINGTON [1989]. For small β, the physician has an incentive to under-report β while the opposite is true for high β. Generally, when a reservation utility constraint is imposed, the solution depends on the sign of V. By contrast, under the minimum profit constraint, the solution only depends on V'.

References

ARMSTRONG, Mark, and Jean-Charles ROCHET (1999): "Multidimensional Screening: A User's Guide", *European Economic Review*, **43**, 959-979 [230]

ARROW, Kenneth J. (1963): "Uncertainty and the Welfare Economics of Medical Care," *American Economic Review*, **53(5)**, 941-969 [231]

BARDEY, David, and Romain LESUR (2006): "Optimal Regulation of Health System with Induced Demand and Ex post Moral Hazard," *Annales d'économie et de statistique*, **83-84**, 279-293 [234]

BAUMGARDNER, James (1991): "The Interaction between Forms of Insurance Contract and Types of Technical Change in Medical Care," *Rand Journal of Economics*, **22(1)**, 36-53 [231]

BESLEY, Timothy, and Maitreesh GHATAK (2005): "Competition and Incentives with Motivated Agents," *American Economic Review*, **95(3)**, 616-636 [232]

CHALKLEY, Martin, and James M. MALCOMSON (1998): "Contracting for Health Services when Patient Demand Does Not Reflect Quality," *Journal of Health Economics*, **17(1)**, 1-20 [229]

DEGROOT, Morris H. (1986): *Probability and Statistics*, second edition, Addison Wesley, Reading Mass [245]

DIXIT, Avinash (2005): "Incentive Contracts for Faith-Based Organizations to Deliver Social Services," in Sajal Lahiri and Pradip Maiti, *Economic Theory in a Changing World: Policymaking for Growth*, New Delhi: Oxford University Press, **2005**, 155-163 [232]

DRANOVE, David (1988): "Demand Inducement and the Physician/Patient Relationship," *Economic Inquiry*, **26**, 251-298 [231]

DRANOVE, David, and Kathryn E. SPIER, (2003): "A Theory of Utilization Review," *Contributions to Economic Analysis and Policy*, **2(1)**, article 9 [229]

DUSHEIKO, Mark; Hugh GRAVELLE, Rowena JACOBS, and Peter SMITH (2006): "The Effect of Budgets on Gatekeeping Doctor Behavior: Evidence from a Natural Experiment," *Journal of Health Economics*, **25(3)**, 449-478 [229]

ELLIS, Randall P. (1998): "Creaming, Skimping, and Dumping: Provider Competition on the Intensive and Extensive Margins," *Journal of Health Economics*, **17**, 537-555 [231]

ELLIS, Randall P., and Thomas G. MCGUIRE (1986): "Provider Behavior Under Prospective Reimbursement," *Journal of Health Economics*, **5**, 129-151 [229]

ELLIS, Randall P., and Thomas G. McGuire, (1990): "Optimal Payment Systems for Health Services," *Journal of Health Economics*, **9**, 375-396 [229]

FRANK, Richard G., Jacob Glazer, and Thomas G. MCGUIRE (2000): "Measuring Adverse Selection in Managed Health Care," *Journal of Health Economics*, **19**, 829-854 [231]

GLAZER, Jacob, and Thomas G. MCGUIRE (2000): "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care," *American Economic Review*, **90(4)**, 1055-1071 [231]

JACK, William (2005): "Purchasing Health Care Services from Providers with Unknown Altruism," *Journal of Health Economics*, **24(1)**, 73-93 [232]

JULLIEN, Bruno (2000): "Participation Constraints in Adverse Selection Models," *Journal of Economic Theory*, **93**, 1-47 [248]

KEELER, Emmett B., Grace CARTER, and Joseph P. NEWHOUSE(1998): "A Model of the Impact of Reimbursement Schemes on Health Plan Choice," *Journal of Health Economics*, **17(3)**, 297-320 [231]

LEWIS, Tracy R., and David SAPPINGTON (1989): "Countervailing Incentives in Agency Problems," *Journal of Economic Theory*, **49**, 294-313 [248]

MA, Ching-to Albert (1994): "Health Care Payment Systems: Cost and Quality Incentives," *Journal of Economics & Management Strategy*, **3(1)**, 93-112 [231]

MA, Ching-to Albert (1998): "Incentius de cost i qualitat en l'assisténcia sanitária. Proveidors altruistes," in Guillem Lopez-Casasnovas, *La Contractació de Serveis Sanitaris*, Generalitat de Catalunya, 65-80 (English version "Cost and Quality Incentives in Health Care: Altruistic Providers" available at http://www.bu.edu/econ/faculty/ma/Papers_Archive/Q-C_ALT.pdf) [229]

MA, Ching-to Albert, and Thomas G. MCGUIRE (1997): "Optimal Health Insurance and Provider Payment," *American Economic Review*, **87(4)**, 685-704 [231]

MA, Ching-to Albert, and Michael H. RIORDAN (2002): "Health Insurance, Moral Hazard, and Managed Care," *Journal of Economics & Management Strategy*, **11(1)**, 81-108 [229]

MCGUIRE, Thomas G. (2000): "Physician Agency," in A.J. Cuyler, and J.P. Newhouse, *Handbook of Health Economics*, North-Holland, 467-536 [229,231]

NEWHOUSE, Joseph P. (1970): "Toward a Theory of Nonprofit Institutions: An Economic Model of a Hospital," *American Economic Review*, **60**, 64-74 [229]

ROCKAFELLAR, R T. (1972): *Convex Analysis, Princeton Univ. Press*, second printing [249]

ROCHAIX, Lise (1989): "Information Asymmetry and Search in the Market for Physician Services," *Journal of Health Economics*, **8**, 53-84 [229]

ROCHET, Jean-Charles, and Philippe CHONÉ (1998): "Ironing, Sweeping and Multidimensional Screening," *Econometrica*, **66(4)**, 783-826 [230]

ROGERSON, William P. (1994): "Choice of Treatment Intensity by a Nonprofit Hospital under Prospective Pricing," *Journal of Economics & Management Strategy*, **3(1)**, 7-51 [229]