DEPARTAMENTO DE ECONOMIA
PUC-RIO

TEXTO PARA DISCUSSÃO
N^{o}. 413

SHOULD WE BE AFRAID OF MANAGED CARE? A THEORETICAL ASSESSMENT

MARCOS B. LISBOA
HUMBERTO MOREIRA

JANEIRO 2000

Should we be Afraid of Managed Care? a theoretical assessment*

Marcos B. Lisboa† Humberto Moreira‡

January 31, 2000

Abstract

Managed care capitation contracts provide monetary incentives for doctors to save medical costs while standard health insurance contracts do not. The paper proposes an alternative model for insurance markets which is used to analyze managed care contracts. In our model, households would like to buy insurance for the possible need of a service. The distinctive aspect of our model is that providers of service have privileged information on the most appropriate procedure to be followed. In the managed care application of the model, doctors are the providers of the service and through a diagnosis have better information of the patient's health condition.

Equilibrium in our model is always constrained efficient. A partial capitation contract arises when both the cost and net benefits of treatment are high enough. We show that a capitation contract provides incentives for doctors: i) to care about the likelihood households will obtain the good state of nature *(altruistic behavior)*; and ii) to save medical costs *(managed care behavior)*. Doctors, in this case, choose less medically efficient treatments as they would choose under a standard health insurance contract. Besides this, household' welfare is increased in comparison to the standard contract. This increased welfare translates into a revealed preference for the capitation contract.

*We would like to thank Francisco Ferreira, Ilan Goldfajn, Marco Bonomo and Monica Viegas Andrade for many useful comments on an early version of the paper. The usual disclaimers apply.

†Graduate School of Economics, Getúlio Vargas Foundation.

‡Department of Economics, Catholic University of Rio de Janeiro.

1 Introduction

1.1 Managed Care

The standard health economics model usually assumes the existence of asymmetric information between insurance companies and patients.[1] Patients may know more about their health than insurance companies do and may tend to use health services more intensively if they have access to a full coverage. These assumptions generate some of the stylized facts observed in health insurance contracts, such as the requirement for co-payment and the supply of partial insurance contracts, even though households typically are risk-averse, while insurance companies may be risk-neutral.

In the traditional health insurance contract, *indemnity plans*, doctors are usually paid per service performed *(fee-for-service)*. In a full insurance contract, however, patients do not have incentives account for the cost of service when deciding whether or not they should consult a doctor. Because of that, they tend to see a doctor more often than it would be optimal.

The moral hazard problem presented in the relation between insurance companies and patients is solved through a risk-sharing contract.[2] Patients are responsible for a share of the service cost each time they see a doctor. This share is usually non-linear, having a lump-sum amount, paid at each visit, plus a percentage of the total expenditure. Contracts may also set a total amount every year to be paid by the insuree before being able to make use of the insurance. On the other hand, contracts between insurance companies and health providers (doctors and hospitals) usually simply establish the payment for services provided. Providers have freedom to propose needed treatments and there are no monetary incentives to limit expenditures.

Since the early sixties a steady increase in private health insurance costs have been observed, going from 3.89 percent of GNP in 1960 to 7.22 in 1990.[3] This tendency was one of the motivations for the Health Maintenance Act of 1973 and some further legal changes in the early 1980s that overturned existing restrictions on specific health insurance contracts. A main motivation for these decisions was to increase competition in the provision of health insurance contracts by allowing a large set of contracts, especially

[1] For a survey on health economics literature, see Zweifel and Breyer (1997) and Newhouse (1996).

[2] Arrow (1963) discusses in detail several possible market failures in the provision of health insurance.

[3] See Health Care Financing Administration (1998).

in the relationship among insurance companies, patients and providers of health services.[4]

Since the Health Maintenance Act, the typical health insurance contract has been subject to several transformations not explained by the standard model.[5] A central aspect of these transformations, usually labeled as the *managed care revolution*, has been the introduction of risk-sharing contracts between doctors and insurance companies.[6] Since doctors are typically risk-averse, this introduction suggests the need to provide incentives for doctors to take into account treatment costs while considering patient health and choosing the appropriate treatment.

Managed care's main objective is to rationalize the use of health services and, with respect to contracts with doctors, is characterized by two central aspects. First, patients have to choose a primary care physician. In order to have access to any medical service, they must first see their primary care physician, who will then determine the need to further exams or treatment.

Second, insurance companies establish a risk-sharing contract with the primary care physician, who receives a fixed amount a month per patient, independently of services provided. Besides that, insurance companies also establish several funds to provide payments for specialists, hospitals and exams. These funds are common to a group of doctors, who use them to pay for their patients ' treatment. Each time a doctor determines the need for a procedure, the payment is withdrawn from the appropriate fund. If at the end of the year there is a positive balance in these funds, this balance is distributed to the doctors in an inverse ratio to the expenditures made by each doctor. These contracts are usually referred to as *partial capitation.*[7]

There are several types of health insurance companies in the managed care system. They differ as to the contract offered and benefits provided. The most common ones are HMOs (Health Maintenance Organizations), PPOs (Preferred Provided Organizations) and POS (Point of Service). Besides specific differences, all types share the basic innovation of managed care, an insurance contract that shares the risk between the insurance com-

[4] For arguments in that direction, see Enthoven (1993).

[5] Brown (1983) summarizes government regulatory impacts on the development of managed care.

[6] Glied (1999) summarizes the empirical literature on managed care.

[7] Fee-for-service contracts and salary payments also exist under managed care in some circumstances (Glied, 1999). Capitation, however, is a major innovation of managed care and will be the focus of the paper. We later on provide sufficient conditions for either capitation or fee-for-service contracts to emerge as the equilibrium outcome. Moreover, full capitation means that doctors bear all the risk. In our model, capitation is *always* partial.

pany and the doctor.[8]

Managed care has been successful in controlling expenses and increasing its market share, displaying not only lower average costs but also lower rate of cost growth. Luft (1981) estimates that in some cases managed care costs are 10 to 40% lower than standard health contract costs. Between 1996 and 1997, for example, the cost of a typical HMO contract increased 1.03% while the typical health insurance contract increased 2.9%.[9] Furthermore, between 1992 and 1997, the managed care market share increased from 51% to 73%.[10] All these results suggest that the risk-sharing contract between doctors and insurance companies has addressed an incentive problem in the relation between doctors and insurance companies which ends up reducing insurance costs and increasing patient' welfare.

There are several controversies whether or not managed care provides an adequate insurance contract for patients and if contracts are well specified or not. In several instances, patients complaint about services not covered or delays in getting treatment. Very often, patients complain that doctors under managed care do not provide the best medical treatment available in comparison to that provided by the standard indemnity plan.[11]

Despite these controversies, however, it remains true that managed care has reduced health insurance cost growth and provided a service preferred by households in the sense that its market share has displayed a steady increase and most patients who choose a managed care contract later on do not regret the choice, not returning to a fee-for-service contract.[12]

What are the reasons for the managed care revolution? Why do these risk-sharing contracts between doctors and insurance companies seemingly increase overall welfare, leading to this increase in market share? Why are these incentives to doctors needed? From standard economic theory models, these risk-sharing contracts tend not to be optimal. Why, then, has managed care introduced them and why have consumers increasingly preferred these contracts?

[8] See Glied (1999).

[9] There is a controversy on to whether managed care reduces total costs or only selects patients with smaller costs. However, independently of the health cost reduction, some houlseholds prefer capitation contracts even with complaints about the quality of treatment. See Glied (1999), Newhouse (1996) and Newhouse, Schwartz, Williams and Witsberger (1985) for a summary of this discussion.

[10] See Health Care Financing Administration (1998).

[11] On the quality of services under managed care, see Miller and Luft (1997). Glied (1999) summarizes the major findings of the literature on the quality of managed care services.

[12] See Newcomer, Preston and Harrington (1996).

There is an extensive theoretical literature that identify some economic reasons for the appearance of capitation contracts. This literature usually relies on the existence of adverse selection problems in the relation of the insurance company with either patients, doctors or hospitals.[13] In this case, capitation may be used as a screening device to select lower risk, or cost, type reducing the efficiency of the equilibrium outcome.[14]

Suppose however that explicit discrimination of patients that need medical care is not allowed and hospitals have access to the same technology. Why would then a single individual choose a capitation contract? In principle, in this case, both the insurance company and the provider of medical service have access to the same information on patient's health condition at the time the contract is signed. What would be the reason for a risk-sharing contract between the insurance company and the provider of medical service in this case? If the provider is risk-averse this contract would cost more than a fee-for-service contract.

Furthermore, the adverse selection literature does not seem to address the problem of treatment choice. One motivation for capitation contracts seems to be to provide incentives for doctors to choose medical procedures different than the ones they would choose under a fee-for-service contract.

This question is addressed by Ellis and McGuire (1986) and Rogerson (1994). In both cases providers of medical services care to some extent for the patient's welfare. In the first paper it is assumed that doctors care simultaneously for the patient's welfare and hospital's profits. In the second, hospitals choose to maximize gross social output.[15] The appearance of capitation contracts in these models, therefore, is associated with a specific assumption about the provider of service utility function.

We propose in this paper an alternative model that generates partial capitation. However, our model does not rely either on the existence of providers of medical service with different but unobserved cost technologies or on an ex-ante asymmetry of information on the patient's health condition. There is to say, capitation in our model arises even if (i) providers of medical service have all access to the same technology; (ii) providers of medical service care only for their income; (iii) both the insurance company and the provider have the same information on the patient's current health condition when the insurance contract is signed. Therefore, capitation arises in our

[13] The basic models used in this literature are variations either of Rothchild-Stiglitz model or Shleifer's yardstick competition model. See Newhouse (1996).

[14] See Ma (1994) and Newhouse (1996).

[15] We refer to the assumption of providers of medical service caring about patient's welfare as *altruistic assumption*.

model even in the absence of selection problems.[16]

1.2 Asymmetry of information in the health insurance market

The paper's main objective is to provide a theoretical model to address the questions raised by the managed care revolution. The model essential assumption is the existence of an information asymmetry in the doctor-insurance company relationship that justifies the establishment of these contracts. Suppose, initially, that doctors care about their income and also about the likelihood of patients being healthy in the future (*good state of nature*). We refer to this as the *altruistic assumption*. Under this assumption, in a fee-for-service contract doctors always choose the treatment that maximizes the likelihood of the good state of nature. In many cases, however, the marginal gain from choosing a better treatment, from a medical point of view, may not offset its cost, from the patient welfare perspective. Indeed, the term "best" here has no economic meaning: the best treatment can actually be very costly and patients may prefer an alternative treatment, which is less effective but cheaper. This result is trivially consistent with the stylized facts from health insurance markets, in particular the overuse, from an economic perspective, of exams and treatments if patients are provided a full insurance system and doctors have a fee-for-service contract.

In order to maximize patient welfare, doctors must take into account not only the likelihood of the good state of nature but also their budget constraints. If doctors always choose the best medical treatment irrespective of its cost, insurance companies anticipate this behavior, including it in the contract's expected cost. Thus, patients end up paying the cost of doctors' behavior.

The capitation contract arises naturally in this framework, which provides doctor incentives to balance the likelihood of good state of nature and the cost of treatment, since their reward depends inversely on how much they spend. This simultaneously reduces the expected cost of treatment and increases the likelihood of the bad state of nature. In an optimal contract, this trade-off is chosen to maximize patient welfare, restricted to the

[16] It is simple to see that the trade-off between efficiency and selection raised by Newhouse (1996) also appears in this model when there are several risk types. Somewhat surpriensinly, the existence of equilibrium in our model generalizes to the case of adverse selection in the providers of medical services. This result contrasts with Rothchild-Stiglitz model. We will deal with that in the sequel of the paper.

required doctor's incentives. Therefore, in this framework the contract optimally generates three basic aspects of managed care: i) incentives for doctors to control costs; ii) the optimal behavior of doctors in some circumstances is not to choose the best medical treatment as they would choose under fee-for-service contracts; and iii) increased patient welfare, which translates into a revealed preference for that contract.

The altruistic assumption, however, is very strong. Why would doctors have direct preferences for patients achieving the good state of nature? We show in the paper that this assumption can be endogenously generated in many circumstances. First, we construct a simple example where doctors may be sued and punished in case bad state of nature happens. The existence of litigation costs leads almost immediately to the altruistic assumption *(defensive medicine)*.[17]

The most interesting case, however, which is the focus of the paper, arises when there is an additional moral hazard problem in the doctor-insurance company relationship.[18] Suppose the quality of the diagnoses performed by the doctor depends upon an effort level which is not directly observable. That effort level may be related with the doctor's effort in keeping up with the medical research literature or the attention he gives to the patient in his office. In any case, in order to provide incentives for the doctor to choose high effort, the optimal contract must offer better rewards in case patients achieve the good state of nature, which generates precisely the altruistic behavior. In doing so, it creates incentives to use the best medical treatment independently of its cost. In order to compensate for that behavior, the optimal contract must also provide incentives for the doctor to take into account total costs: the doctor's payment must be inversely related to the expenditures on the patient treatment.

1.3 Insurance markets for specialized service

The model proposed here is quite general and seems to apply to a variety of circumstances. We consider a model with providers of a service and households who would like to buy insurance against the chance of having to use such a service. Providers have access to privileged information on the

[17] For the evidence that doctors practice defensive medicine, see McClelan and Kessler (1996).

[18] Equilibrium in our model exists even if there are different types of doctors and each doctor's type is private information (adverse selection model). As we show later, contrary to the Rothschild and Stiglitz (1976) model, in our framework the equilibrium outcome can be obtained as a solution of a planner's problem, which always has a solution.

most appropriate procedure to be followed. There is always a chance, however, that providing the service may not be successful. Suppose households have access to a full insurance contract that pays for whatever procedure providers find necessary. If providers may be sued in case the procedure is not successful, or if they are concerned about their reputation, they have incentives to always propose the procedure most likely to be successful irrespective to its cost. In equilibrium, however, insurance companies anticipate this behavior, which results in higher insurance premiums.

We propose that this asymmetry of information lies at the core of the managed care revolution. The model's basic features, however, may also be used to understand several insurance for service arrangements, such as possible repair of durable goods or fixed assets. A central property of the model is that a risk-sharing contract between an insurance company and providers of service may arise every time a household would like to buy insurance for possible services needed and the provider has privileged information. *The rationale of this contract is to make providers internalize not only households ' desire for a successful outcome, but also the cost of alternative procedures.* In some cases, a procedure with a lower likelihood of success may be chosen over one with higher likelihood that nevertheless costs more.

Being more precise, we consider a model with three types of parties: households, providers of a service and insurance companies. Households have uncertainty about the future need for a service, which affects their level of income.[19] Providers, if hired by a household, perform a diagnosis which provides a signal about which service may reduce the likelihood of a loss of income due, for example, to the need to buy another good or a worse health condition. This signal is observed only by the provider of the service. For each signal there is a procedure that maximizes the likelihood of households obtaining the good state of nature in the final period. Before providing the diagnosis, however, providers have to choose an effort level. High effort levels increase the probability of the good state of nature for every signal and procedure chosen. There are risk-neutral insurance companies that offer contracts to both the providers and households.[20]

[19] From the formal point of view, in our model uncertainty on future levels of income is equivalent to uncertainty on future utility levels. Therefore, the bad state of nature can be interpreted either as a loss in income or, more generally, as a loss in welfare due to any other motive, including a health condition.

[20] We abstain in the paper from the moral hazard problem in the relationship between patient and insurance company. Such a problem is well treated by the standard model and it could be easily added in our model without any additional cost, except making the notation even more complex.

The paper provides sufficient conditions for the existence of equilibrium. In order to induce providers to make effort, the optimal contract must specify higher payments in the good state of nature and thus induce them to care about the outcome for the household. This last equilibrium property result is equivalent to the *altruistic assumption.* The equilibrium outcome is characterized by a loss of welfare for the household, in comparison with the first-best outcome, due to the need to encourage providers to choose, on one hand, effort, and the least expensive treatment for some signals on the other. However, the equilibrium outcome is still constrained optimal: given the asymmetry of information, the market outcome is the most efficient one. Risk-sharing contracts do not always produce an equilibrium outcome in our model. We show, however, that they do if costs and net benefits from procedures get very high. This result seems to fit particularly well in the health insurance case, where the growth of managed care has been increasing precisely while new medical technologies have become available, new technologies that are simultaneously more expensive and more effective in treating diseases.[21]

The next section provides a precise description of the model. Section 3 focuses on two leading examples: defensive medicine and risk-neutral providers. Section 4 describes the optimal contract and summarizes our central results. In order to simplify the model's interpretation and convey the central message as a proposal to understand theoretically managed care, we stick to the health insurance market interpretation of the model where service providers are labelled as *doctors* and households are often referred to as *patients.*

In the sequel to this paper we will extend the model to a dynamic setting where insurance companies are restricted in introducing contracts with monetary payments independent of the state of nature. This generalization seems natural since in several instances one does not observe providers's payments to be contingent upon the state of nature. In particular, providers of medical service payments do not seem to depend on whether or not patients get better after being treated. We show that by allowing companies to threaten not to renew the contracts, they may still be able to induce providers to make an effort. In this case, however, providers' expected utility must be higher than their reservation level.

[21] On the increasing cost of new medical technologies, see Glied (1999) and the references there summarized.

2 The basic model

Consider a partial equilibrium model with a single commodity, three periods - *ex-ante, interim, ex-post* - and three types of individuals: *patients, doctors*[22] *and insurance companies.* A patient faces uncertainty about the initial endowment in the ex-post period: there are two individual states of nature and the endowment in the second period is strictly larger than that in the first period. We refer to the patient's endowment in state s by W_s, $s \in \{B, G\}$. Preference regarding consumption bundles is represented by a state-independent utility function $u : \Re_+ \to \Re$, strictly increasing and concave in the differentiable sense:[23]

(H1) $u \in C^2$ *and satisfies: i)* $du > 0$ *; ii)* $d^2u < 0$*. Moreover,*

$$0 < W_B < W_G$$

Notice that from the formal point of view, uncertainty concerning future income levels is equivalent to uncertainty about future utility levels. To see this equivalence, one has only to allow patient utility to be state-dependent. Therefore, a bad state of nature can be interpreted either as a loss in income or as a bad health condition that reduces the patient utility. Under these assumptions of random endowments and strict risk-aversion, the patient is willing to buy an insurance contract, provided that the price of the contract is not much higher than the actuarially fair price. The standard insurance literature assumes the existence of a finite collection of risk-neutral insurance companies, while offer insurance contracts simultaneously and independently. The outcome of the model with at least two companies is to offer actuarially fair contracts, which are accepted by the patients.

The model proposed here departs from the standard literature by assuming the existence of a third type of individual, whom we refer to as *doctor*. A doctor examines a patient and chooses an action, or treatment, in the interim period that affects the patient's probability of the good state of nature. This action is supposed to be perfectly observable, but its effectiveness depends upon a signal privately observed by the doctor.

The time line of the model can be summarized as follows.

[22] There is no change in the results that follow if the service provider is a doctor or as an hospital.

[23] From now on, the symbol d represents the differential of the respective function.

Contracts offered	Choice of effort	Signal	Choice of treatment	State realized Contracts enforced
-------|	-------|	-------|	-------|	-------|
ex-ante	interim 1	interim 2	interim 3	ex-post

In the ex-ante period, contracts are proposed and chosen. The interim period is divided into three sub-periods. In the first doctors choose an effort level. In the second a signal is observed and, in the third, a treatment is proposed. Finally, in the ex-post period the state of nature is revealed and contracts are enforced.

To make the argument precise, suppose that in the interim period the doctor observes a private signal, $s \in [0,1]$. The probability of a signal s is described by the cumulative probability function $F(s) \in C^1$, $dF(s) > 0$ for every $s \in (0,1)$. Without any loss of generality, we assume that F is the uniform distribution. There are 2 types of actions available for the doctor

$$A := \{a_0, a_1\}$$

Action i costs a_i to be implemented in addition to the doctors' payments, $0 = a_0 < a_1 = a$. Action a_0 should be interpreted as the doctor choosing *no action.* Given this signal, the probability of the good state of nature for a patient depends on the action chosen and it is given by $\pi(a,s)$ for every a. We assume:

(H2) For each $a \in A$ the function $\pi(a,\cdot) \in C^2$. Furthermore, for every $s \in (0,1)$

$$0 > d\pi(a_1, s) > d\pi(a_0, s)$$

and

$$\begin{aligned} 0 &< \pi(a_1, 0) < \pi(a_0, 0) < 1 \\ 0 &< \pi(a_0, 1) < \pi(a_1, 1) < 1 \end{aligned}$$

Therefore, the higher the signal the lower is the probability of the good state, and there is a signal $s^* \in (0,1)$ such that both actions generate the same probability of the good state of nature

$$\pi(a_1, s^*) = \pi(a_0, s^*)$$

Moreover, for every signal $s < s^*$, action a^0 generates a higher probability of the good state of nature in the second period, while for every signal $s > s^*$ the reverse happens. We refer to the action that maximizes the probability of the good state of nature as the *best action* or *the best medical treatment.*

Before observing the signal, doctors have to choose an *effort level,* $e \in \{0,1\}$, with associated costs given by

$$0 = c_0 < c_1 = c$$

Let

$$\pi^0 (a, s) = \pi (a, s) \in (0, 1)$$

be the probability of the good state given a, s and $e = 0$, and

$$\pi^1 (a, s) = \Pi (a, s) \in (0, 1)$$

be the probability of the good state given s, a and $e = 1$. We assume that $(H2)$ holds for each π^i and effort always increases the likelihood of the good state of nature. In particular, there exists a uniform diagnosis lag between effort and non-effort: making an effort is equivalent to observing a more favorable signal on the patient's health condition. Furthermore, we also assume that the treatment technology has decreasing returns on the signal.

(H3) (i) $\pi(a_j, \rho s) = \Pi (a_j, s)$, *where* $\rho \in (0, 1)$, *for all* j *and* s;
(ii) the function

$$\frac{d\Pi (a_0, s)}{d\Pi (a_1, s)}$$

is non-increasing.

The second part of assumption $(H3)$ establishes that as the signal increases the marginal gains from treatment with respect to non-treatment does not increase. Indeed, $d\Pi$ describes the loss in the likelihood of the good state as the signal increases. Thus (ii) states that this loss cannot be increasing in no treatment with respect to treatment.

Let $v : \Re_+ \to \Re \in C^2$ be the doctor's utility function, concave in the differentiable sense:

(H4) $\quad dv > 0,\ d^2 v \leq 0,\ \lim_{x \to 0} v(x) = -\infty$ *and* $\lim_{x \to \infty} v(x) = \infty.$

A doctor has a reservation utility denoted by $\bar{v}$, where if $\bar{r}$ satisfies $v(\bar{r}) = \bar{v}$ then $\bar{r} > 0$.

Finally, there are $I > 1$ principals, or *insurance companies*, that provide insurance for the patients and intermediate their relation with the doctors. We suppose that these principals are risk-neutral. In the ex-ante period they offer contracts for the doctors and patients. As in Rothschild and Stiglitz (1976), we assume that the companies are free to enter or exit the market. As we shall see, this assumption implies that in equilibrium, insurance companies make zero expected profits.

3 Leading examples

In this section we consider two simple examples that provide the basic intuition of our model. The need to provide incentives to doctors to save on patient's treatment arises when doctors care about the likelihood of the good state of nature. This care may result either from some exogenous reason (e.g., litigation costs in case bad state happens), or from an endogenous one. In order to provide incentives to doctors to choose high effort levels, the optimal contract must provide larger payments in the good state of nature than in the bad state.[24] We shall exploit both cases in the simple examples that follow and then develop a general model for the second one.

In the first example, doctors may be punished in case they do not choose the best medical treatment *(defensive medicine)*. It is shown that in this case, even without the choice of effort, the basic results of the paper carry over. Punishment, in this case, makes doctors care about the likelihood of patients achieving the good state of nature and, in the absence of any other incentive, always to choose the best treatment. The efficient outcome, however, may require doctors not always to choose the best treatment, since its cost may not compensate the marginal benefit of increasing the probability of the good state of nature. The optimal contract, in this case, will offset the punishment cost, providing larger payments in case doctors choose no treatment.

The second example does not rely on exogenous incentives to doctors to care about patients' good sate of nature. Following our general model, doctors have to choose a non-observable effort level before performing the diagnosis. In order to provide incentives for doctors to choose effort, the optimal contract must provide larger payments in case the good state of nature happens. But then, a problem similar to the one discussed regarding

[24] A second endogenous reason in a dynamic model would be reputation.

defensive medicine arises: doctors will always tend to choose the best medical treatment available, irrespective of its cost. Once more, the efficient outcome requires the optimal contract to also provide incentives so that doctors save medical costs by offering larger rewards in case they choose no treatment.

3.1 Defensive medicine

In this section we consider the leading example of defensive medicine. In this case, doctors may be sued if they do not choose the best medical treatment and the patient does not obtain the good state of nature. We abstract the effort choice in this example. Let α be the expected cost of being sued and $\pi(a, s)$ the probability of the good state of nature given action a and state s. We assume that doctors' loss in income due to not choosing the best medical treatment is proportional to the patient loss in the likelihood of the good state of nature. Let r_j^ω be the doctor payment in case he chooses treatment j and state of nature ω happens. Let $v_j^\omega := v\left(r_j^\omega\right)$ be the doctor indirect utility. The doctor's utility level given a contract r, signal s and choice of treatment a_j is given by

$$\pi(a_j, s)v_j^G + (1 - \pi(a_j, s))\, v\left(r_j^B - \alpha\left(\pi(s) - \pi(a_j, s)\right)\right)$$

where $\alpha > 0$ and $\pi(s) := \max_a \pi(a, s)$ is the probability associated with the best medical treatment. Therefore, if the doctor's payments are not contingent on the choice of treatment, he always chooses the best medical treatment, that is to say a_j so that $\pi(a_j, s) = \pi(s)$ for every s.

Let us consider the critical signal where, for a given contract, doctors change from no treatment to treatment. Suppose doctors always use no treatment when this is the best medical option, but may also use it when treatment is best, that is, when $s \geq s^*$. It is simple to verify that at such a signal the doctor must be indifferent between using treatment or not:

$$\pi(a_0, s)v_0^G + (1 - \pi(a_0, s))\, v\left(r_0^B - \alpha\left(\pi(s) - \pi(a_0, s)\right)\right) = \pi(a_1, s)v_1^G + \pi(a_1, s)\, v_1^B$$

Consider a contract that maximizes patient welfare given the needed incentives for the doctors. We claim that in an efficient contract, payments must be contingent upon the treatment chosen.

Suppose the claim is not true and the efficient contract does not have payments contingent upon the choice of treatment, $r_0^\omega = r_1^\omega$ for every ω. It is simple to verify that for each signal s doctors always choose a_j so that $\pi(a_j, s) = \pi(s)$. Therefore, the critical signal is given by $s = s^*$, where for $s \leq s^*$ no treatment is the best medical option and for $s \geq s^*$ treatment maximizes the likelihood of the good state of nature. In this case, there are no welfare gains for the patient in proposing a contract with payment contingent in the state of nature. Therefore, if the efficient contract implements signal s^*, we must have $r_j^\omega = \bar{r}$ for every ω and j, where $v(\bar{r}) = \bar{v}$.

The expected social benefit for this contract is given by

$$\int_0^{s^*} \left(\pi(a_0, s) W_G + (1 - \pi(a_0, s))\, W_B\right) ds + \int_{s^*}^{1} \left(\pi(a_1, s) W_G + (1 - \pi(a_1, s))\, W_B\right) ds$$

where its cost is given by

$$\bar{r} + a\,(1 - s^*)$$

Therefore, if the contract is efficient, it must satisfy the following first order conditions:

$$\left(\pi(a_0, s^*) W_G + (1 - \pi(a_0, s^*))\, W_B\right) - \left(\pi(a_1, s^*) W_G + (1 - \pi(a_1, s^*))\, W_B\right) = -a$$

Since $\pi(a_0, s^*) = \pi(a_1, s^*)$, we get

$$0 = -a$$

But this is absurd since $a > 0$. This means that an efficient contract must necessarily have payments contingent upon the treatment chosen.

3.2 Risk neutral doctors

In this example we keep all assumptions presented in section 2, including the effort problem, but we assume that the doctor's utility function is given by $v(x) = x$. Consider the following problem for the insurance company: to minimize the cost of implementing effort such that the doctor chooses no treatment for every signal $s \leq \bar{s}$ and treatment afterwards. Let $r := \left(r_j^{\omega}\right)$ be the contract. Let $V_i^{\bar{s}(i)}(r)$ be a doctor utility level given a contract r and with a choice of effort level i and a change from no treatment to treatment at signal $\bar{s}(i)$.

$$\begin{aligned} V_i^{\bar{s}(i)}(r) &= \int_0^{\bar{s}(i)} \left[r_0^B \left(1 - \pi^i \left(a_0, \bar{s}(i)\right)\right) + r_0^G \pi^i \left(a_0, \bar{s}(i)\right)\right] ds + \\ & \int_{\bar{s}(i)}^{1} \left[r_0^B \left(1 - \pi^i \left(a_0, \bar{s}(i)\right)\right) + r_0^G \pi^i \left(a_0, \bar{s}(i)\right)\right] ds \end{aligned}$$

In order for the doctor to accept this contract r, his utility level must be higher than the reservation utility:

$$V_i^{\bar{s}(i)}(r) \geq \bar{v} \quad (IR)$$

Moreover, the doctor will make an effort if the associated utility is not smaller than that of no effort:

$$V_1^{\bar{s}(1)}(r) \geq V_0^{\bar{s}(0)}(r) \quad (IC_1)$$

Finally, if the doctor changes from no treatment to treatment at signal $\bar{s}$, then following condition holds:

$$r_0^G \pi^i \left(a_0, \bar{s}\right) + r_0^B \left(1 - \pi^i \left(a_0, \bar{s}\right)\right) = r_1^G \pi^i \left(a_1, \bar{s}\right) + r_1^B \left(1 - \pi^i \left(a_1, \bar{s}\right)\right) \quad (CT)$$

It is easy to see that the cost of implementing the change of treatment at $\bar{s}$ with effort is

$$c^1(\bar{s}) = \bar{r} + c + \int_{\bar{s}}^{1} a ds$$

while the cost with no effort, $c^0(\bar{s})$, is simply $\bar{r}$.

Suppose social benefit is maximized when doctors choose effort, but as in the previous example, the optimal contract is not contingent upon the choice of treatment, $\Delta r^{\omega} = 0$ for every ω. Then, from (CT) either the critical signal is $s = s^*$ where

$$\Pi(a_0, s^*) = \Pi(a_1, s^*)$$

or $\Delta r_j = 0$ for every j, in which case the same equation must hold and doctors choose no effort. Suppose $s = s^*$. Since the efficient outcome requires effort, $\Delta r_j \neq 0$ for some j.

Since at the efficient outcome, social net benefit must be maximized as in the previous example, we must have

$$0 = d\left(\int_{\bar{s}}^{1} a ds\right) = -a$$

Therefore, once more at the efficient outcome with effort, contracts must specify payments contingent upon the choice of treatment. In the next sections we show that the efficient outcome is actually supported as an equilibrium in our model and that this basic result holds true for the general model. We will also provide several equilibrium properties associated with the optimal contract. Moreover, contrary to the standard moral hazard model, where the risk neutrality of the agent and the principal leads to no agency cost (i.e., the first best allocation is attainable), here this is not true because the signal is non observable.

4 The optimal contract

4.1 Incentives and contracts

We start the analysis of the model investigating the principal's offers to the doctor which specify how much he receives in the interim period. As we will see later, in the optimal contract it may be optimal to make the doctor's payments contingent upon the action chosen. Let r_j^w be the payment the principal makes to the doctor if he chooses a treatment $j \in \{0, 1\}$ and nature chooses $w \in \{B, G\}$. As usual in the literature of moral hazard problems, we will write the contracts in terms of the doctor's utility, i.e., $v_j^{\omega} := v(r_j^{\omega})$. We also define the *power* of the contract: $\Delta v_j = v_j^G - v_j^B$ and $\Delta v^{\omega} := v_1^{\omega} - v_0^{\omega}$, which give the incentives for effort and choice of treatment.

Suppose a doctor has accepted a contract $\left(r_j^\omega\right)$, with indirect utility $\left(v_j^\omega\right)$. His optimal behavior is to choose effort and, for a given signal s, treatment a_j. His expected utility in the interim period, $V_j^i(s)$, is then given by

$$\pi^i(a_j,s)\,v_j^G + \left(1-\pi^i(a_j,s)\right)v_j^B = v_j^B + \pi^i(a_j,s)\,\Delta v_j.$$

Definition 1 *We say that a contract* $\left(v_j^\omega\right)$ *generates a partition* $\left(S_0^i, S_1^i\right)$ *if the treatment* a_j *is chosen in* S_j^i *given the effort* i*, i.e.,*

$$a_j \in \arg\max\ v_j^B + \pi^i(a_j,s)\,\Delta v_j \text{ if and only if } s \in S_j^i.$$

If $\bar{s}$ is a signal where the doctor changes the treatment, then

$$V_0^i(\bar{s}) - V_1^i(\bar{s}) = 0$$

which gives

$$\Delta v^B - \pi^i(a_0,\bar{s})\,\Delta v_0 + \pi^i(a_1,\bar{s})\,\Delta v_1 = 0.$$

Moreover, if $\bar{s}$ represents the change from no treatment $(j=0)$ to treatment $(j=1)$, then

$$-\ d\pi^i(a_0,\bar{s})\,\Delta v_0 +\ d\pi^i(a_1,\bar{s})\,\Delta v_1 \geq 0.$$

The doctor's ex-ante expected utility under the contract $\left(r_j^\omega\right)$, given the partition $\left(S_0^i, S_1^i\right)$ and effort i, V^i, is:

$$\int_{S_0^i}\left[\left(1-\pi^i(a_0,s)\right)v_0^B + \pi^i(a_0,s)\,v_0^G\right]ds + \int_{S_1^i}\left[\left(1-\pi^i(a_1,s)\right)v_1^B + \pi^i(a_1,s)\,v_1^G\right]ds - c_i$$

which reduces to

$$V^i = [\alpha^i - \bar{\pi}_0^i]v_0^B + \bar{\pi}_0^i v_0^G + [1-\alpha^i - \bar{\pi}_1^i]v_1^B + \bar{\pi}_1^i v_1^G - c_i$$

where $\alpha^i := \int_{S_0^i} ds$ is the *ex-ante* probability that the doctor chooses no treatment given a choice of effort i and

$$\bar{\pi}^i_j := \int_{S^i_j} \pi^i \left(a_j, s\right) ds$$

is the *ex-ante* probability of the good state given treatment a_j and effort i.

There are $I > 1$ principals, or *insurance companies*, that provide insurance for the patients and intermediate their relation with the doctors. We suppose that these principals are risk-neutral. In the ex-ante period they offer contracts for the doctors and patients. If an insurance company offers a contract $r = \left(r^w_j\right)$ to the doctor and $d = \{d_{B,} - d_G\}$ to the patient and both accept the contract, then its expected profit will be

$$\begin{aligned} L^i(r,d) &= G^i d_G - (1 - G^i) d_B - \{\left(\alpha^i - \bar{\pi}^i_0\right) r^B_0 + \bar{\pi}^i_0 r^G_0 \\ &\quad + \left(1 - \alpha^i - \bar{\pi}^i_1\right) \left(r^B_1 + a\right) + \bar{\pi}^i_1 (r^G_1 + a)\} \end{aligned}$$

where $G^i = \bar{\pi}^i_0 + \bar{\pi}^i_1$ is the probability of the good state given the contract and effort i. If the patient accepts this offer, his utility is then given by

$$U^i(r,d) = \left(1 - G^i\right) u\left(W_B + d_B\right) + G^i u\left(W_G - d_G\right)$$

Suppose the remaining companies, $k \neq j$, offer contracts (r_k, d_k). The patient would accept the company's offer if

$$U^i(r,d) \geqslant U^i\left(r_k, d_k\right) \quad \text{for all } k \neq j$$

and

$$U^i(r,d) \geqslant U^i(0,0)$$

since the patient may always reject all offers. It is simple to verify that we can restrict the optimal contract analysis to contracts that smooth patient consumption across the states of nature (the wealth of the patient is constant).

On the other hand, the company should guarantee the participation of the doctor and induce the doctor's optimal effort.[25] A doctor who decides

[25] If the service provider is an hospital, the utility function is linear and the (IR) constraint is the break-even condition and everything is the same.

to make an effort i will accept this contract if the participation constraint is satisfied:

$$\alpha^i v_0^B + (1-\alpha^i) v_1^B + \bar{\pi}_0^i \Delta v_0 + \bar{\pi}_1^i \Delta v_1 - c_i \geq \bar{v} \quad (IR)$$

and it induces effort i if it satisfies the incentive constraint:

$$\begin{aligned} &\alpha^i v_0^B + (1-\alpha^i) v_1^B + \bar{\pi}_0^i \Delta v_0 + \bar{\pi}_1^i \Delta v_1 - c_i \geq \\ &\alpha^{-i} v_0^B + (1-\alpha^{-i}) v_1^B + \bar{\pi}_0^{-i} \Delta v_0 + \bar{\pi}_1^{-i} \Delta v_1 - c_{-i} \quad (IC_i) \end{aligned}$$

where $-i \neq i$ is the alternative effort. The insurance company may also decide not to hire a doctor, in which case it can simply offer $r = 0$ and the probability of the good state of nature is given by π (null contract). In this case, the household just buys a full insurance contract that transfers income from the bad to the good state of nature and does not go to see a doctor. Let R be the set of contracts that satisfy the incentive and participation constraints and the null contract.

Therefore, besides the participation constraint on the doctor, there will also be incentive constraints that must be satisfied. In the next subsections we deal with this problem. What is important here is that the decision for effort is taken after he signing the contract and before the doctor sees the signal in the interim period.

4.2 Existence of equilibrium and welfare

Definition 2 *An **equilibrium** is a collection of strategies* $\{(r_i, d_i)_i\}$ *where for each* i *the contract* (r_i, d_i) *solves the problem*

$$\begin{gathered} \max L^i(r,d) \\ \textit{subject to } U(r,d) \geq U(r_k, d_k), \ k \neq i \\ U(r,d) \geq U(0,0) \\ r \in R \end{gathered}$$

Consider an alternative model with the same primitives, except that the doctor's effort level and signal are observable. In this case one can easily verify that at equilibrium doctors receive a fixed payment independently of the treatment chosen, which provides them the reservation utility. Insurance companies, due to competition, make zero profits and the optimal signal is chosen to maximize patient welfare. As usual, this outcome is referred as *Pareto optimal or first best.*

In our model, on the other hand, in some cases insurance companies have to induce doctors to choose a high effort level by offering more rewards in case the good state of nature occurs. These incentives induce doctors to choose the most expensive treatment when that choice maximizes the probability of the good state of nature. However, this behavior may not be optimal from the patient perspective since he also takes into account the effect on the expected cost of treatment and, therefore, the expected cost of insurance. Therefore, the optimal contract may also have to provide incentives for doctors to choose the least expensive treatment in some cases by reducing their payment in case they choose the more expensive (and more effective) one from the health perspective.

We will show later, however, that the market outcome is always constrained optimal, in the sense provided below. Let R^0 denote the set of incentive-compatible contracts which can be contingent upon both the effort level and signal observed by the doctor, and let R^s be the set of incentive compatible contracts which can be contingent only upon the signal observed. The set R^0 corresponds to the model with perfect, symmetric information, and R^s correspond to the standard moral hazard model.

Definition 3 *We say that an equilibrium $\{(r_i, d_i)_i\}$ is **first-best** if there is no other contract (r, d) satisfying*

$$r \in R^0,\ L(r,d) \geq 0 \text{ and } U(r,d) > \max_i U(r_i, d_i)$$

*We say that an equilibrium $\{(r_i, d_i)_i\}$ is **second-best** if there is no other contract (r, d) satisfying*

$$r \in R^s,\ L(r,d) \geq 0 \text{ and } U(r,d) > \max_i U(r_i, d_i)$$

*Finally, we say that an equilibrium $\{(r_i, d_i)_i\}$ is **third-best, or constrained optimal,** if there is no other contract (r, d) satisfying*

$$r \in R,\ L(r,d) \geq 0 \text{ and } U(r,d) > \max_i U(r_i, d_i)$$

Proposition 4 *Every equilibrium is constrained optimal.*

Proof. This proposition follows immediately from the continuity of the insurance companies problem and the assumption that $I > 1$. Suppose there is an equilibrium which is not constrained optimal. Without loss of generality, suppose the patient is buying a contract from insurance company 1 while insurance company 2 is offering an alternative contract which provides the same or less indirect utility for that patient. Since the equilibrium is not constrained optimal, there is an alternative contract, $\{d_B, -d_G\}$, that strictly increases the patient welfare and satisfies the incentive restrictions. By charging $\{d_B, -(d_G + \varepsilon)\}$, where $\varepsilon > 0$ is small enough, this contract still strictly increases the patient welfare and provides strictly positive profits. Since company 2 was making zero profits in trading with this patient, it strictly prefers to offer this contract rather than the one it was offering originally. Therefore, the proposed set of strategies was not an equilibrium, which is the desired result. ■

Proposition 5 *There is an equilibrium.*[26]

Proof. The set R is trivially bounded below and closed. Moreover, it always contains at least the null contract. It is simple to see that if a sequence $\{r_n\}$ satisfies $\|r_n\| \to \infty$, then any contract (r_n, d_n) that satisfies non-negative profits must also satisfy $\|d_n\| \to \infty$. But then, for n large enough, $U(r_n, d_n) < U(0,0)$. Therefore, we can restrict the set of feasible contracts to a bounded set and there is at least one third-best contract. It is simple to verify that this contract must provide zero profit. Consider the following strategy profile for the insurance companies. All companies offer this third-best contract, and patients buy only the contract from company 1. The existence of a profitable deviation for any company would violate the very definition of third-best contract. Therefore, the proposed profile is an equilibrium. ■

4.3 When no effort is the efficient outcome

We start the analysis of the optimal contract studying the case when no effort by the doctor maximizes the social welfare. In this case the optimal contract between doctors and insurance companies has no incentive. More precisely:

Lemma 6 *If $i = 0$ is the constrained optimal effort, then the optimal contract is constant: $r_j^\omega = \bar{r}$ for all ω and j. In this case, the equilibrium is first-best.*

[26] We can show an analogous result here even if there are different types of doctors and each doctor's type is private information (adverse selection model). The existence is guaranted by a generalization of proposition 4, which shows that an equilibrium outcome can be obtained by the solution of a monopolist agency problem where the principal is the patient and the agent is the doctor.

The lemma is proved in the appendix.

By lemma 6, the doctor is indifferent when he is going to change the treatment. In this case we assume that, in equilibrium, the doctor chooses what is the best for the patient.

It is easy to see that the cost of implementing the change of treatment at $\bar{s}$ with no effort is

$$c^0(\bar{s}) = \bar{r} + \left(1 - \alpha^0\right) \; a$$

By assumption $H2$, the best way to implement this change is to use the partition: $S_0 = [0, \bar{s}]$ and $S_1 = [\bar{s}, 1]$.

Lemma 7 *If there is a constrained optimal equilibrium with no effort and the optimal change of treatment s^0 is interior, then*

$$\pi\left(a_1, s^0\right) - \pi\left(a_0, s^0\right) = a / \left(W_G - W_B\right)$$

Reciprocally, if $s^0 \in [0,1]$ satisfies the equation above and no effort is constrained optimal, then s^0 is the optimal change of treatment.

Proof. It is enough to show that such s^0 satisfies the first order condition of the following program:

$$\max_{s \in [0,1]} W_B + \left(W_G - W_B\right) G^0(s) - c^0(s)$$

which is straightforward. Observe that assumption $(H2)$ implies that the program above is a concave one. ■

4.4 The optimal contract with effort

In order to analyze the optimal contract with effort, we follow the usual approach of the firm problem. We start by considering a cost minimization problem and later we discuss the optimal contract problem. The cost minimization problem in this case is as follows. Suppose an insurance company would like to provide incentives so that doctors choose high effort, and to change from no treatment to treatment at $\bar{s}$. Define the inverse of the doctor's utility function: $h = v^{-1}$. The minimum cost to induce this behavior is given by the following optimization problem:

$$c^1(\bar{s}) = \min_{\{v_j^w\}} \alpha^1 h\left(v_0^B\right) + (1-\alpha^1)\left[h\left(v_1^B\right) + a\right] + \bar{\Pi}_0\left[h\left(v_0^B + \Delta v_0\right) - h\left(v_0^B\right)\right]$$
$$+\bar{\Pi}_1\left[h\left(v_1^B + \Delta v_1\right) - h\left(v_1^B\right)\right]$$

$$\text{s.t}\begin{cases} -(\alpha^1-\alpha^0)\Delta v^B + (\bar{\Pi}_0 - \bar{\pi}_0)\Delta v_0 + (\bar{\Pi}_1 - \bar{\pi}_1)\Delta v_1 \geq c & (IC_1) \\ \alpha^1 v_0^B + (1-\alpha^1)v_1^B + \bar{\Pi}_0 \Delta v_0 + \bar{\Pi}_1 \Delta v_1 - c \geq \bar{v} & (IR) \\ v_0^B + \Pi\left(a_0, \bar{s}\right)\Delta v_0 = v_1^B + \Pi\left(a_1, \bar{s}\right)\Delta v_1 & (CT) \end{cases}$$

where the last constraint is the first order condition of the change of treatment at $\bar{s}$. [27]

The next lemma shows that doctors choose no treatment for every signal $s \leq \bar{s}$ and choose treatment for an interval $[\bar{s}, \hat{s}]$. If $\hat{s} < 1$, then they choose no treatment for every signal $s \geq \hat{s}$. In particular, once they switch from treatment to no treatment, they never switch back to treatment.

Lemma 8 *In the cost of implementation program, the optimal contract induces just one change to treatment: no treatment below $\bar{s}$ and treatment from $\bar{s}$ up to some signal $\hat{s}$. If $\hat{s} < 1$, then doctors choose no treatment for every signal larger than $\hat{s}$.*

This lemma is proved in the appendix.

The next lemma shows the cost minimization problem has a solution and at the solution the (IR) restriction must be binding and thus doctors receive the reservation utility. Moreover, we also show that at the optimal contract doctors always receive more in the good state of nature than in the bad.

Lemma 9 *For each $\bar{s}$, there exist a solution for the program above and the (IR) constraint is binding at the optimal contract. Moreover, we must have $\Delta v_j \geq 0$ for $j = 0, 1$.* [28]

The lemma is proved in the appendix.

[27] The equality is substituted by $\geq$ $(\leq)$ when $\bar{s} = 0(1)$.

[28] Actually, a close investigation of the next lemma's proof shows that one must have $\Delta v_1 \geq \Delta v_0$ and $\Delta v^B \leq 0$.

Definition 10 *We say that a contract r is* ***fee-for-service*** *if payments are not contingent on the treatment chosen, $\Delta r^{\omega} = 0$ for every ω. We say that r is* ***capitation*** *if payments are contingent upon the treatment chosen, $\Delta r^{\omega} \neq 0$ for some ω. If doctors do not always choose the best medical treatment, we say that* ***incentives are provided to doctors to control costs****.*

The next lemma shows that the optimal contract with effort is capitation and it also describes when the optimal contract requires effort. We shall use the following notation. Let s^* be the critical signal where treatment becomes the best medical option, given that the doctor has chosen to make an effort. Therefore, the best medical option is to choose no treatment for every signal $s \leq s^*$ and to choose treatment thereafter (returning or not to no treatment again).

Lemma 11 *i) Suppose doctors are strictly risk-averse and the optimal contract is capitation. Then equilibrium is not second-best.*
ii) Suppose at the optimal contract doctors choose effort. The optimal contract is capitation and $\bar{s} > s^*$.
iii) If both $W_G - W_B$ and a are large enough and their ratio, $(W_G - W_B)/a$, is also high enough, then in the optimal contract doctors choose effort.

The lemma is proved in the appendix.

The fee-for-service contract corresponds to a contract with payment not contingent upon the treatment chosen. From the (CT) condition, given such a contract doctors will always choose the best medical treatment, irrespective of its cost. As we saw in the last section, if no effort is constrained optimal, then the optimal contract is fee-for-service and the equilibrium outcome is actually first-best.

Suppose, however, it is optimal to make effort. From the previous lemma and once more from (CT), the optimal contract will specify a critical signal $\bar{s}$ such that: doctors choose no treatment before $\bar{s}$ and treatment for an interval after that. Given that $\bar{s} > s^*$, there is an interval of signals such that doctors do not choose the best medical treatment. However, since equilibrium in our model is always constrained optimal, the welfare associated with this capitation contract is larger than that associated with a fee-for-service contract.

These results can be summarized in the following main proposition of the paper.

Proposition 12 *If no effort is constrained optimal, then the optimal contract is fee-for-service and the equilibrium outcome is first-best.*
If both the cost of treatment, a, and its benefit, $(W_G - W_B)$, are high enough and benefit over cost, $(W_G - W_B)/a$, is also high enough, then effort is constrained optimal. Moreover, in this case:
i) the optimal contract is partial capitation;
ii) doctors are provided incentives to control costs and, hence, do not always choose the best medical treatment;
iii) patient welfare is strictly larger under capitation than under fee-for-service.

In our model partial capitation (e.g., risk-sharing contract) arises when effort is constrained optimal. One should notice that the conditions specified in the proposition seems to fit the stylized facts from the health insurance market: the increased market share of capitation contracts is associated with the appearance of new medical technologies that are simultaneously more expensive and more efficient in the treatment of several health conditions, which is precisely a sufficient condition in our model for capitation contracts to be optimal, provided that the benefit over cost ratio is also high enough. In this case, at the optimal contract doctors are provided incentives to save medical costs and do not always choose the best medical treatment available for every signal. In particular, for some signals doctors choose no treatment when treatment provides a better likelihood of the patient achieving the good state of nature. Besides this, patient welfare is higher under capitation than under a fee-for-service contract. Therefore, even though the best medical treatment is not always chosen under capitation, patients, in this case choose this contract instead of the standard fee-for-service contract.

5 Concluding remarks

The paper proposed a contract model to analyze managed care and, specifically, capitation contracts. The model's basic feature is the existence of an asymmetry of information between insurance companies and patients, on one side, and the providers of service, doctors, on the other. Doctors know more about patients' health than insurance companies or even patients do. Moreover, doctors success in treating patients may depend on a non-observable effort level. In order to induce doctors to make an effort, optimal contracts have to provide larger payoffs in case patients achieve the good state of nature. However, this induces doctors to disregard treatment costs, leading to high insurance premiums and reducing patients' expected welfare. Therefore, in equilibrium, it may be optimal to provide incentives to doctors to save medical costs through capitation contracts.

We have shown that fee-for-service contracts might be first-best. This happens when the gains from providing incentives for doctors to choose effort are not high enough. However, in many cases the first-best is not attainable and, in particular, capitation contracts may provide larger patient expected utility than fee-for-service contracts. Capitation contracts arise precisely when both the benefits and costs from treatment are high enough: in this case, patients benefit from high effort but also from doctors not always using the best medical treatment. This equilibrium property seems to fit the stylized facts from the growth of managed care and capitation contracts in the US.

Since capitation contracts are not first best, the transition from fee-for-service to capitation contracts, due for example, to a change in the treatment technology, may generate a sense of loss in efficiency. However, every time the optimal contract is capitation, it welfare dominates a fee-for-service contract. Even though they are not first-best, capitation contracts, whenever they arise, are the best possible outcome, maximizing patient welfare subject to the given information asymmetry.

6 Appendix

6.1 Proof of lemma 6

The proof follows immediately from the concavity of v. Consider the contract:

$$\begin{aligned} r_j &= \left(1 - \frac{\pi_j^i}{\alpha^i}\right) r_j^B + \frac{\pi_j^i}{\alpha^i} r_j^G \quad j = 0, 1 \\ r &= \alpha^i r_0 + (1 - \alpha^i) r_1 \end{aligned}$$

We have:

$$\begin{aligned} L^i(r, d) &= G^i d_G - (1 - G^i) d_B - [r + (1 - \alpha^i) a] \quad \text{and} \\ V^i &\leq v(r) - c_i \end{aligned}$$

We showed that every contract is dominated by a constant one, since the (IC_0) is satisfied for this contract. Thus, $(r_j^w = \bar{r})$ is the least expensive contract. $\square$

6.2 Proof of lemma 8

Consider the function $T : \left[\Pi_0^{-1}(1), \Pi_0^{-1}(0)\right] \to [0,1]$, $T(p) = \Pi_1(\Pi_0^{-1}(p))$, which gives the trade-off between the probability of the good state of nature given treatment and no treatment. Observe that by $(H2)$ and (ii) of $(H3)$, this function is increasing and concave.

Remark 13 *Restriction (CT) in the cost minimization problem*

$$\Delta v^B - \Pi(a_0, \bar{s})\, \Delta v_0 + \Pi(a_1, \bar{s})\, \Delta v_1 \geqslant 0$$

is satisfied if and only if

$$(\Pi(a_0, \bar{s}), \Pi(a_1, \bar{s})) \in H^+ := \{(x, y) \in \Re^2; \Delta v^B - \Delta v_0 x + \Delta v_1 y \geqslant 0\}$$

Since the graph of T is an increasing concave curve, the hyperplane H crosses the graph of T twice at most.

Suppose that at the optimal contract the hyperplane H crosses the graph of T at another signal $\widehat{s} \neq \bar{s}$. There are two cases to consider:

(i) $\bar{s} > \widehat{s}$, and treatment is chosen on $[0, \widehat{s}] \cup [\bar{s}, 1]$ and no treatment on $[\widehat{s}, \bar{s}]$.

(ii) $\bar{s} > \widehat{s}$, and no treatment is chosen on $[0, \widehat{s}] \cup [\bar{s}, 1]$ and treatment on $[\widehat{s}, \bar{s}]$.

We claim that case *(i)* cannot happen.

Since H crosses the graph of T at $\widehat{s}$ and $\bar{s}$,

$$\frac{\Delta v_0}{\Delta v_1} = \frac{\Pi(a_1, \bar{s}) - \Pi(a_1, \widehat{s})}{\Pi(a_0, \bar{s}) - \Pi(a_0, \widehat{s})} > 0$$

It is simple to verify that one cannot have $\Delta v_j < 0$ for both j's. Thus, $\Delta v_j \geq 0$ for all j[29]. By the second order condition of change of treatment (see the inequality after definition 1),

$$-d\Pi(a_0, \widehat{s})\Delta v_0 + d\Pi(a_1, \widehat{s})\Delta v_1 \leq 0$$

and thus

$$\frac{d\Pi(a_0, \widehat{s})}{d\Pi(a_1, \widehat{s})} \leq \frac{\Delta v_1}{\Delta v_0}$$

and by (ii) of $(H3)$ the function on the left-hand side is non-increasing. If doctors strictly prefer treatment for some signal s, we must have

$$-d\Pi(a_0, s)\Delta v_0 + d\Pi(a_1, s)\Delta v_1 > 0$$

[29] See the IC_1 constraint in the proof of lemma 9.

which implies

$$\frac{d\Pi(a_0,s)}{d\Pi(a_1,s)} > \frac{\Delta v_1}{\Delta v_0}.$$

But this is impossible for any $s \geq \hat{s}$, which is the desired contradiction. $\square$

6.3 Proof of lemma 9

By (i) of $(H3)$, $\alpha^1 - \alpha^0 = (1-\rho)\alpha^1$ and $\bar{\Pi}_j - \bar{\pi}_j = (1-\rho)\bar{\Pi}_j$ $, j = 0, 1$. Thus, using (CT), (IR) and (IC_1) are equivalent to

$$\begin{aligned} v_1^B + [\bar{\Pi}_0 - \alpha^1 \Pi(a_0, \bar{s})]\Delta v_0 + [\bar{\Pi}_1 + \alpha^1 \Pi(a_1, \bar{s})]\Delta v_1 - c - \bar{v} &\geq 0 \\ [\bar{\Pi}_0 - \alpha^1 \Pi(a_0, \bar{s})]\Delta v_0 + [\bar{\Pi}_1 + \alpha^1 \Pi(a_1, \bar{s})]\Delta v_1 - \frac{c}{1-\rho} &\geq 0 \end{aligned}$$

Observe that we can easily construct a contract that satisfies (CT). We can also find Δv^B such that (IC_1) is true and (CT) is preserved. Finally, we can choose v_1^B to satisfy the (IR). It is also easy to see that (IR) constraint is binding with the optimal contract.

Suppose the problem has no solution. Then, there is a sequence r_n such that $c(v(r_n))$ is a strictly decreasing function. If r_n has a convergent subsequence with limit $r \gg 0$, then it is easy to verify that r would solve the minimization problem, which contradicts the assumption. Thus, taking a sub-sequence if necessary, there are j and ω such that $r_j^\omega \to 0$. By (IR), there must exist ω^* and j^* such that $r_{j^*}^{\omega^*} \to \infty$. But then $c(v(r_n)) \to \infty$, which is the desired contradiction.

By (IC_1), Δv_0 and Δv_1 cannot be negative at the same time. Moreover, if they have opposite signs, the derivative of the (CT) condition

$$-d\Pi(a_0, \bar{s})\,\Delta v_0 + d\Pi(a_1, \bar{s})\,\Delta v_1 \geq 0$$

implies that $\Delta v_0 \geq 0$. We now prove that Δv_1 is also non-negative. Suppose this is not the case. Then, by the first order condition associated with the choice of treatment, doctors change from no treatment to treatment at a single signal. Moreover, under (IR), (IC_1) is equivalent to

$$v_1^B \leq \bar{v} - \tfrac{c\rho}{1-\rho}$$

Thus, the cost of implementing program $c^1(\bar{s})$ with just one crossing is:

$$\begin{aligned} &\min \alpha^1 h\left(v_0^B\right) + (1-\alpha^1)\left[h\left(v_1^B\right) + a\right] + \bar{\Pi}_0\left[h\left(v_0^B + \Delta v_0\right) - h\left(v_0^B\right)\right] \\ &+ \bar{\Pi}_1\left[h\left(v_1^B + \Delta v_1\right) - h\left(v_1^B\right)\right] \end{aligned}$$

$$\text{s.t.} \begin{cases} v_1^B + [\bar{\Pi}_0 - \alpha^1 \Pi(a_0, \bar{s})]\Delta v_0 + [\bar{\Pi}_1 + \alpha^1 \Pi(a_1, \bar{s})]\Delta v_1 - c - \bar{v} \geq 0 & (IR) \\ \bar{v} - \dfrac{c\rho}{1-\rho} - v_1^B \geq 0 & (IC_1) \\ \Delta v^B - \Pi(a_0, \bar{s})\Delta v_0 + \Pi(a_1, \bar{s})\Delta v_1 = 0 & (CT) \end{cases}$$

Taking the Lagrangian derivative with respect to v_0^B, v_1^B, Δv_0 and Δv_1, we get respectively

$$\begin{aligned} \alpha^1 h'(v_0^B) + \bar{\Pi}_0(h'(v_0^G) - h'(v_0^B)) + \lambda_3 &= 0 \quad (1) \\ (1-\alpha^1)h'(v_1^B) + \bar{\Pi}_1(h'(v_1^G) - h'(v_1^B)) - \lambda_1 + \lambda_2 - \lambda_3 &= 0 \quad (2) \\ \bar{\Pi}_0 h'(v_0^G) - \lambda_1[\bar{\Pi}_0 - \alpha^1 \Pi(a_0, \bar{s})] + \lambda_3 \Pi(a_0, \bar{s}) &= 0 \quad (3) \\ \bar{\Pi}_1 h'(v_1^G) - \lambda_1[\bar{\Pi}_1 + \alpha^1 \Pi(a_1, \bar{s})] - \lambda_3 \Pi(a_1, \bar{s}) &= 0 \quad (4) \end{aligned}$$

where $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and λ_3 are the Lagrange multipliers of $(IR), (IC_1)$, and (CT), respectively. The equations (3) and (4) are equivalent to

$$\begin{aligned} h'(v_0^G) &= \lambda_1(1 - \alpha^1 \frac{\Pi(a_0, \bar{s})}{\bar{\Pi}_0}) - \lambda_3 \frac{\Pi(a_0, \bar{s})}{\bar{\Pi}_0} \\ h'(v_1^G) &= \lambda_1(1 + \alpha^1 \frac{\Pi(a_1, \bar{s})}{\bar{\Pi}_1}) + \lambda_3 \frac{\Pi(a_1, \bar{s})}{\bar{\Pi}_1} \end{aligned}$$

Thus,

$$(\lambda_1 - h'(v_0^G))\frac{\bar{\Pi}_0}{\Pi(a_0, \bar{s})} = \alpha^1 \lambda_1 + \lambda_3 = (h'(v_1^G) - \lambda_1)\frac{\bar{\Pi}_1}{\Pi(a_1, \bar{s})},$$

which implies that

$$h'(v_0^G) > \lambda_1 > h'(v_1^G) \text{ or } h'(v_1^G) > \lambda_1 > h'(v_0^G)$$

when $h'(v_0^G) \neq h'(v_1^G)$. Moreover,

$$v_0^G > v_1^G \Leftrightarrow h'(v_0^G) > h'(v_1^G) \Leftrightarrow \alpha^1 \lambda_1 + \lambda_3 < 0.$$

From equations (1) and (3) above we get, after a few manipulations,

$$(\alpha^1 - \bar{\Pi}_0)h'(v_0^G) = (\alpha^1 - \bar{\Pi}_0)h'(v_0^B) - (\alpha^1 \lambda_1 + \lambda_3)\left(\alpha^1 \frac{\Pi(a_0, \bar{s})}{\bar{\Pi}_0} - 1\right)$$

Suppose $v_0^G > v_1^G$. Since $\alpha^1 - \bar{\Pi}_0 > 0$, $\alpha^1 \lambda_1 + \lambda_3 < 0$ and $\alpha^1 \Pi(a_0, \bar{s}) - \bar{\Pi}_0 < 0$, we get $h'(v_0^B) > h'(v_0^G)$ and so $v_0^B > v_0^G$. But this contradicts $\Delta v_0 \geq 0$. Therefore, $v_0^G \leq v_1^G$.

Notice that (CT) is equivalent to the equality of a convex combination of $\{v_0^B, v_0^G\}$ and of $\{v_1^B, v_1^G\}$, which implies that the intervals $[v_0^B, v_0^G]$ and $[v_1^B, v_1^G]$ must overlap. Thus, $v_1^B < v_0^G \leq v_1^G$ or, in particular, $\Delta v_1 > 0$. The proof is then complete. $\square$

6.4 Proof of lemma 11

(i) The fact that any equilibrium where doctors choose effort is not first best follows from the standard moral hazard model. In that model, since signals are observable, $\Delta v^\omega = 0$. Under capitation, however, we must have $\Delta v^\omega \neq 0$ for some ω. The result then follows from doctors' risk aversion.

(ii) We have to show that if with the optimal contract doctors choose effort, then $\Delta v^\omega \neq 0$ for some ω and in this case $\bar{s} > s^*$.

If this is not true, then $v_1^G = v_0^G =: v^G$, $v_1^B = v_0^B =: v^B$ and $\Delta v_1 = \Delta v_0 =: \Delta v$. We also use a similar notation for the contract r. From (CT) we have

$$(\Pi(a_0, \bar{s}) - \Pi(a_1, \bar{s}))\, \Delta v = 0$$

and thus either $\Delta v = 0$, in which case doctors choose no effort by the previous section, or

$$\Pi(a_0, \bar{s}) = \Pi(a_1, \bar{s}) \Rightarrow \bar{s} = s^*.$$

By (CT), in this case there is just one change from no treatment to treatment at $s = s^*$. From the proof of lemma 9, at $\bar{s} = s^*$, the (CT) condition implies that $v_1^B < v_0^B$. Using the envelope theorem we get the following derivative of c^1 :

$$\begin{gathered} h(v_0^B) - (h(v_1^B) + a) + \Pi(a_0,\bar{s}) \left[h\left(v_0^B + \Delta v_0\right) - h\left(v_0^B\right)\right] \\ -\Pi(a_1,\bar{s}) \left[h\left(v_1^B + \Delta v_1\right) - h\left(v_1^B\right)\right] - (\lambda_1 \alpha^1 + \lambda_2)(-d\Pi(a_0,\bar{s})\,\Delta v_0 + d\Pi(a_1,\bar{s})\,\Delta v_1 < 0 \end{gathered}$$

since $\lambda_1 \alpha^1 + \lambda_2 > 0$, and the derivative of (CT), $-d\Pi(a_0,\bar{s})\,\Delta v_0 + d\Pi(a_1,\bar{s})\,\Delta v_1$, is positive.

Since the derivative of social benefit at s^* is zero, s^* can not be the unique second-best change of treatment. Moreover, we must have $\Delta v^\omega \neq 0$ for some ω. Therefore, if effort is constrained optimal, capitation is the optimal contract and $\bar{s} \neq s^*$.

We claim that $v_1^B < v_0^B$, for $\overline{s} \neq s^*$. We have to consider two cases:

(1) $\overline{s} > s^*$. We have $\Pi(a_1, \bar{s}) > \Pi(a_0, \bar{s})$ and consequently

$$\Pi(a_1, \bar{s})\, v_1^G + (1 - \Pi(a_1, \bar{s})) v_1^B > \Pi(a_0, \bar{s})\, v_0^G + (1 - \Pi(a_0, \bar{s})) v_0^B \quad (*),$$

which is absurd by (CT).

(2) $\overline{s} < s^*$. We have $\Pi(a_1,\bar{s}) \leq \Pi(a_0,\bar{s})$ and consequently the lottery $\{r_1^B, r_1^G, 1-\Pi(a_1,\bar{s}), \Pi(a_1,\bar{s})\}$ strictly stochastically dominates (in the first order sense) the lottery $\{r_0^B, r_0^G, 1-\Pi(a_0,\bar{s}), \Pi(a_0,\bar{s})\}$. This implies the inequality $(*)$ again.

Therefore, we have that the derivative of c^1 is negative. Since the derivative of benefits is negative if and only if $\overline{s} > s^*$, at the optimal contract we must have $\overline{s} > s^*$. [30]

(iii) Observe that, for the contract defined in (i), the (IC_1) and (IR) constraints give:

$$\Delta v = \frac{c}{(1-\rho)(\bar{\Pi}_0^* + \bar{\Pi}_1^*)} \text{ and } v^B \leq \bar{v} - \frac{c\rho}{1-\rho}.$$

From proposition 7, the fee-for-service contract has an expected cost given by:

$$c^f = \bar{r} + \left(1 - s^f\right) a$$

where $s^f = s^0$ is the optimal signal under a fee-for-service contract, and its benefits are given by

$$b^f = W_B + (W_G - W_B)\left(\bar{\pi}_0^f + \bar{\pi}_1^f\right)$$

where $\pi\left(a_1, s^f\right) - \pi\left(a_0, s^f\right) = a/\left(W_G - W_B\right)$. Therefore, s^f is increasing in $a/\left(W_G - W_B\right)$.

The cost of the contract defined in (i) is:

$$c^c = r^B + (1 - s^*)a + (\bar{\Pi}_0^* + \bar{\Pi}_1^*)\Delta r$$

and benefits, on the other hand, are given by

$$b^c = W_B + (W_G - W_B)\left(\bar{\Pi}_0^* + \bar{\Pi}_1^*\right)$$

Thus, capitation will be implemented if $b^f - c^f < b^c - c^c$, i.e.,

$$(W_G - W_B)\left(\bar{\Pi}_0^* - \bar{\pi}_0^f + \bar{\Pi}_1^* - \bar{\pi}_1^f\right) > r^B - \bar{r} + \left(s^f - s^*\right)a + \left(\bar{\Pi}_0^* + \bar{\Pi}_1^*\right)\Delta r.$$

Therefore, $s^f < s^*$ if $a/\left(W_G - W_B\right)$ is small enough and the strict inequality holds provided that a and $(W_G - W_B)$ are large enough relative to Δr, and $r^B - \bar{r}$ since $\left(\bar{\Pi}_0^* - \bar{\pi}_0^f + \bar{\Pi}_1^* - \bar{\pi}_1^f\right) \backsimeq (1-\rho)\left(\bar{\Pi}_0^* + \bar{\Pi}_1^*\right) > 0$. □

[30] Observe that in this case there might be two changes of signals and the Lagrangian that was presented above is just for one change of signal. However, it is immediately verifiable that the inclusion of the other (CT) does not change the result.

7 References

Arrow, K. J. (1963): "Uncertainty and the welfare economics of medical care"; *American Economic Review*; **53:941-973.**

Brown, L. D. (1983): "Exceptionalism as the rule? US health policy innovation and cross-national learning"; *Journal of Health Politics, Policy, and law;* **23:35-51.**

Ethhoven, A. (1993): "The history and principles of managed competition"; in: *Health Affairs*, Supplement: **24-48.**

Glied, S. (1999): "Managed Care"; NBER working paper #7205.

Health Care Financing Administration (1998): "Brief Summaries of Medicare and Medicaid" and "National Health Expenditures". Available: http://www.hcfa.gov/Medicare (accessed August 1998).

Luft, H. S. (1981): *Health Maintenance Organizations: dimensions of performance*; John Wiley and Sons, New York.

Mc Clelan, M. and D. Kessler (1996): "Do Doctors Practice Defensive Medicine?" *Quarterly Journal of Economics*; Cambridge; **CXI:353-390**.

Miller, R. H. and H. S. Luft (1997): "Does Managed care lead to better or worse quality of care?"; *Health Affairs*; **16:7-25.**

Newcomer, R.; S. Preston. and C. Harrington (1996): "Health Plan Satisfaction and Risk of Disenrollment Among Social/HMO and Fee-for- Service Recipients"; *Inquiry : a journal of medical care organization, provision and financing*; **33:144-154**.

Newhouse, J. (1996): "Reimbursing health plans and health providers: efficiency in production versus selection"; *Journal of Economic Literature*; **7:285-288.**

Newhouse, J., W. Schwartz, A. Williams and C. Witsberger (1985): "Are fee-for-service costs increasing faster than HMO costs?"; *Medical Care;* **23:960-66.**

Pauly, M. (1986): "Taxation, Health Insurance, and Market Failure in Meidcal Care"; *Journal of Economic Literature*; **24(3):629-75.**

Rothschild, M. and J. Stiglitz (1976): "Equilibrium in competitive insurance markets"; *Quarterly Journal of Economics*; **90:629-650**.

Zweifel, P. and F. Breyer (1997): *Health Economics;* Oxford University Press.