CEFOS REPORT 10

Quality and Efficiency in Health Care Production

Mattias Lundbäck

ISRN GU-CEFO-R--10--SE
ISSN 1104-327x

Contents

– PART ONE –

Incentives for Quality Maintenance and Efficiency Enhancement in Health Care Production

– *PART TWO* –

Imperfect Agency and the Regulation of Hospitals[1]

[1] An earlier version of this paper is also published in The Geneva Papers on Risk and Insurance Theory (Lundbäck, 1997).

– *PART THREE* –

Non-linear Incentives in Not-for-profit Hospitals

Foreword

In the spring of 1996, the project EUS (Economic Evaluation of Schizophrenia Treatment) was initiated by Professor Lars Söderström at the Department of Economics at Lund University. The project concerns the evaluation of psychiatric care and the study of cost and quality issues in the provision of health care. The papers in this report provides a theoretical background for the analysis of this topic.

This report is number ten in the series of CEFOS reports. The report *Quality and Efficiency in Health Care Production* is also a licenciate dissertation at the Department of Economics, Göteborg University.

Göteborg, June 1998

Professor
Lars Strömberg
Head of the Center for Public Sector Research

Preface

While doing the research that has resulted in this dissertation, I have had great help from many people. First and foremost, of course, my supervisor Lars Söderström with whom I innumerable times during the last two years discussed, and received comments about, this thesis. I would also like to thank him for providing me financial and moral support for this work.

I would like to thank CEFOS and Nationalekonomiska Institutionen at Göteborg University for providing me financial support. I would also like to thank all people working at CEFOS for comments on my work. I also want to say that I have enjoyed the friendly atmosphere at CEFOS very much.

I have benefited greatly from comments about the papers in this thesis from a large number of people. I would especially like to mention Klas Bergenheim, Fredrik Carlsson, Dinky Daruvala, Thomas Ericson, Bengt J. Eriksson, Bengt Haraldsson, Per-Johan Horgby, Karin Lönnerstedt, Anna Persson, Klas Rikner and Knut Sydsäter.

I am grateful for comments from participants at seminars held at Lund University, University of Karlstad, CEFOS at Göteborg University and Göteborg School of Economics and Commercial Law.

In the academic world outside Sweden, I would like to thank Prof. Harris Schlesinger, Prof. Christian Gollier, Prof. Michael Rothschild, Prof. Richard D. MacMinn and Prof. Johann-Matthias Graf von der Schulenburg for accepting my paper for the 23rd Seminar of the European Group of Risk and Insurance Economists. I would also like to thank the participants of the conference, among those my commentator Prof. Mark J. Browne and also Prof. Hugh Gravelle for comments on my paper.

I would also like to thank the editors of The Geneva Papers on Risk and Insurance Theory, Prof. Harris Schlesinger, Prof. Christian Gollier and two anonymous referees for valuable comments on my paper (Part two of this thesis).

Last, but not least, I would like to thank my parents and the rest of my family for supporting me in my work with this thesis and my academic work here at CEFOS and Göteborg School of Economics and Commercial Law.

– Part One –

Incentives for Quality Maintenance and Efficiency Enhancement in Health Care Production

1 Introduction

> "A physician's character is injured when we endeavour to pursuade the world he kills his patients instead of curing them, for by such a report he loses his business." (Adam Smith)

1.1 Background and purpose

Since the introduction of fixed, predetermined, payments for hospitals treating patients in the goverment financed Medicare reimbursement system in the United States, a recurring question has been whether these prospective payments will cause a deterioration of quality in health care production. Setting payments in advance will give doctors, and/or hospitals, incentives to reduce cost by reducing the quality of treatments. Quality is defined as expenses that increases the health (and consequently also the utility) of patients. Another term often used to describe the same variable is *treatment intensity*. Often, the definition of quality is scaled so as to make it identical to "the cost of increasing the quality of treatments".

Quality has been of concern in the study of less regulated markets than the hospital industry. One example is Akerlof´s (1970) famous article about the market for lemons.[2] The dichotomy between payment for hospital services (often made by insurance companies or the state) and the consumption, makes quality problems more acute in the

[2] Akerlof's article concerns the fact that it is difficult to observe the quality of cars in the used car market. The concept "lemon" is used to denote a car that turns out to need a lot of repairs after it is bought.

hospital industry than in most other markets. The extreme informational advantage of the physician vis-à-vis the patient exaggerates this problem further. Arrow (1963, p. 966) notes that licensing and educational standards are a way of obviating the problem of informational uncertainty. These methods are designed to reduce the uncertainty in the mind of the consumer as to the quality of the product.

One obvious remedy to the problem of how to maintain quality in health care production would be to increase available resources until quality reaches the desired level. However, a problem with this approach is that the payers (taxpayers or insurance companies) have insufficient means to make sure that an increase in resources to the health care sector automatically generates higher quality. The result may well be that the production processes gets less efficient and that managerial perks, instead of higher quality in production, consume resources.

Another problem is that the paying part (insurance company, local government, etc.) might not know how efficient the hospital could really become. This fact causes problems when the regulator wants to reimburse the hospital for its incurred costs. Hospitals with a high efficiency level might try to masquerade as low efficiency ones to get more money. This could be done by investing resources in amenities that mainly benefit the employees at the hospital or by refraining from undertaking cost reduction efforts. This fact makes it necessary to adapt incentive schemes that promotes quality, but at the same time limits slack in the organisation. Quality and efficiency are thus the central concerns when deciding governance systems for health care production and thus also the concern of this thesis.

1.2 Outline of the thesis

In Part one, I will present an overview of the theory about regulation of health care production, with special emphasis on hospitals. In Part two, a model of hospital regulation in an environment of hidden information, hidden action and imperfect agency, is presented. Part three concerns the difference between regulating for-profit hospitals and not-for-profit hospitals. This is an important distinction as the incentives for managers in the two organisational forms are likely to be quite different.

Part one begins with an overview of problems in the production and financing of health care (section 2). This is followed by a discussion of how reimbursements are done in reality, starting from an overview of reimbursement systems used for hospitals in different environments (3.1). The implementation of prospective payments is discussed in section (3.2). This is followed by a discussion about the contrariety between different types of health care production and the performance of different types of reimbursement systems in these variable production settings (3.3).

The discussion about actual reimbursements is followed by an introduction to the principal agent theory (4.1). The following section concerns asymmetric information and its consequences in principal-agent relations (4.2). This is ensued by a discussion about yardstick competition and how "hidden information" and "hidden action"[3] complicates the task of regulating hospitals (4.3). A general discussion about how the models of section 4.3 could be combined concludes the

[3] "Hidden information" is often used as a synonym for the more common term "adverse selection" and "hidden action" is similarly used as a synonym for "moral hazard". I prefer to use the terms "hidden action" and "hidden information" because they are more intuitive and also because there often is a confusion between the use of these concepts in the insurance literature and in the regulation literature.

chapter. The outlines of Part two and Part three are given in the introduction to their respective parts.

1.3 Methodology

The study is firmly based on traditional neo-classical approach; rational choice, stable preferences and equilibrium. Since the study deals with asymmetric information and non contractible actions, some of the traditional neo-classical assumptions are not always assumed to hold. These deviations are explicitly defined in the model.

My concern in this report is not to illustrate all different ways to organise production of health care, but to show how economic incentives affect quality and efficiency. To make this task manageable, it is necessary to limit the discussion somewhat. The focus is on the economic incentives for the individual hospital, and the distinction between prospective reimbursements and the traditional cost reimbursement method will be stressed. Cost reimbursement implies that the regulator reimburses the hospital for its incurred costs when treating a patient. Prospective payment methods are based on different ways to measure productivity and make reimbursements on the basis of these productivity measures. I wish to distinguish between these two main categories in the analysis. Aspects about the demand side will to a large extent be ignored.

2 Why regulation of health care?

2.1 General background

The main reason for government intervention in the market for goods and services is usually assumed to be the existence of market failures. More specifically, market failure refers to a situation in which a market equilibrium does not attain a Pareto optimal allocation of resources. The point of departure for the theory of market failure is the first theorem of welfare economics (Arrow, 1951, Debreu, 1959). This proposition states that in the absence of external effects and with free information, every competitive equilibrium[4] constitutes a Pareto optimum.

There are several reasons why the first welfare theorem does not apply to the market for health care. Some of these are:

- Imperfect competition
- External effects
- Merit wants
- Imperfect information

The fact that monopolies cause welfare loss in the economy is a well known proposition in economic theory. The same applies if consumption of a good increase/decrease the utility of other individuals in society. The individuals should be made to take these effects into consideration when deciding on their consumption of health care. If

[4] The notion of a competitive equilibrium also presumes that there are many agents on both the demand and supply side, and that they each are of small size relative to the overall market size.

there is such a positive externality, the consumption of the individual would, from a social point of view, be too low. A public subsidy to the consumption of health care would thus be motivated (Vaccination against contagious diseases could be one example).

The *merit want* argument (Musgrave, 1959) is based on the idea that the individual's own valuation of his/her utility from a good should not be accepted for some reason. When consumption of some goods have been prohibited by the state, one example is illegal drugs, this reasoning seems to be at the bottom-line.

The fourth reason, which will be dealt with most thoroughly in this thesis, is that there exists imperfect information. The lack of perfect information is the reason for existence of principal-agent problems, as well as many other issues treated in regulation and insurance theory. Asymmetric information makes contract terms hard to verify and enforce. Lack of contractability is thus related to the existence of asymmetric information. Coase (1960) used the general term "transaction costs" to describe the problems that evolved in the process of constructing, monitoring and enforcing the terms of a contract. Under perfect competition, and with no transaction costs, private and social costs will always be equal, as pointed out by Stigler (1966). The modern use of asymmetric information in formalised economic models begins with Mirrlees (1974, 1976) and Stiglitz (1974, 1975) among others.

2.2 Imperfect information

The nature of demand for medical services is highly irregular and unpredictable. Kenneth Arrow noted:

> "It is hard, indeed, to think of another commodity of significance in the average budget of which this is true. A

> portion of legal services, devoted to defence in criminal trials or to lawsuits, might fall in this category but the incidence is surely much lower (and, of course, there are, in fact strong institutional similarities between the legal and medical-care markets)." (Arrow, 1963, p. 948)

The demand for health care is thus based on the consumers "state contingent" utility. Health care is a state contingent good (especially in the case of injuries and diseases, routine health controls are different). Specifying a contract to determine consumption and create a buffer to finance health care in every possible contingency is an impossible task. Another important aspect is that there is a large uncertainty about the product. Recovery from a disease is often as unpredictable as its incidence. The large degree of unpredictability about health care costs makes it necessary to find methods to finance health care that puts the patients in a more favourable position, i. e. less exposed to risks.[5]

Implementing health insurance, however, raises new and difficult problems. Two of these problems are moral hazard and adverse selection. Moral hazard, or "hidden action", means that the patients who does not pay the full cost of their health care themselves, might have incentives to consume more of the good (health care) in each possible contingency, than if it had been possible to specify the means and measures that should be taken in case of an illness, ex ante (i. e. the consumer, being fully informed, has too limited incentives to avoid being in need of health care).

Adverse selection, or "hidden information", bring about that only patients who are especially likely to be ill would be interested in buying

---

[5] In the world of Arrow-Debreu, health is a contingent good and risk-averse consumers sign contracts that insure against financial risk in states of the world where they need expensive medical treatment.

health insurance if premiums are valued at average risk. Because a high proportion of very ill patients buy insurance, the relatively healthy people will have to pay a too high insurance premium and may thus be reluctant to buy insurance at all. This may result in a complete market failure for private health insurance.

Nevertheless, it is common that health care is financed by some kind of insurer, or other "third party intermediator", like the state or local authorities. This "roundabout" method of financing health care has implied that the control of producers has been limited from the payers perspective. The producer and the patient may thus form a kind of implicit collusion. This implicit collusion will be in the interest of the patient as well as the provider. The provider will see a greater demand for health care and the patient will get more services from the provider.

These concerns is the background to the issue of hospital regulation. Regulation may thus be seen as a second best solution to a problem originally created in the relation between three groups; patients, providers and payers.

3 Hospital regulation in practice

3.1 Reimbursement for health care providers

The most common way to reimburse hospitals in the United States was, until recently, the fee-for-service system. *Fee-for-service* essentially means the physicians and hospitals send a bill of their treatment costs, concerning a patient, to the insurer. This system does obviously not stimulate cost consciousness on behalf of the providers. Modifications of the fee-for-service system have implied the use of *fixed fees* per visit or according to the type of procedure performed. *Capitation* is another reimbursement method commonly used. The hospital is given a fixed sum per associated patient and year, regardless if the patient uses services of the provider. *Global budgets* are a fixed sum to a hospital, regardless of the number of patients treated. In recent years, *case based,* or *prospective payments,* have also become common in many countries. The prime method to measure productivity in hospitals is to use *DRGs*[6] *(diagnosis related groups).* This productivity measure is then used to form a basis for reimbursement by prospective payment. DRGs are not the only system for categorisation used as a basis for prospective payments. The systems that exist differ by complexity. There exist simple systems based on the average cost per admission. There are intermediaries, like

[6] Diagnosis related groups (DRGs) were originally created as a way of tracking costs within hospitals and allocating resources. DRGs were thus not originally introduced as a way of making it possible to reimburse hospitals prospectively. Each patient is classified in a DRG group at the time of admission. The classification mainly depends of what diagnosis the patient gets at admission. This DRG classification can then be used to compare the services performed by different hospitals and ultimately also to determine appropriate reimbursements for a hospital.

the 16 group reimbursement system in Zaire (Shepard et al., 1990) and more complex ones, like the DRG based system, with 485 categories, used in the Medicare[7] system in USA (ProPAC, 1992), the Brazil 266 group system (World bank, 1993) or the 433 category system in Hungary. In Sweden, the DRG-system is often used for reimbursements to hospitals, with weights calculated for swedish average costs of treatment. Combinations of fee-for-service and case based payments also exist. Some HMOs[8] in the United States make capitated contracts with hospitals, but "carve out" some especially costly services and reimburse then on a fee-for-service basis (Barnum et al., 1995).

The lack of productivity measures for hospitals has probably been a factor that made the services by hospitals in many countries a matter for the state or local authorities. The reason is that it is hard to write contracts when performance cannot be measured precisely. Reimbursements to hospitals have often been made by using global budgets or, negotiated or ad-hoc, fee-for-service systems. Since fee-for-service in the past often was the only possible reimbursement system in a system based on competition, fee-for-service is more common where private or semi-private insurance companies are responsible for financing health care. Today, managed care and HMOs are replacing fee-for-service as the premium finance method. Since fee-for-service is common in the United States, the high proportion of health care

---

[7] A system for financing the health care for elderly in the United States. Medicare introduced prospective payments as a reimbursement mechanism on a large scale in 1983.

[8] HMOs are a type of prepaid medical plans which people can join. The HMOs have contracts with health care providers that specifies what should be done in case of an illness. The HMO is in that sense a kind of vertical integration between insurance companies and health care providers. The consumers choice is limited to providers that have a contract with the HMO. It can thus maintain larger control over the treatment procedures provided for patients and the cost of those.

expenditures, relative to GDP, there, has been ascribed to this reimbursement method.

Barnum et al. (1995) use the following table to show the effects of different types of reimbursement systems on some different aspects of health care delivery.

Table 1 Summary of incentives in pure reimbursement systems

Reimbursement Type	Underlying Incentives for: Cost/Unit	Services/Case	Quantity (of cases)	Risk Selection
Global budget	– –	– –	–	0
Fee for Service				
Unconstrained	–	+ +	+	0
Fixed	– –	+ +	+	+
Capitation	– –	– –	– –	+ +
Case Based	– –	– –	+ +	+

Legend: – – strong incentive to reduce; – moderate incentive to reduce; 0 no clear incentive; + moderate incentive to increase; + + strong incentive to increase.

Source: Barnum et al. (1995)

These different reimbursement methods can be combined in different ways and they are indeed used in many different combinations for reimbursing hospitals all over the world. There has been a global movement from fee-for-service and cost-reimbursement, to case-based payment systems, at least for hospitals. Capitation systems based on "managed care", e. g. HMOs, PPOs[9], etc., have also grown increasingly popular in the United States. This reimbursement method is based on a

[9] Preferred Provider Organisations. A kind of HMO.

vertical integration between insurance companies and health care providers.

3.2 Implementing prospective payments

Prospective payment is a fixed sum given to the hospital for the treatment of a patient, based on his/her diagnosis. Prospective payments are, however, often mitigated by different exemptions from the "fix-price" principle. One example is outlier rules often used for patients who are unusually costly or stay a long time at the hospital. These patients are often reimbursed by their costs, instead of their diagnosis. McClellan (1997) also show that many aspects of reimbursements in practice are not fully prospective, but that many factors that determine reimbursements are decided on by the physicians, i. e., through the choice between different treatment methods for a given diagnosis. McClellan asks why these retrospective elements are used in the PPS system and why they are not replaced by more sophisticated prospective payments, such as refined diagnosis-related classifications? The reason for this is most likely that the regulators want to avoid negative effects of the prospective payment mechanism on the quality of health care. This could be tendencies of hospitals to dismiss patients too early. There are in general two ways to solve this problem:

1] Using a reimbursement system based on capitation, like the HMOs in the USA and let the physicians alone make the decisions about resource use.

2] Further specialise DRGs to make it possible to separate efficiency in production of the intermediate products from their use by different patients.

The first route is also quite similar to the traditional method of financing hospitals in many contries, that is, global budgets decided on by politicians. Sweden is one example of the implementation of such a financing method. The second route has also been developed for fields where prospective payments traditionally have been regarded as difficult to implement, such as for psychiatry. These models are still mainly in the developing stage and have not yet been implemented on a large scale (Sharfstein, 1991, Frank & Lave, 1986, MH-CASC, 1995).

A reflection on this is that there might be a trade-off where the gains from efficiency enhancement by economic incentives, due to incentive payments, is outweighed by the adverse consequences for quality and loss of rent to the providers (see Part two of the thesis). This discussion is closely related to recent advances in the theory of the firm, specifically about the limits of the firm (Holmström, 1996), but beyond the scope of this thesis.

Prospective payments will only perform well if there is a clear division of responsibility between cost reduction and resource use, that is, a type of division between the production of intermediate products and the use of these products in the process of health production. The responsibility could then be divided between managers and physicians. Implicit here is the notion that every clinical cost will have some sort of 'intermediate product line'. For a laboratory or radiology cost centre, this is straightforward conceptually, since the intermediate product line consists of tests or procedures performed by the department. The same is true for departments, such as surgery or obstretics, where the product line consists of surgical procedures or normal births, and also hours of preoperative or postoperative nursing care. With departments such as psychiatry or medicine, things are a bit more problematical. Here, the patient is often being treated for chronic illness, and may be in the

hospital for a period of observation. As a contrast, the work that internists or psychiatrists do on the patients is not often easy to quantify in terms of 'procedures'. Fetter et al. (1991) suggest that the best measure to use in this case is just the number of consults or number of minutes or hours spent by physicians with patients, as well as number of hours or days of nursing care.

But if the accounting unit (the intermediate input) is just minutes, there will be small gains made by trying to make the input produced more efficiently. Since it is impossible to shorten a minute or an hour, the idea of responsibility division between managers and physicians will seem rather superfluous. The management level seems to loose its role as a co-ordinating factor, since the reimbursement could as well be made directly to the physicians, and thus only be based on the number of minutes of health care produced.

3.3 Two examples

Contrasting the analysis of surgical procedures, with the treatment of psychiatric illnesses, we see that the treatment procedures for psychiatric illnesses are less well defined and that the adverse consequences of too early dismissal is less clear than in the case of surgery. It is at least not as easy to prove the connection between adverse consequences and too early dismissal. These observations suggest that it is harder to infer the quality level than in surgery. The cost dispersion is very high for psychiatric patients, and the likelihood of systematic differences between different clinics and hospitals is larger in psychiatry. Psychiatric care is also often small-scale operations, and risk sensitivity can be high for care-givers. A prospective system thus puts caregivers at a large financial risk. These factors would theoretically

make a system of cost reimbursement, or global budgets, perform better than prospective payments in the treatment of psychiatric disease.

Surgery and psychiatry are two examples among the health care disciplines, that represent two extremes in hospital treatment and illustrate that one system for reimbursement is not optimal for all different types of health care. It is most likely the case that the principles for reimbursement in reality are adapted to this fact. The reimbursement methods used for psychiatry is generally different from the reimbursement methods for surgery. Surgery is generally thought to be characterised by fairly well-defined procedures, precise diagnoses and fairly homogeneous costs in a single diagnosis group.

While prospective payment is often used in surgery, it is less commonly used in psychiatrics. One reason is that it is difficult to make prognoses for the cost of psychiatric patients. Another reason is that it is hard to define the optimal length of treatment. The decision to dismiss a patient from a psychiatric unit is rather arbitrary, compared to the situation in surgery. In surgery, it is likely that a best practice concerning when to dismiss patients develops. The consequences of dismissing a newly operated patient too early can be severe and it is rather easy to measure the result (mortality rate for example) of the surgical procedures. In psychiatrics, too early dismissal of patients can be fatal, but it is hard to say whether a relapse into psychosis is a result of too early dismissal from the clinic or something that would have happened anyway. These factors make it harder to observe the quality level in psychiatric care, and hence, to apply prospective reimbursement without diminishing the quality of care. Three important elements make prospective payment difficult to apply in psychiatric care.

These are:

1] It is hard to observe the relation between inputs and outputs in psychiatric care.

2] Large differences in the treatment cost of different patients and a lack of methods to limit these differences by using diagnosis related groups or similar systems. This is critical, since most prospective payment systems rely on yardstick competition and try to compare the cost of different hospitals to set the prospective rate appropriately.

3] A large number of care-givers need to be co-ordinated, since modern psychiatric care often takes place within out-patient departments, with exception for severe cases of psychosis.

The follow-up process is extremely important, as well as the reduction of stress factors in the patients' environment. The treatment process is a "multi principal-agent" problem. The existence of multi principal problems make psychiatry even less suitable for prospective payment, since prospective reimbursements can induce co-ordination problems when a task is performed by several principals[10] (Laffont & Martimort, 1997, p. 214).

Ten different diagnosis groups (from about 500) in the DRG system concerns psychiatric illness. The cost variations within these groups are,

---

[10] An example might be where a physician is working as a case manager to perform two different complementary tasks, which are, achieving medical compliance and integrating the patient into society. The principals might then be the health care provider and the social service of the local community. If the marginal cost of integrating the patient into society decreases when the medical treatment is intensified, and vice versa, the two tasks are complementary. The situation introduces a free-rider effect in incentives, since the case manager can extract more rents from the principals if the two tasks are complements. The end result is that optimal output (service level) is lower than in the case of one integrated principal and that optimal incentives are lower powered.

however, very large. The COV (coefficient of variation) is over 1.0 in almost every psychiatric DRG-group (Frank & Lave, 1986). This makes the DRG system rather difficult to use within the psychiatric specialities. Some providers reimbursed by Medicare are exempted from the prospective payment system. Psychiatric speciality hospitals and qualified psychiatric speciality units in general hospitals has been exempted on grounds that there are; (1) large systematic differences in mean treatment costs among major groups of providers, particularly between general and psychiatric speciality hospitals, and; (2) unusually high levels of variation around their average value in treatment costs within psychiatric DRGs, compared to other services (Dada et al., 1992, p. 483–484).

However, there is not only the financial risk to consider. Frank & Lave (1984) found a significant impact of financial incentives on psychiatric inpatients. They noted that a limit of 25 days reimbursement per admission reduced length of stay by about 28 percent compared to states in the U. S. without such a limit. Rupp et al. (1984) found that early discharges, due to the prospective Medicare reimbursement system, lead to offsetting costs associated with re-admission of the patients.

Since the problem of early dismissal in psychiatric care is well known, different methods has been proposed to limit the negative effects of prospective payments on psychiatric patients (Frank & Lave, 1986, p. 90, Sharfstein, 1991). Common for these proposals is that they aim for mitigating the incentives to shorten the length of stay for the patient. This can be accomplished by combining the diagnosis related reimbursement with a payment based on the patients length of stay at the hospital, i. e., effectively a kind of mixed system of prospective payment and cost reimbursement.

4 Theoretic analysis

The first section (4.1) is an introduction to the principal-agent problem formed between payor and provider in health care production. Section 4.2 is an analysis of the effects of asymmetric information in a principal-agent relationship. Section 4.3 deals with how yardstick competition is used as a method to regulate health care providers and the problems introduced by "hidden information" and "hidden action". A discussion about the prospects of combining the models of section 4.3, concludes chapter 4.

4.1 Principal and agent

The need for third-party payment mechanisms naturally turns the problem of paying providers of health care into a principal-agent problem. The term "principal-agent" is due to Ross (1973). The principal-agent literature is concerned with how one individual, the principal, can design a compensation system (a contract) which motivates another individual, his agent, to act in the principal's interest (Stiglitz, 1987). A principal-agent problem arises when there is imperfect information about what action the agent has taken or should take. This is the case when an insurance company pays for the costs of health care consumed by a patient. In this case, the insurance company is the principal, the health care provider is the agent and the benefits accrue to the patients. There is also another agency problem involved in the production of health care. This problem origin in the relation between the physician and the patient. The physician can be stated to act as an agent for the patient.

There are three important reasons for the existence of principal-agent relations. One source of principal-agent problems is moral hazard/hidden actions. Another reason for the existence of principal agent problems is the need for risk sharing between the principal and the agent. If one party, say the agent is risk averse, a contract that puts all the responsibility for the outcome on the agent might not be pareto-optimal. However, if the principal consequently assume some of the responsibility for the outcome the agent might then not have as strong incentives to achieve a good result. A third reason for the existence of principal-agent problems is that the principal might not know how efficient the agent is. This lack of information might make the principal pay too much to the agent to make him accomplish the task set out for him. If there is a social cost of public funds, and the agent is paid by public funds, this might entail a welfare loss.

To summarise the issue, there are generally three problems that together, in one or the other combination, make up the principal agent problem;

1] Hidden action (moral hazard)

2] Risk aversion

3] Hidden information (adverse selection)

A fourth issue will be added in our discussion about the optimal regulation of health care providers. This is the issue of "unobservable quality". This problem is essentially only another type of "hidden action", but in this exposé I will mainly distinguish between situations where the agents "hidden actions" could decrease the costs of production and when the actions of the agents could increase the benefits of the consumers.

4.2 Asymmetric information – Imperfect productivity measurement

The discussion about the optimality of different reimbursement systems is based on a notion that there exist only imperfect productivity measures. It is practically impossible to formulate a first-best reimbursement method when there is information asymmetry. The simplest way to model the imperfection of reimbursement schemes in practice is to use two variables, slack and quality, to signify the undesired investments by hospitals and the desired ones, respectively. The observation that imperfect productivity measures lead to problems with hidden action takes us to a more general discussion. The regulation models in health economics analyses mainly two reimbursement methods, fixed prospective payments for a patient, cost reimbursement, and linear combinations of those two. However, there are many variables, more or less easy to observe, that can be used to base reimbursements on. Patient days, capital investment and personnel density are some other factors that can be used. It is not necessarily true that a more complex reimbursement system performs better. The complexity might in itself preclude new solutions and when the number of variables become very large, the information problem probably becomes overwhelming.

Productivity measurement in health care production is difficult. Measures such as patient days and number of appendectomies all have their drawbacks. Variables that are easy to measure are often not meaningful to use as a goal for the organisation (or society). Weisbrod (1992) examines the properties of different productivity measures for hospitals. His conclusion is that using imperfect productivity measures as a basis for reimbursement, will introduce different forms of sub-

optimising behaviour among providers. The production function of health care quality can be expressed as

$$s = s(L, E, P, C) \tag{1}$$

where L is the average length of stay for patients, E is the number of employees per patient, P is the performance of employees and C is health care following discharge (external effects). It is possible to express the effects of different reimbursement systems as a vector of the monetary amount of resources channelled through different forms of reimbursement. The intermediary inputs can be written as functions of vectors of the different reimbursement methods

$$L(X),\ E(X),\ P(X),\ C(X) \tag{2}$$

where X_1 could be the amount of prospective reimbursements, as one example. Substituting the production functions for intermediary inputs into the production function of health care quality then yields

$$s = s(X) \tag{3}$$

where X is a vector of different reimbursement methods, each measured in monetary terms. The problem of the regulator is then

$$\begin{aligned} &Min\ X \\ &\text{s. t.} \\ &s(X) = s^* \end{aligned} \tag{4}$$

where s^* is the regulators desired quality level of care at the hospital. If the number of regulatory instruments (reimbursement methods) is less than the number of factors the regulators want to influence indirectly, if some variables are difficult to observe or to contract on, the solution will be second-best with regard to the use of intermediary inputs. The choice

of regulatory instruments thus has to be made by assessing the negative and positive effects of each reimbursement method on the use of intermediary production factors. Prospective reimbursements could decrease the length of stay of patients, which might have a negative effect on quality, but perhaps also increase cost reduction efforts from the staff at the hospital, which would make it possible to achieve higher quality for a given cost. If *quality*, thus, is what is desired in production, there will be negative effects from introducing productivity measures not perfectly correlated with *quality*. These measurement problems increase the scope of hidden actions within the agents realm. Although the problem might be multidimensional, a two dimensional analysis of the problems has its advantages in terms of simplicity and tractability.

4.3 From cost reimbursement to prospective payments

4.3.1 Yardstick competition

The basic approach to regulation by yardstick competition is to try to explain the differences between firms by using a regression model in order to create a "shadow firm" (Schleifer, 1985). Using yardstick competition makes it possible to replace the cost-reimbursement method with fixed, prospective, prices derived from observations on other similar firms in the economy. The regulator ought to use all available information to reduce informational asymmetry. Instead of using Schleifer's original model, I will here follow the simpler approach of Laffont & Tirole (Laffont & Tirole, 1993, p. 85).

Assume a regulator who wants to limit the cost of reimbursing a population of hospitals in the economy. The costs of running the hospitals are all different, but these cost variations can be traced back to

certain exogenous differences that can be observed by the regulator. The cost of a hospital is denoted $C_j = \beta_j - e_j$ and is a function of a fixed cost parameter, β_j, and the managers effort to reduce the costs of the hospital, e_j. The manager has a utility function

$$U = t_j - \psi(e_j) - C_j, \tag{5}$$

which is a function of the reimbursement to the hospital, t_j, the managers disutility of effort to reduce the costs of the hospital, $\psi(e_j)$ ($\psi' > 0$, $\psi'' > 0$) and the cost of the hospital, C_j. His participation constraint is $U \geq 0$. If there is a cost of leaving rent to the hospital, the participation constraint should be binding at the optimal solution.

Assume also that the regulator can construct a "shadow hospital" by running a regression on the costs of similar hospitals in the economy, with characteristics such as urban or rural location and the number of patients of a certain type or with a certain diagnosis. The predicted values of this regression can then be used to "construct" a hospital similar to hospital *j*. We can denote the cost of this "shadow hospital" as C_i. If the regulator then offers hospital *j* a transfer of

$$t_j = \psi(e^*) + C_i \tag{6}$$

the manager of hospital *j* then maximises

$$\psi(e^*) + C_i - \psi(e_j) - C_j, \tag{7}$$

where $\psi(e^*) + C_i$ is constant and chooses $e_j = e^*$.

Since all other hospitals also choose $e_j = e^*$, the hospital earns no rent and production is thus carried out efficiently.

4.3.2 *Prospective payments under hidden information*

The question now arises about what will happen when there are no "shadow hospitals" that can be used as a comparison and benchmark to set the optimal price (or reimbursement level). The problem with hidden information in health care production (on the supply side) has been dealt with by a number of researchers. Two of the most well known models are Pope (1990) and Siegel et al. (1992). Pope analyses the problem of hidden information in a framework where the provider of health care is assumed to be risk neutral. Siegel et al. extends this model to a setting where the provider is risk averse. Pope (1990) uses a model where cost is a function of some exogenous factors, the quality level and managerial slack. He shows that the optimal payment is a linear function of the hospital's realised cost. Popes model differs from traditional models of regulation, where there is no hidden information parameter (exogenous factor) and it therefore is possible to use yardstick competition to enhance efficiency. The existence of unobservable cost differences is the starting point of Pope (1990) who assumes a simple cost-function of the type

$$C_j = \beta_j + s_j + \hat{e}_j, \tag{8}$$

where s is investment in quality, β is exogenous cost differences between hospitals (not possible to reduce through yardstick competition), and $\hat{e}_j$ is slack in the organisation.[11] The index j denotes the firm (hospital).

---

[11] Pope uses the concept "slack in the organisation" and have a positive sign for e. In the model presented in part two of this report, the concept "effort to reduce slack" is used instead. There is no practical difference between the two cost functions, however, since the cost function also could be expressed as $C = (\beta + e) + s - e$, where $\bar{e}$ is the maximum slack conceivable. β could then be redefined as $\beta = \beta + \bar{e}$ and e

The regulator wants to reimburse β and s. If the reimbursement is restricted to a linear function of the realised cost, it is possible to write it as

$$t_j = a + rC_j. \tag{9}$$

The r parameter denotes the degree of retrospectiveness in the reimbursement system. An r of zero denotes a fully prospective system and an r of one denotes a cost-reimbursement system. Pope (1990) makes the assertion that the regulators aim is to minimise squared deviations between actual reimbursements and the "desired" payment. The desired reimbursement is assumed to be compensation for intrinsic cost differences between hospitals plus an extra payment, sufficient to make the hospital achieve average quality level among the population of hospitals in the sample. The desired payment would then be

$$t^d = \beta_j + E(s_j). \tag{10}$$

In this setting, minimising squared errors of the difference between optimal and actual reimbursements means to

$$\min_{a,r} \frac{1}{n}\sum_j \left[\beta_j + E(s_j) - a - rC_j\right]^2. \tag{11}$$

This minimisation process gives us the optimal parameter values of a and r.

Siegel et al. (1992) extends Pope (1990) to account for the financial risk aspects of patients and providers, and develops a model to incorporate risk into the reimbursement formula. They call the reimbursement

---

would be the managers effort to reduce slack.

model "an integrated risk-based prospective payment system". The desirable payment under full information would be

$$P_{ij}^{d} = \beta_{ij} + s_j \tag{12}$$

for patient i in hospital j, where β_{ij} is the unobservable hospital cost, including differences in treatment costs for different patients, and s_j is the quality level of treatments at the hospital. The cost of patient i at hospital j can thus be decomposed into three terms

$$C_{ij} = \beta_{ij} + s_j + \hat{e}_j, \tag{13}$$

where β_{ij} is the level of exogenous cost differences for a patient and hospital specific costs, s_j is the quality (treatment intensity) at the hospital and $\hat{e}_j$ is a slack.

Siegel et al. (1992) postulate that the desired payment should converge to the hospital mean and thus also that the desirable payment should be a linear combination of hospital average cost (ϕ_j), individual patient costs (X_{ij}) and national average cost (μ), according to the formula

$$P_{ij} = \phi_j + r_p X_{ij} + r_n \mu, \tag{14}$$

where the coefficients for patient cost and national average cost are denoted by r_p and r_n respectively. ϕ_j is a constant. The coefficients are, as in Popes model, calculated by minimising a loss function, that is, the deviation between the hypothesised optimal costs and the reimbursements. The formula is thus essentially also a yardstick competition model, but reimbursements are here also adjusted to compensate the hospitals for risk.

4.3.3 Hidden action – The physicians' behaviour

The analysis in this thesis builds on the notion that altruism is a product of the social environment and not given by nature. This makes the analysis congruent with most economic models of human behaviour. Physicians are assumed to be subjected to social feedback as a consequence of their behaviour towards the patients. A model of how judicial feedback towards physicians might be put to use as a regulatory instrument is the approach by Blomqvist (1991). This model assumes that a stochastic performance guarantee is used to control the behaviour of physicians. This performance guarantee is thought of as a penalty imposed on the physicians if the outcome (health of the patient) falls below a certain level. Consideration of enforcement costs may mean that it is second-best efficient to accept a system which leads to only partial elimination of the losses due to imperfect agency (Blomqvist, 1991). If the control task is performed efficiently and responsibility is put directly on the physicians, we will in effect have a situation where the physicians are forced to act as double agents. A double agent, of course, has two principals. One of these is the hospital manager; the other might be some abstract entity acting on behalf of the patient, such as the judicial system, quality controlling authorities, peer reviews, etc. The feedback should also come directly from patients.

Most regulation models assume a unitary firm organised to maximise the manager's utility. If we abandon this "black-box approach" to regulation, we find that a natural starting point is to use "agency-models"[12] to explain the behaviour of health care providers. This

[12] The idea of imperfect agency bears a strong resemblance to the theories of supplier induced demand (Anderson et al., 1981, Reinhart, 1987).

literature started with Ellis & McGuire's articles in the Journal of Health Economics (Ellis & McGuire, 1986, 1990). They view the physician as an agent of the patient. This is done by putting the utility function of the patient into the utility function of the physician. Technically, this is a simple solution, but it creates some problems as the model becomes more complex.

The hospital manager is assumed to maximise the profit function

$$\pi(s) = t(s) - C(s), \tag{15}$$

where s is the quality of treatments at the hospital, t is the transfer from the regulator to the hospital and C is the hospitals cost.

The physicians maximise their utility functions, in which the profit of the hospital is an argument.

$$U = U(\pi(s), B(s)) \tag{16}$$

$B(s)$ is here the benefits of the patients.

The first order conditions of the physician's utility maximisation can be written as

$$\frac{\partial U}{\partial B}\frac{dB}{ds} + \frac{\partial U}{\partial \pi}\frac{d\pi}{ds} = 0. \tag{17}$$

The physician's marginal rate of substitution between profits of the hospital and utility of the patients can then be expressed as

$$MRS_{\pi,s} = \frac{(\partial U / \partial B)B'(s)}{(\partial U / \partial \pi)} = MRS_{\pi,B}B'(s) = \alpha B'(s). \tag{18}$$

The term $B'(s)$ is the first derivative of the benefit function with respect to quality. The parameter $\alpha = MRS_{\pi,B}$ is then interpreted as to what degree the physician is a "perfect agent " for the patient. A perfect agent

is in the health economics literature defined as an agent who weights the social value of the profit of the hospital equal to the social value of the benefits of the patients.

The main result of the Ellis & McGuire model is that the optimal reimbursement might not be either prospective payments or cost-reimbursements, but something in-between. For a linear cost function, the optimal incentive scheme is characterised by

$$\alpha B'(s) = (1-r)c, \tag{19}$$

where c is the marginal cost of quality and r is the degree to which the hospital's costs are reimbursed retrospectively, according to the formula

$$t(s) = a + rC(s). \tag{20}$$

4.4 Discussion

Laffont & Tirole (1989)[13] discuss the problem of quality in regulation. Their models in this area are based on the division between search and experience goods, that is, when quality indirectly affects demand or future contracts. Quality in health care may often not so much be a concern of the hospitals as profit maximising units, but more of the employees working in the hospitals (the physicians) since quality itself has small effects on demand in settings without explicit competition between providers. Often hospitals also refrain from trying to affect the quality level directly, but instead leave that task to physicians working

[13] A brief introduction to the regulation theory of Laffont and Tirole can be found in section 2.3 in Part two of this thesis. The regulation model is too technically complicated to be presented in some short sentences here.

in the hospital. Another way of putting it could be to say that the "black-box" approach (Laffont & Tirole, 1993, p. 667) to the regulation of hospitals might not be as fruitful as in many other areas. Many other elements of the Laffont & Tirole model are, however, applicable to the health care context. Unverifiable differences in case mix is parallel to unverifiable exogenous differences in efficiency among firms, and hidden action is likely to be a problem in the health care industry, as well as in most other industries.

To concentrate only on adverse selection, as in Pope (1990), in the discussion about the regulation of hospitals will ultimately be too limiting. Realistically, cost sharing will have strong dynamic effects on the behaviour of the regulated firms. A way of introducing hidden action into the model would be to make use of the regulation theories of Laffont & Tirole (1993). Newhouse (1996, p. 1247) notes that the reimbursement system proposed by Pope would leave some hospitals incurring losses, unless the fixed payment is sufficiently high. If there is a marginal cost of public funds (as assumed in most regulatory literature) this also poses a problem in the model. Laffont & Tirole (Laffont, 1987, Laffont & Tirole, 1986, 1993) have however, earlierly addressed that problem, as Newhouse notes.

> "Like the health economics literature, firms are not assumed to be homogeneous with respect to cost. Although observed total cost varies with managerial effort , as in both Schleifer and Pope. Laffont and Tirole assume that observed cost is also a function of a cost parameter that varies by firm and is unknown to the regulator, analogously to Popes β_j." (Newhouse, 1996, p. 1247)

In the Laffont & Tirole model, a regulator cannot tell from realised costs whether the manager is lazy or if the hospital's cost is high for some

other reasons. It is assumed, however, that the regulator has prior knowledge of the distribution of the β_j parameter.

Unlike the Pope model, the Laffont & Tirole approach implies that firms break even and that there is a marginal cost of public funds. Newhouse (1996, p. 1248) thus argue that it would be possible to combine the approaches of Pope and Laffont & Tirole. The underlying cost parameter could, for example, arise from unobserved case mix variation, Popes β_j, which a regulator wishes to reimburse. The framework of bidding fits well with the notion of HMOs contracting for the care of the employees of a firm, according to Newhouse. He notes that HMOs do not necessarily know their true cost parameter when bidding. This information is, however, not essential in the Laffont & Tirole model.

> "The possibility of decentralising through a menu of linear contracts shows that introducing noise (accounting or forecast error) in the cost function has no effect. The intuition for this result is that the introduction of noise does not affect expected cost." (Laffont & Tirole, 1993, p. 72)

The existence of noisy cost information does not affect the optimality of incentive schemes and it does not affect the social welfare either. This is due to the fact that firms (agents) are assumed to be risk–neutral. However, it is essential that the agent submitting a bid have better information about the expected costs than the regulator. Otherwise, the relation between principal and agent would be reversed. Newhouse (1996, p. 1248) notes that it would be possible to imagine each HMOs bidding a price schedule, a fixed amount for each enrolee's use. Selecting winning bids then requires a method to combine these two pieces of information. The model presented in Part two is much in line with the ideas of Newhouse. It is also a synthesis between Ellis &

McGuire (1986) and Laffont & Tirole (1993) with all the elements; quality, efficiency and adverse selection. The integration of the models of imperfect agency (unverifiable quality), hidden information (unverifiable costs) and hidden action with respect to productive efficiency, is the aim of Part two of this thesis.

5 Conclusions

The theoretical literature and the empirical evidence of hospital regulation show us that the issue of hospital regulation is complex and that several concepts in the theory of asymmetric information must be applied simultaneously to make it possible to grasp the issues at hand. In some instances we can see that existing solutions have antedated the economic theory about what the optimal regulation should be. An example of this is the difference between the use of prospective payments in psychiatry and surgery respectively. In surgery, the procedures are fairly well defined and the prospects of dumping patients on other caretakers are small. This could be thought of as a large degree of perfection in the agency relationship between patients and physicians, that is, a high α parameter, in the terms of the Ellis & McGuire (1986) model. Not only is the α parameter high, but the cost variation is fairly small as well, within a certain diagnosis related group. This makes risk exposure by physicians smaller. This fact is perhaps not so important in surgery anyway, since surgical procedures build on a long driven specialisation and fairly large amounts of capital, which makes the optimal reimbursement tilt further towards prospective payment, as seen in the model by Siegel et al. (1992). The small variation in costs, also makes the adverse selection parameter, β, less variable, and according to the model by Pope (1990), this is in favour of prospective payments, relative to cost-reimbursement. The production process in surgery can be divided between the production of intermediate products and their use by patients. The well-defined production of intermediate products, and the larger degree of standardisation of these intermediate outputs, makes it possible to gain efficiency by implementing steep

incentive schemes for the hospital. Although there is a trade off between efficiency and quality, the potential to increase efficiency makes it more advantageous to try to gain efficiency by sacrificing some quality in production.

In section two of this thesis, I will formulate a theoretical model to combine the concepts of hidden action, hidden information and imperfect agency, and present a normative model of what the optimal regulation of a hospital should look like. This model is a combination of the aspects analysed in this first section and is intended to show that the problems analysed combine the problem of optimal reimbursement as a task of balancing those aspects.

– Part Two –

Imperfect Agency and the Regulation of Hospitals

1 Introduction

Many problems in health care production can be related to lack of cost control, especially the dichotomy between costs relating to activities of administrative staff and medical staff. Given their large responsibility with respect to medical issues, physicians control a large part of the costs, but have no responsibility for the over-all cost development. It has been shown in a number of studies that medical staff controls between 70 and 80 percent of hospital costs (Enthoven, 1980, Saltman, 1985), an important factor to take into account when trying to develop ways of controlling costs. However, decentralised decision-making in the hospital hierarchy has made it hard to track aggregate costs and to co-ordinate different units. In some countries, especially in the United States, privately practising physicians can admit and treat their own patients at a hospital. Although this is also practised in Europe, hospitals there usually employ physicians for a fixed salary. This difference can perhaps partly be explained within the model itself,[14] but the type of hospital I use as a basis for the analysis is one where physicians are employed by the hospital.

When discussing optimal incentive systems, it is necessary to ask whose incentives we are talking about. Often an expression like "the incentives of the hospital" is used. Is it the manager's incentives that is meant or perhaps the incentives of doctors (Glaser, 1987, p. 57)? Nobody knows, since the ownership role of the hospital is often not clear. When

[14] A system where physicians have been assigned total cost-responsibility for patients can be seen as a contractual arrangement to give as many incentives as possible to physicians in order to ensure cost-effectiveness.

incentives are given to hospitals, regulators do not let managers run off with the extra money earned by making the hospital more efficient. Rather, the money is often reinvested into the hospital, and the manager might get a rise in pay or status. This may well mean that the manager is directly interested in making the hospital cost-effective and prosperous. One problem is that there is no direct connection between the owners' objectives and the expected reaction by the manager. Suppose that the manager does not get paid extra for making the hospital prosperous. What use will it then be of giving the hospital incentives when the manager not in some way is affected by these also?

When studying private companies, it is often assumed that managers are trying to make the company profitable. This is because it is also implicitly assumed that owners put pressure on the manager and, by carrot and stick methods, induce him to take actions that are good for the company's shareholders. However, hospitals are often publicly owned, there is thus no guarantee that managers will want to make the hospital a profit maximising unit. They might have other objectives in mind, such as size maximisation or fringe benefit maximisation.

Expense preference theory[15] is an alternative to profit-maximisation assumptions. Williamson (1963) develops a model to account for

---

[15] While expense preference theories are perhaps a little off the mark, when we discuss profit maximising firms with strong owners, they can still be useful when studying proprietary hospitals. The responsibility of doctors for quality of health care remains even when hospitals turn into profit maximising units, although. There might be some differences in weights put on different aspects of hospital performance. However, the differences between not-for-profit and proprietary hospitals should not be overstated. It has been concluded that the prospective reimbursement system for Medicare, introduced in the United States in 1983, has affected non-profit as well as for-profit hospitals (Oswald et al., 1994). In fact, the not-for-profit hospitals have increased efficiency even more than proprietary hospitals. We must, however, remember that physicians only indirectly are affected by incentives directed to the hospital and that it ultimately is them who make the decisions about tests, treatments and when to dismiss the patients.

differences in the objectives of the company and its managerial staff. A common assumption is that managers maximise their own utility instead of company profits. This behaviour should be most prevalent in non-profit companies, making the theory a natural candidate to the study of hospitals. The clear division of responsibility within the hospital itself also raises questions whether profit maximisation is a reasonable assumption. The physicians have a large ethical and legal responsibility for many aspects of the hospital production process. This makes it relevant to ask whether profit generation for the hospital is the only relevant objective for the physician. The physician might want to balance the profits of the hospital with the benefits of the patients, when deciding how long to keep a patient in the hospital and the resources to spend on each patient.

When a prospective reimbursement system is used, every extra day of care for a patient can potentially inhibit admission of another patient. This means that every day of care has a substantial opportunity cost, at least if the hospital's capacity is limited. The decision to keep a patient an extra day is ultimately made by the physician. This means that hospital management cannot control the total level of costs in the hospital, but will have to adjust to what they expect that physicians will decide. Does this mean that managers of the hospital have no influence at all or that they cannot influence the profitability of the hospital? Not necessarily, since managers may put different sorts of pressure on the medical staff. This being the case, it can still be socially optimal to give managers of the hospital incentives to reduce cost. These pressures may force physicians also to try to minimise costs, at the expense of the quality of care in the hospital.

A common view of hospital organisation is therefore to look at the physician as a kind of "double-agent" who weigh interests of the

hospital against interests of the patients (or in some cases, perhaps even interests of the society at large). It is a model based on this view that I wish to develop in this paper. The model will also be applicable to situations where there are large external effects of health care production. The problem of imperfect agency can be treated in the same way as the problem caused by external effects of quality in health care. The agency problem can be redefined, imagining the doctor as a less than perfect agent of society. As an example, bad quality of mental health care can, besides the negative effect on patients, cause costs for relatives, law enforcement and the social security system. This is of course also true for ordinary health care to some degree. It is easy to adjust my "double agency model" to take account of such effects.

2 The model

2.1 The patients utility

Following Chalkley & Malcomson (1995) I specify a net utility function for the patients of the type $U(s,r)=u(s)-r$ where s is the quality level in the hospital and $r\in[\underline{r}-\bar{r}]$ is the patients reservation utility of treatment.[16] The utility function of the patients is assumed to be monotone and concave in s. It is also necessary to make the assumption that quality and cost reduction effort needed is independent of the number of patients treated. This condition can, of course, only be approximately satisfied and only for minor changes in demand. Demand for treatment is given by

$$x(s)=\int_{\underline{r}}^{u(s)} f(r)dr\,. \tag{1}$$

$f(r)$ is the distribution function of the patients' reservation utility. The social benefit function for a utilitarian is thus

$$B(s)=u(s)\int_{\underline{r}}^{u(s)} f(r)dr \tag{2}$$

if we assume that only patients with a positive net-utility of treatment want to be treated at the hospital.

---

[16] The reservation utility can be thought of as the utility received by the patients from the best alternative treatment. This could either be treatment at another hospital or some other kind of health care available.

This is equivalent to the expression $B(s) = u(s)F(u(s))$. A condition for a stable equilibrium is that the benefit function has a negative second derivative. If demand is sensitive to quality, we could get local convexities. If this is the case, it will anyway be possible to find a globally optimal solution to the situation where there is only imperfect agency problems or moral hazard and imperfect agency. When adverse selection is introduced, the consequences of local convexities are not clear. What can be noted, however, is that the existence of local convexities might mean that large quality differences might appear between different hospitals, due only to small changes in the variables.

If the capacity of the hospital, for some reason, is fixed, the social welfare function will be

$$B(s) = xu(s) - \int_{\underline{r}}^{r(x)} rf(r)dr\,. \tag{3}$$

The difference with (2) is that the demand is fixed at a predetermined level, x. In this case, it is obvious (by the concavity of the individual utility functions) that the benefit function will be well behaved.

The objective of different rationing methods used in health care is to make the patients choose health care in a manner that maximises social welfare. An optimal rationing system can be constructed either by patient fees or queues. The social aim of these systems is, even if their distributive consequences can be somewhat different, rather similar. The patients with the highest net utility of health care should be the ones who are admitted to the hospital. This means that, with or without rationing, as long as the cost of quality is independent of demand, higher quality will make it socially optimal to treat more patients, since more patients will find the treatment at this hospital superior to treatment in another way. In practice, however, the assertion that the

cost of quality is independent of the number of patients is of course disputable. Relaxing this assumption slightly does not change the properties of the benefit function to any large extent.

2.2 The physicians maximisation problem

The problem of finding the optimal incentive scheme can be conceived as a game between the regulator, the hospital managers and the physicians. In order to be able to reach a "regulatory Bayesian Nash-equilibrium",[17] I have to make some explicit assumptions about the game. The following is assumed to be true:

1] Hospital managers write a contract with the regulator (just as in the basic regulatory model by Laffont & Tirole (1993) referred to above).

2] Once the contract is signed with the regulator, neither the hospital, nor the regulator, have any direct influence over the quality of care chosen by the physicians.

3] The physicians always maximise utility and cannot commit themselves to exercise a certain quality.

4] Both the regulator and the hospital managers know the utility function of the physicians and can predict what actions the physicians will take once the contract has been signed.

The physicians' utility function can be expressed as

---

[17] A Bayesian Nash-equilibrium is a sequential equilibrium in strategies of the players concerned. This means that a player cannot credibly commit to exercise an action that is detrimental to his own objectives just to force a previous player to make a choice that is against his interests. In this example, physicians can't make a threat to depart from their utility maximising positions to make it optimal for the hospital to choose a different type of contract or exert a different level of effort.

$$u_D(U_H - U_H^*(\beta), B(s), \mathbf{n}), \qquad (4)$$

where U_H is the rent of the hospital, U_H^* is the rent-target of the hospital[18] and $\mathbf{n}$ is a vector of other factors that affect the physicians' utility. I will here include the utility of the patients directly into the utility function of the physician because it is the easiest way to model his behaviour.

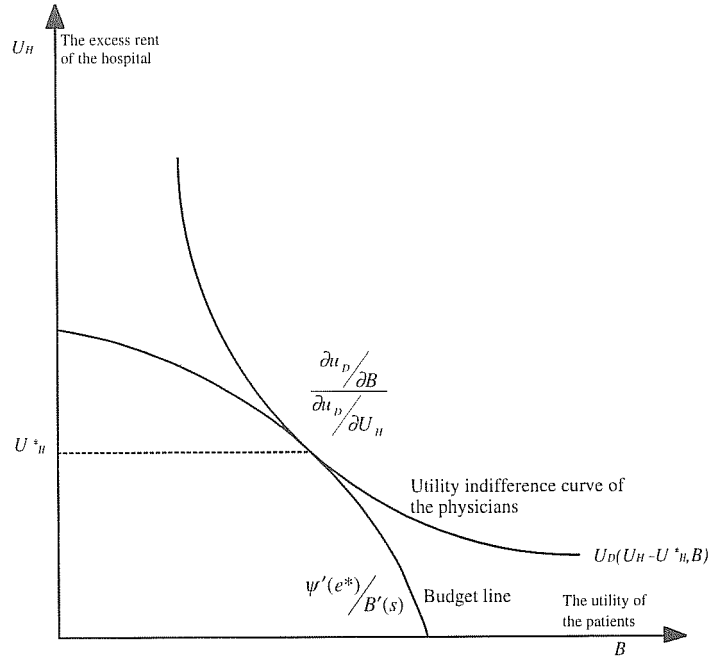
The benefits of the patients is a function of the quality level exercised in the hospital, $B(s)$ ($B'(s)>0$, $B''(s)<0$). The utility function of the physicians is based on the fact that the hospital's work-incentives directed towards the physicians are independent of the profit-target set for the hospital. This makes sense, since there is no obvious reason why a physician working in a high-profit hospital would be happier than a physician in a low-profit hospital would, *ceteris paribus*.

Assuming usual properties of the utility function, I can express the physicians maximisation problem graphically in Figure 1.

---

[18] When solving the regulatory maximisation problem, under all appropriate constraints, the resulting rent-level of the hospital will be U_H. This level is either decided by the IR constraint of the hospital, as in the first two models, or, as in section 2.6, by a combination of IR and IC constraints. The solution to the maximisation problems imply that the physicians always maximise their utility at the point where $U_H=U^*_H$. However, to reach that conclusion we have to make it possible for the physician to make decisions that would mean that this rent-level was not achieved in the end. This means that we can treat the variable U_H in the same way as the rent variable in the basic regulatory model by Laffont and Tirole, as long as we make sure that imposed restrictions make the physician choose a quality level that results in $U_H=U^*_H$.

Figure 1 The trade-off between quality and profits



Since the physician only can adjust the benefits of the patient indirectly via higher quality, the budget line is concave. This is a direct consequence of the fact that the benefit function, *B(s)*, is assumed to be monotone and concave. The budget line must also be adjusted by a factor depending on the slope of the incentive scheme used. A decrease in cost, due to lower quality, will only affect profits of the hospital by a factor proportional to the amount of cost sharing between the regulator and the hospital. Since the slope of the incentive scheme is equivalent to

$\psi'(e^*)$[19] it is possible to write the slope of the physicians budget constraint at any point as $\psi'(e^*)/B'(s)$.

The equilibrium condition can then be written as

$$\frac{\partial u_D / \partial B}{\partial u_D / \partial U_H} = \frac{\psi'(e^*)}{B'(s)}. \tag{5}$$

If we Taylor-expand the utility function of the physician around some suitable values, we get the linear expression:

$$u_D = R + \gamma_1 \left[U_H - U_H^*(\beta) \right] + \gamma_2 B + \gamma \mathbf{n}, \tag{6}$$

where γ is a vector of coefficients. This means that the marginal rate of substitution between rents of the hospital and benefits of patients, for a given rent-target, can be written as $\alpha = \gamma_2 / \gamma_1$. We can then rewrite the equilibrium condition (5) as $\alpha B'(s) = \psi'(e^*)$.

2.3 Regulation under asymmetric information

The Laffont-Tirole model is based on the assumption of risk neutrality on behalf of both the principal and the agent. This assumption makes the model a lot more convenient than models where risk aversion is a crucial element, either in the case of the agent or the principal. Some of the advantages of the Laffont-Tirole approach are:

- The model will be robust to specification errors, as, for example, randomly distributed errors in the cost function.

---

[19] The rent gradient, $\psi'(e)$ is equal to the degree of marginal cost sharing implied by the contract between the regulator and the hospital. This is further explained in section 2.5.

- The regulation can be based upon expected costs, and not necessarily on actual costs.

This makes the model applicable in circumstances where the variance of outcomes is very large. As long as uncertainty is purely symmetric, e.g. neither the principal nor the agent knows the outcome, but they have one common expectation of the cost outcome, the optimal regulatory scheme will be the same as if both principal and agent would know the cost outcome in advance.

Another important feature of the model is that both adverse selection and moral hazard are addressed. In fact, without either of these items, we would not have a regulatory problem at all. Without risk aversion and adverse selection, responsibility would be put on the contractual party who has control over the moral hazard parameter, e.g. usually the agent. The socially optimal solution would then be achieved and no type of risk sharing would occur.

When there is both adverse selection and moral hazard, it will generally not be optimal to give the agent full responsibility for variations in outcomes. To do this would yield the first best effort level from the agent, but only at the cost of leaving very high economic rents to him. If public funds (and I assume that the hospital is publicly financed) are used, the principal will want to limit the amount of rent received by the agent. The social cost of public funds will, in line with Laffont & Tirole, be denoted λ. It is normally assumed that λ is in the range 0.2–0.4 (Ballard et al., 1985).

The solution to the problem of finding the optimal method to regulate a hospital is based on implementation of incentive schemes that imply a higher or lower degree of marginal cost sharing between the regulator and the regulated. The degree of cost sharing is here referred to as the slope of the incentive scheme. This slope can take values between zero

and one, where zero is a scheme where the hospital bears no cost overruns, compared to a cost target specified in the contract between the regulator and the hospital. A scheme where the hospital has full responsibility for deviations from the cost target, has a slope of one. The optimal incentive scheme, as derived by Laffont & Tirole, can be described as a menu of linear incentive schemes where the regulated firm selects its preferred alternative. This selection reveals the intrinsic efficiency level of the firm (the adverse selection parameter). In order to have the agent select the contracts presented, it must not be optimal for the agent to pretend to be less efficient than he really is by exerting lower effort.

The problem of finding an optimal regulatory scheme can be looked upon as a problem of making the agent assert a high effort, but at the same time limiting rents of the agent. Assuming that the agent has an intrinsic efficiency ranging from $\underline{\beta}$ to $\overline{\beta}$, distributed $f(\beta)$, and a cost function $C = \beta - e$ with a disutility of effort function[20] $\psi(e)$, $(\psi'(e)>0, \psi''(e)>0)$, we can use the steepest incentive scheme for the most efficient types of agents (lowest β). If we would use the same incentive scheme for an agent with an intrinsic efficiency slightly less than the most efficient one, e.g. $\underline{\beta} + d\beta$, we would risk that the most efficient type imitates the slightly less efficient by not exerting optimal effort and thereby capturing some extra rent. If we want to avoid this problem, it is

[20] Since both principal and agent are assumed to be risk neutral, we denote all variables in the cost function in monetary terms. This means that effort, e, quality, s, and disutility of effort, ψ, all will be defined in monetary terms. Effort, e, will be the monetary cost of cost decreases. Quality, s, will be the level of quality enhancement that marginally increases cost by one monetary unit. Disutility of effort, ψ, will be the monetary compensation the agent wants to exert a certain level of effort. These transformations make it easier to model the regulatory problem and since there are no obvious scales on which to measure these three variables, this approach will have no clear disadvantages.

necessary to bribe the most efficient type to make him stick to the optimal (first best) effort level.

It is fundamental, when implementing this regulatory model that the assumption of a monotone likelihood ratio holds. This means that

$$\partial\left(\frac{F(\beta)}{f(\beta)}\right)\Big/\partial\beta \geq 0. \tag{7}$$

This is not a very restrictive assumption. The basic technology, known by everyone, is characterised by parameter $\overline{\beta}$. There can be improvements of this technology. In Laffont & Tiroles notation $\overline{\beta}-\beta$ is the number of improvements. $F(\beta)$ is the probability that there are at least $\overline{\beta}-\beta$ improvements. The probability that there are more than $\overline{\beta}-\beta$ improvements and less than $\overline{\beta}-\beta+d\beta$ improvements is thus $f(\beta)d\beta$. Decreasing β from $\overline{\beta}$, $f(\beta)/F(\beta)$ is the conditional probability that there are no more improvements, given that there already have been $\overline{\beta}-\beta$ improvements. By earlier assumption this conditional probability increases as the firm becomes more efficient. This is thus a kind of "decreasing returns" assumption, satisfied by the most usual distributions; normal, uniform, logistic, chi-squared, exponential and Laplace.[21] The utility of the principal is:

$$S-(1+\lambda)(\beta-e+\psi(e))-\lambda U_H \tag{8}$$

where S is the social utility of the project and $U_H(\beta)$ is the rent given up to the agent in order to bribe him not to pose as less efficient than he

---

[21] For a proof, see Laffont & Tirole (1993)

really is. The optimal effort level for each type of agent must satisfy the condition (derivation, see Appendix 1):

$$\psi'(e(\beta)) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e(\beta)) \qquad (9)$$

This condition can be derived by using optimal control theory or simply by maximising the integral of the principal's utility function over β while maintaining the individual rationality constraint. The interpretation of this equation is further explained:

> "Equation (9) has a straightforward interpretation. Raise effort of types in $(\beta, \beta+d\beta)$ (in number $f(\beta)$) by δe. Productive efficiency increases by $(1-\psi'(e(\beta)))\delta e$ for these types, which yields social gain $(1+\lambda)(1-\psi'(e(\beta)))\delta e f(\beta)d\beta$. However, this also raises the rent of types in $(\underline{\beta}, \beta)$ (in number $F(\beta)$). From $[\dot{U}(\beta) = -\psi'(\beta - C(\beta))]$ the rent of type β is increased by $\psi''(e(\beta))\delta e d\beta$, and so is the rent of types $\tilde{\beta} < \beta$. The social cost of the extra rents is $\lambda\psi''(e(\beta))(\delta e)(d\beta)F(\beta)$. At the optimum the marginal cost must equal the marginal benefit which yields (9)" (Laffont & Tirole, 1993, pp. 65–66).

The Laffont-Tirole approach is easy to expand to cover many other problems related to regulation. Among those problems are: yardstick competition, incentives to provide quality, multi-product regulation and auctioning of incentive contracts. It is, as we will see, also possible to combine their approach with the agency theory often used in health economics.

2.4 No moral hazard

In this section, the cost function of the hospital is defined as $C=\beta+s$ since investment in quality is possible, but since moral hazard is absent, there

is no effort variable. The regulator knows the correct value of β. Note that all costs are, by assumption, reimbursed by the regulator. We can derive the optimal incentive scheme by first observing that a social optimum must satisfy the condition[22]

$$\begin{aligned} &\underset{s}{\text{Max}}\ U_H - (1+\lambda)(t+C) + B(s) - \theta(s); \\ &\text{s. t. } U_H^* = 0\ ,\ u_D = 0\ ,\ U_H^* = U_H\ ,\ \theta(s) = \alpha B(s) \end{aligned} \tag{10}$$

Since we assume that there exist a social cost of feedback to the physicians (the physicians are not "pure" altruists), the variable $\theta(s)$ is introduced, which for simplicity is assumed to be equal to the physicians valuation of quality. I will also assume that the participation constraints are satisfied, see Appendix 4. The restriction on the utility function of the physician can be expressed as

$$u_D^* = R + \gamma_1\left[U_H - U_H^*(\beta)\right] + \gamma_2 B + \gamma\mathbf{n} - \tau = 0 \tag{11}$$

at the optimal solution, where R is the constant of the Taylor-expansion and τ is a lump sum transfers that makes the physician indifferent between working or not. Since the wage for physicians is determined on a market where perfect competition reigns, we can regard the variable τ as exogenously determined henceforth. This means that the hospitals have no other option than to offer the physicians the market wage. This wage will always be equal to $\varpi - \alpha B(s^*)$ in equilibrium. I define the constant $\varpi = (\tau - R - \gamma\mathbf{n})/\gamma_1$ to simplify the derivations later on.

Because of the restrictions, the rent of the hospital can be written as

---

[22] We have transformed the physicians utility function, assuming that u_D=0 is the lowest utility at which the physicians wants to continue working.

$$U_H = t - \varpi + \alpha B(s). \tag{12}$$

The two last terms can, in equilibrium, be interpreted as the cost of salaries to the physicians. If the quality deviates from the optimal level, the utility level of the physician can be either negative or positive. In equilibrium, however, the utility level is always zero.

If we substitute U_H for (12) and substitute the ratio γ_2/γ_1 for α in the maximisation problem (10), we get

$$\begin{aligned} &\underset{s}{\text{Max}}\, B(s) - (1+\lambda)(\beta + s + \varpi - \alpha B(s)) - \lambda U_H - \alpha B(s) \\ &\text{s. t. } U_H = 0 \end{aligned} \tag{13a}$$

The last term is the social cost of extra precaution from the physicians, as a response to social and judicial feedback from society. Simplifying this we have

$$\begin{aligned} &\underset{s}{\text{Max}}\,(1+\alpha\lambda)B(s) - (1+\lambda)(\beta + s + \varpi) - \lambda U_H \\ &\text{s. t. } U_H = 0 \end{aligned} \tag{13b}$$

The first order condition of the maximisation problem is

$$B'(s) = \frac{1+\lambda}{1+\alpha\lambda}. \tag{14}$$

Since all variables here are known, it is possible for the regulator to implement the optimal solution. Adjusting the slope of the incentive scheme for the hospital so that it equals the marginal rate of substitution between profits of the hospital and quality for the patients in the physicians utility function does this. This can be implemented by giving the hospital a transfer that is a linear function of the achieved cost level, according to the formula

$$t = \frac{\alpha + \alpha\lambda}{1 + \alpha\lambda}[C^* - C] + \varpi - \alpha B(s^*), \tag{15}$$

where C is the achieved cost level s^* is the optimal quality level, and C^* is the cost that would be achieved if optimal quality were exercised. We can by substitution write the physicians utility function as

$$u_D = R + \gamma_1\left[\frac{\alpha + \alpha\lambda}{1 + \alpha\lambda}[C^* - \beta - s] + \varpi - \alpha B(s^*) - U_H^*\right] + \gamma_1 U_H^*(\beta) + \gamma_2 B(s) + \gamma\mathbf{n} - \tau \tag{16}$$

Maximising this with respect to s and dividing by γ_1, we get:

$$-\frac{\alpha + \alpha\lambda}{1 + \alpha\lambda} + \alpha B'(s) = 0 \Rightarrow s = s^*. \tag{17}$$

Second order conditions are satisfied, since $B''(s) < 0$ and α is a positive constant. Since the physicians maximise their utility at point s^*, the incentive scheme will, when applied to the hospital, indirectly lead to the achievement of social optimum.

It might seem strange that the marginal social utility of quality is not set equal to one in this case. The explanation for this is that social cost is higher than the purely monetary cost. The marginal social cost of quality is $(1+\lambda)s$, since the hospital and the physician are paid tax-money. Taxes have a distortionary effect, and thus the marginal utility of quality is more than one in optimum.

The condition (17) means that the slope of the incentive scheme should be approximately equal to α. This is roughly the same result as the one Ellis & McGuire (1986) reached. The difference is that I include the social cost of transfers to the hospital.

2.5 Moral hazard

The hospital's cost is here $C = \beta + s - e$. When we introduce moral hazard, we have to take into account that it is impossible for hospital management to control the actions of the physicians. From section 2.1 we have the restriction $\psi'(e) = \alpha B'(s)$ where α is the marginal rate of substitution between rent of the hospital and benefit for the patients in the physicians' utility function. The physicians will always adjust quality so that the condition $\alpha B'(s)=\psi'(e)$ is fulfilled.

With the same substitutions as in the previous section, I can now write the maximisation problem as

$$\begin{aligned} \underset{e,\xi,s}{\text{Max}} \; & (1+\alpha\lambda)B(s) - (1+\lambda)(\beta + s - e + \varpi + \psi(e)) \\ & - \lambda U_H^* + \xi(\alpha B'(s) - \psi'(e)) \\ \text{s. t. } & U_H^* = 0 \end{aligned} \tag{18}$$

Since it is possible for the hospital to vary the effort level, the hospital could, theoretically, achieve economic rents. The application of the optimal incentive scheme, however, precludes this possibility, since both lump-sum transfers and marginal cost incentives are used to induce the hospital to exert optimal effort (constrained optimality). The solution is

$$\psi'(e) = 1 - \frac{\xi\psi''(e)}{1+\lambda} \tag{19}$$

and

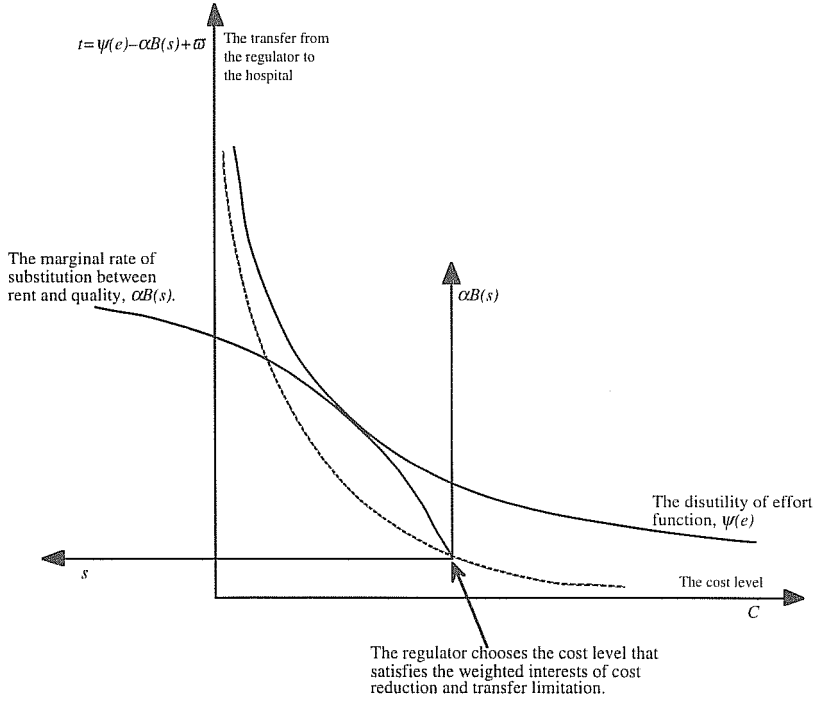
$$\alpha B'(s) = \frac{\alpha((1+\lambda) - \xi\alpha B''(s))}{(1+\alpha\lambda)}. \tag{20}$$

This means that the optimal incentive scheme is steeper than when moral hazard is not a problem. This is easily seen, since the Lagrange

multiplier is positive and the second derivative of the benefit function is negative.

The solution is illustrated graphically in Figure 2.

Figure 2 The regulator's choice of quality



The dotted line is derived by sliding the "benefit of quality" function along the marginal disutility of effort function to reach a level of cost that matches a certain transfer from the regulator, while the hospital rent is maintained at zero. The regulator can choose a point on this dotted line where the marginal utility of cost reduction is equal to the marginal cost of transfers to the hospital plus the quality deterioration that is associated with extracting higher effort levels from the hospital. The

implementation of the incentive scheme is as simple as in the previous section, at least in theory. The regulator adjusts the slope of the incentive scheme so that it is equal to the right hand side of (20).

2.6 Moral hazard and adverse selection

My aim here is to identify some of the effects on the optimal incentive schemes when we have both moral hazard and adverse selection. The basic model used to clarify these problems is based on theoretical foundations laid by Laffont & Tirole[23] and my previous derivation of the optimal regulatory scheme when adverse selection is not a problem. The cost function is the same as in the previous section, i. e. $C = \beta + s - e$.

My aim is to maximise expected utility of the regulator, under the restriction that physicians must not sabotage the regulators' aim of reaching a certain cost level by adjusting the quality in health care. The reasoning is that if the hospital is given a certain contract, physicians must not be tempted to decrease quality, thereby causing the hospital to decrease effort and increase rent. Another situation arises if doctors are tempted to increase quality. This will force the hospital to increase effort to remain at the contracted cost level. If the hospital anticipates this problem before contracting with the regulator, it would like to try to masquerade as less efficient than it really is in order to get a more advantageous contract. The only solution to this problem is to make sure that the contract offered to the hospital is compatible with the utility-maximising behaviour of the physicians.

The utility function of the physicians is, as before, written as:

---

[23] See review problem 3, p. 673 in Laffont and Tirole (1993)

$$u_D = R + \gamma_1\left[U_H - U_H^*(\beta)\right] + \gamma_2 B + \gamma\mathfrak{n} - \tau. \tag{21}$$

Solving for U_H we get:

$$U_H = t - \frac{\tau - R - \gamma\mathfrak{n}}{\gamma_1} + \alpha B(s) + U_H^*(\beta). \tag{22}$$

Since the regulator now maximises the expected utility over the different types of hospitals, we get the maximisation problem

$$\max_{e(.),U(.),\xi,s} \int_{\underline{\beta}}^{\bar{\beta}} \left[\begin{array}{l} [(1+\alpha\lambda)B(s) - (1+\lambda)(\beta - e(\beta)) \\ +s + \psi(e(\beta)) + \varpi) - \lambda U_H^*(\beta)) \end{array} \right] f(\beta) + \xi(\alpha B'(s) - \psi'(e)] d\beta \tag{23}$$

s. t.

$$\dot{U}_H^*(\beta) = -\psi'(e(\beta)),\ \dot{C}(\beta) \geq 0,$$

where U^*_H is the rent that accrues to the hospital in order to satisfy the incentive compatibility constraint. A high β means that optimal effort must be low. A low effort is connected with a low marginal disutility of effort, since $\psi'(e)>0$. This means that marginal benefit of quality must be low and therefore that quality must be high, by the equality $\psi'(e)=\alpha B'(s)$. The conditions for optimum are (derivation, see Appendix 2)

$$\psi'(e) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e) - \frac{\xi}{f(\beta)(1+\lambda)}\psi''(e) \tag{24}$$

and

$$\alpha B'(s) = \frac{\alpha((1+\lambda) - \xi\alpha B''(s)/f(\beta))}{(1+\lambda\alpha)}. \tag{25}$$

The size of the Lagrangian multiplier shows to what extent quality is a substitute for effort. We can also see that the optimal incentive scheme is

flatter for all but the most efficient type of hospital. Just as in the usual Laffont-Tirole model, the optimal regulatory system is to offer hospitals a menu of linear incentive schemes.

The optimal incentive schemes will also be flatter than when investments in quality are not feasible. Comparing marginal costs and benefits of increasing effort we can see this.

$(1+\lambda)[1-\psi'(e(\beta))]\Delta Ef(\beta)d\beta$ is the social gain from increasing effort one unit for types in $[\beta,\ \beta+d\beta]$ and $F(\beta)\lambda\psi''(e(\beta))\Delta Ed\beta$ is the social cost of giving rent to types more efficient than β. $\xi\psi''(e(\beta))\Delta Ed\beta$ is the social cost of lowering quality (which is necessary when effort is increased). When the optimal incentive scheme is applied, the following equality must hold

$$(1+\lambda)[1-\psi'(e(\beta))]\Delta Ef(\beta)d\beta = F(\beta)\lambda\psi''(e(\beta))\Delta Ed\beta+\xi\psi''(e(\beta))\Delta Ed\beta. \quad (26)$$

It is easy to see that optimal effort is lower when investment in quality is feasible. This means that incentive schemes must be flatter, since $\psi'(e)>0$. The equations (19) and (20) are the limits of expressions (24) and (25) as adverse selection disappears (see Appendix 3).

3 Interpretation

3.1 The optimal reimbursement system

One of the conclusions from the model is that putting the entire marginal responsibility for cost overruns on the hospital often is a less than optimal strategy in the regulation of health care production. When steep incentive schemes are used, the physicians will be put under great pressure from the hospital management to reduce the costs of the hospital by reducing the level of quality in health care. This conclusion is valid at least as long as the physicians have control over the financial resources in the hospital and is less than perfect agents in the provision of health care.[24]

It is also my conclusion that the optimal contractual form will be somewhere in-between cost-plus and fixed reimbursement and that the existence of adverse selection makes different incentive schemes for different hospitals the optimal solution. The problem is, though, that while prospective payment in the DRG system, where prospective payments are used to reimburse hospitals, is simple and easy to understand, a prospective payment where both achieved quality and reimbursement schemes differ between hospitals, is rather difficult to operationalise. It is also not an advantage that the variables used in the optimal regulatory scheme are not known with certainty. Even if the marginal substitution between rent of hospitals and patient benefit can perhaps be estimated, it would still be necessary to estimate the adverse

[24] It is also possible to imagine a situation where physicians are more than perfect agents for patients, e.g. $\alpha > 1$. This is, however, not very likely.

selection parameter and the marginal disutility of effort function. Neither of these variables can be directly observed.

In situations where quality observation problems can be expected to be severe, we can assume that the physicians marginal rate of substitution between hospital rent and patient benefit is low (low α parameter). This is natural, since the feedback to the physicians will be less powerful if quality is hard to measure. If this problem is combined with a severe adverse selection problem, it might be sub optimal to use systems like DRGs. If moral hazard is the more serious problem, on the other hand, steeper incentive schemes are optimal. If these problems exist, we can say something about, not only the effect of each problem, but also the interaction between them as well as their combined effects and the optimal regulation to be applied.

3.2 Regulation in practice

The regulation of hospital services has taken many different routes, with both demand side and supply side regulation. On the supply side, the prospective payment system of Medicare is perhaps the best known. However, there are many other examples, more or less well functioning. There are not many systems with a mixed reimbursement system, like the one derived in my paper. In California, on the other hand, it has been observed that reimbursements for medical services correspond to the theoretically optimal incentive schemes under adverse selection. Vistnes (1994) investigated contracts between California's Medicaid program and hospitals and found empirical support for the assertion that high-cost hospitals receive contracts closer to cost reimbursement and that low cost hospitals get fixed price contracts.

It has, however, been observed that the reimbursement methods for different types of health care differ a lot. If this is due to political or economic considerations remains to be investigated. In the U.S. Medicare system, psychiatric hospitals, rehabilitation hospitals; psychiatric or rehabilitation units which are distinct parts of a hospital, alcohol or drug-abuse hospitals, children's hospitals, long term hospitals and hospitals outside the 50 states and the district of Columbia (Green-book, 1995) are provided with a cost limit. Their costs are reimbursed up to a certain limit. Costs above that level are reimbursed at a rate of 50 percent, up to a maximum of 110 percent of the limit. It is perhaps not a coincidence that these rules apply for non-surgical hospitals, where quality is hard to measure and results are loosely connected to inputs. It can also be observed that it is more common to use capitation payments in out-patient health care. This would also be in line with my model, since economic incentives for treatment would not be as effective as in in-patient care.

The hypothesis implied by the model can be tested against empirical data and the behaviour of regulators. The prospect of regulating health care from the supply side, must however be weighted against other forms of regulation, like capitation, fee for service, line item budgeting, global budgets and fixed fee schedules. The main advantages of supply side regulation, with prospective payments, can perhaps be found within medical and surgical health care. As similar systems are applied to other types of health care, like the proposed application of DRG to mental care in Australia (MH–CASC, 1995), we will be able to evaluate the results to draw more general conclusions.[25]

---

[25] For a more complete overview of the different reimbursement models and their effects on costs and outcomes, see Zweifel & Breyer (1996), chapter 9.

4 Conclusion

We can conclude that problems of adverse selection, moral hazard and imperfect agency interact to make the problems in regulation of hospital services complex. The problem of adverse selection has partly been handled by dividing patients into different diagnosis-related groups, DRGs. Many problems remain, though. If physicians are not regarded as perfect agents for patients, or the society at large, it will still not be optimal to use steep incentive schemes for any type of hospital.

If adverse selection is present, the optimal regulatory model is a menu of incentive schemes whose steepness is mitigated by the concern for quality in health care. It might not be obvious that the linear menu of incentive schemes is maximising social utility. Once it is recognised that the deterioration of quality implied for the most efficient hospitals, is the price that has to be paid to make it possible to extract rents and making the hospital managers exert effort, this is more understandable.

What makes the model used here different from others on hospital regulation, is the result that incentive schemes should be decided by the trade-off between effort inducement, rent extraction and quality in provision of health care services. My results are more general than the conclusions of Ellis & McGuire, who assert that incentive schemes should be adjusted by a factor that depends on the degree of perfection in the agency relationship between the doctor and the patient. If moral hazard is a problem, we see that it is rational to use steeper incentive schemes than the ones prescribed by the Ellis & McGuire model. On the other hand, if adverse selection rears its ugly head, my conclusion is that the steepness of incentive schemes should in general be reduced.

Appendix

A1

Derivation of the optimal incentive scheme when investment in quality is not feasible (Laffont & Tirole, 1993, p 67).

Let U be the state variable and e the control variable. The Hamiltonian is

$$H = (S - (1+\lambda)[\beta - e(\beta) + \psi(e(\beta))] - \lambda U(\beta)f(\beta) - \mu(\beta)\psi'(e(\beta)), \qquad (1)$$

where $\mu(.)$ is the Pontryagin multiplier. By the maximum principle

$$\dot{\mu} = -\frac{\partial H}{\partial U} = \lambda f(\beta). \qquad (2)$$

The boundary $\beta = \underline{\beta}$ is unconstrained. Hence the transversality condition at $\beta = \underline{\beta}$ is $\mu(\underline{\beta}) = 0$. Integrating (2) yields $\mu(\beta) = \lambda F(\beta)$. Last, maximising H with respect to e gives

$$\psi'(e(\beta)) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e(\beta)). \qquad (3)$$

A2

The derivation of the optimal regulatory scheme when the doctor is a double agent.

The maximisation problem is

$$\max_{e(.),U_H^*(.),\xi,s} \int_{\underline{\beta}}^{\bar{\beta}} \left[\begin{array}{l} [(1+\alpha\lambda)B(s)-(1+\lambda)(\beta- \\ e(\beta)+s+\psi(e(\beta))+\varpi)-\lambda U_H^* \end{array} \right] f(\beta)+\xi(\alpha B'(s)-\psi'(e(\beta))]d\beta \quad (1)$$

s. t.

$$\dot{U}_H^*(\beta) = -\psi'(e(\beta)),$$
$$U_H^*(\bar{\beta}) = 0,$$
$$\alpha B'(s) - \psi'(e > (\beta)) = 0,$$
$$\dot{C}(\beta) \geq 0$$

This problem can be solved by optimal control and by ignoring the monotonicity constraint we get

$$H = [(1+\alpha\lambda)B(s)-(1+\lambda)(\beta - e(\beta)+s+\psi(e(\beta))+\varpi)-\lambda U_H^*)]f(\beta) \quad (2)$$
$$-\mu(\beta)\psi'(e(\beta))+\xi(\alpha B'(s)-\psi'(e(\beta)))$$

where $\mu(.)$ is a Pontryagin multiplier and ξ is a Lagrange multiplier. The Lagrangian gives us the first order conditions:

$$\frac{\partial H}{\partial U_H^*} = -\dot{\mu}(\beta) \Rightarrow \dot{\mu}(\beta) = \lambda f(\beta), \quad (3)$$

$$\frac{\partial H}{\partial e} = -(1+\lambda)(-1+\psi'(e(\beta)))f(\beta) - \mu(\beta)\psi''(e(\beta)) - \xi\psi''(e(\beta)), \quad (4)$$

$$\frac{\partial H}{\partial \xi} = 0 \Rightarrow \alpha B'(s) = \psi'(e(\beta)), \quad (5)$$

$$\frac{\partial H}{\partial s} = 0 \Rightarrow 0 = (-(1+\lambda)+(1+\alpha\lambda)B'(s)))f(\beta)+\xi\alpha B''(s), \quad (6)$$

$$\dot{U}_H^*(\beta) = \frac{\partial H}{\partial \mu} = -\psi'(e(\beta)). \quad (7)$$

Transversality condition at $\beta = \underline{\beta}$ is $\mu(\underline{\beta}) = 0$. (8)

Integrating (3) and using (8) implies $\mu(\beta) = \lambda F(\beta)$. Substituting this into (4) gives us

$$\psi'(e) = 1 - \frac{\lambda}{1+\lambda}\frac{F(\beta)}{f(\beta)}\psi''(e) - \frac{\xi}{f(\beta)(1+\lambda)}\psi''(e). \tag{9}$$

From (6)

$$(1+\alpha\lambda)B'(s) - (1+\lambda) = -\frac{\xi\alpha B''(s)}{f(\beta)}. \tag{10}$$

A3

By replacing $\frac{\xi}{f(\beta)}$ by ξ in equations 24 and 25 we see that equations 19 and 20 is the limit of equations 24 and 25 as $\overline{\beta} - \underline{\beta} \Rightarrow 0$ and $f(\beta) \Rightarrow \infty$, since the Lagrange multipliers in equations 24 and 25 will be proportional to $f(\beta)$ as it increases. The transversality condition $\mu(\beta)=0$ and equation A2 (4) gives us this proportionality and since $\overline{\beta} = \underline{\beta}$ in the limit it can be seen that the expressions 19 and 20 will be the same as 24 and 25.

A4

Participation constraints for the maximisation problems:

No moral hazard

The condition is that the maximisation problem

$$\begin{aligned} &\underset{s}{\text{Max}}\ U_H - (1+\lambda)(t+C) + B(s) - \theta(s) \\ &\text{s. t. } U_H^* = 0\ ,\ u_D = 0\ ,\ U_H^* = U_H\ ,\ \theta(s) = \alpha B(s) \end{aligned} \tag{1}$$

is non-negative at the optimal quality level. This means that

$$(1+\alpha\lambda)B(s^*)-(1+\lambda)(\beta+s^*+\varpi)\geq 0. \qquad (2)$$

Moral-hazard

$$(1+\alpha\lambda)B(s^*)-(1+\lambda)(\beta+s^*-e^*+\psi(e^*)+\varpi)\geq 0 \qquad (3)$$

Moral-hazard and adverse selection

For the least efficient hospital to be considered as a producer, the following condition must hold

$$(1+\alpha\lambda)B(s^*)-(1+\lambda)(\overline{\beta}-e^*+s^*+\psi(e^*)+\varpi)\geq 0. \qquad (4)$$

Only hospitals that fulfil this condition will be considered, since only those hospitals would contribute to social utility. The distribution will otherwise be truncated, so that the least efficient type of hospital will just barely fulfil the condition. For an explanation of this, see Laffont & Tirole (1993), pp. 73). The difference here with their model is that I do not have a constant utility for the project, but instead a benefit-function of quality *B(s)*. This does not, however, alter the basic conclusions made.

– Part Three –

Non-linear Incentives in Not-for-profit Hospitals

1 Introduction

A common feature of non-profit organisations is that their revenue often come from voluntary and state contributions (Weisbrod, 1977, p. 173). This means that their revenue is only indirectly related to their production. The state or other contributors will first have to observe the accomplishments of the organisation and then decide if they want to contribute to it. When output is difficult to observe, it is hard for the contributors to know if their contributions are used in a productive manner. The non-existence of payment mechanisms closely correlated with observable aspects of the goals of an organisation is, most likely, an important reason for its non-profit status. Easley & O'Hara (1983) suggest that one reason for the existence of non-profit firms is that output is hard to observe. Pure monetary transfers from the regulated firm to its owners could in that sense be a threat to quality and thus indirectly also to the welfare of society.

What are then the guarantees that the not-for-profit organisation will pay any attention to the goals of the contributors? One explanation could perhaps be that ethical principles among the employees of the organisation will make certain that resources are used to promote the interests of the contributors. Within the hospital sector, these ethical principles are often assumed to be defended by physicians. There is a strong ethical code within the profession. The Hippocratic oath is one manifestation of this. In economic theory this idea is conveyed by the proponents of the "agency theories", like Ellis & McGuire (1986, 1990).

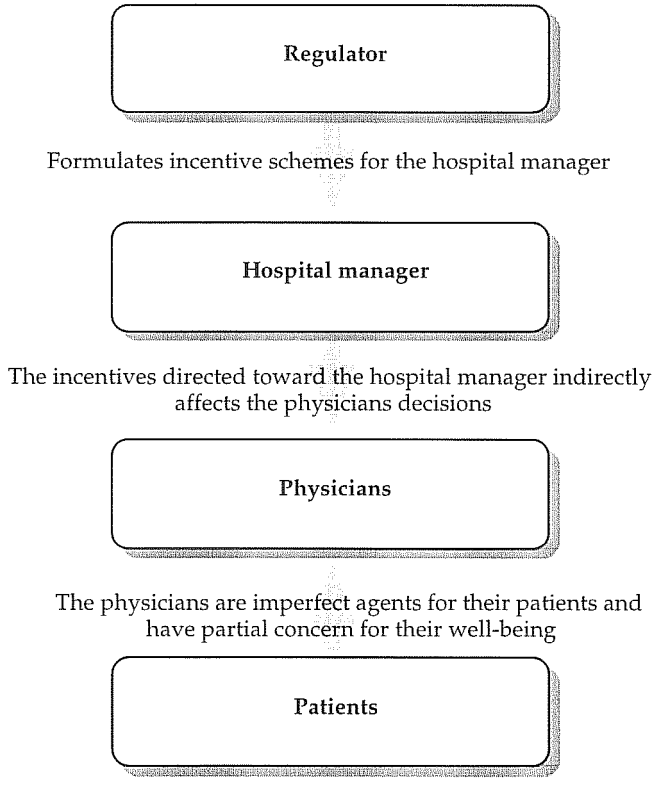
The agency models in health economics are based on the idea that physicians are acting on behalf of patients. This has implications for the division of responsibility for the quality at the hospital. If the physicians

have the prime responsibility for the quality there might be negative consequences from putting financial pressure on the physicians, since that might induce them to pay less attention to the needs of the patients and more to the profitability of the hospital. The model presented in this paper is based on the agency theory and is an attempt to describe the behaviour of hospitals under the assumption of a large discretion for physicians. I use a double agency model to describe both the behaviour of not-for-profit (NFP) and for-profit (FP) hospitals and their reaction to prospective (fixed fee) reimbursement schemes. The model is also applicable when there are intrinsic differences in efficiency across hospitals. It is shown that FP hospitals preserve the same quality level, regardless of their intrinsic efficiency differences. NFP hospitals, in contrast, are more likely to use the quality level to regulate costs. These effects origin from the non-linear incentives for the managers in the hospitals. In FP hospitals efficiency differences are mirrored by the hospitals proceedings, but in NFP hospitals these differences are instead reflected in quality variance. I also expand the model to deal with a milieu of moral hazard on behalf of the hospital managers. The model show that there is a positive correlation between quality and slack in the hospital organisation, and that the ratio of these two is independent of the ownership status of the hospital. Two hospitals with the same cost level thus always uphold the same quality and efficiency level, independent of their ownership status. This inference has direct implications for the appropriateness of prospective reimbursements. A proprietary structure is consequently urgent, for at least some fraction of the hospital population, if marginal economic incentives is to be used as a regulatory tool. The alternative, auctioning and purchaser-provider models, put a lot more strain on the regulators, who then will have to get involved in quality assessment and other time consuming activities.

2 The model

The model is based on the following assumptions. The regulator formulates incentives for the manager of the hospital. The hospital manager is assumed to be risk neutral.

Figure 1 The decision hierarchy



The physicians in the hospital are acting independently of the hospital managers, but their behaviour is indirectly affected by the incentives directed toward the manager, because the rent of the manager is

included in the utility function of the physician.The physicians are also acting as agents for their patients. The set-up is described in Figure 1.

2.1 Quality

The hospital's cost can be expressed as

$$C = \beta + s \tag{5}$$

where β represents an exogenous fixed cost and s is the hospitals investment in quality. As in Part two of this thesis, the physicians are assumed to be responsible for investments in quality.

The utility of the hospital manager is

$$U_H = \tau - C \tag{6}$$

where τ is the gross transfer from the regulator to the hospital.

The benefits of patients is a function of quality that satisfies the Inada conditions, that is

$B(0) = 0$
$B'(0) = \infty$
$B' > 0 \wedge s \geq 0$
$B'' < 0$

The physicians utility is expressed as a function of the rent of managers and the benefit of the patients, according to the agency theory by Ellis & McGuire (1986).

$$u_D = u_D(U_H, B(s)) \tag{7}$$

The first order condition for the physicians utility maximisation with respect to quality, s, can be expressed as

$$\frac{\partial u_D}{\partial U_H}\frac{\partial U_H}{\partial s}+\frac{\partial u_D}{\partial B}B'(s)=0. \tag{8}$$

We will now look at the effects of a fixed prospective payment on the quality level. The variable τ is the gross transfer from the regulator to the hospital and is defined as (from (6))

$$\tau = C + U_H. \tag{9}$$

The physicians utility can under entirely prospective reimbursements then be written as

$$u_D = u_D(\tau - \beta - s, B(s)). \tag{10}$$

Differentiating with respect to s gives the first order condition

$$-\frac{\partial u_D}{\partial U_H}+\frac{\partial u_D}{\partial B}B'(s)=0. \tag{11}$$

Dividing by $\frac{\partial u_D}{\partial U_H}$ yields

$$\alpha B'(s)=1 \tag{12}$$

where

$$\alpha = \frac{\partial u_D}{\partial B}\Big/\frac{\partial u_D}{\partial U_H}. \tag{13}$$

This is the same expression as derived by Ellis & McGuire in 1986. The marginal social benefit of quality, $B'(s)$, is thus more than one. This means that investment in quality is less than the socially optimal level if the benefit function is defined as the social benefits in monetary terms and the cost of quality, s, also is defined in monetary units.

2.2 Ratchet effects

The ratchet-principle is based on setting the cost-target of later periods as a function of the profit made in the present period. Weitzman (1980) shows that the effect of ratchet principles can be captured in a single variable, the ratchet price. In fact, hospitals are not always "for-profit" and can be subject to large ratchet effects. This makes it interesting to study what happens if the hospitals rent target is made dependent on the previous results. The ratchet price converts the dynamic optimisation problem under ratchet effects into a single period maximisation problem. The ratchet price is a function of two variables; the discount rate between two periods and the change in the target as a function of the difference between the target and actual performance. It is shown that the firm behaves optimally under the influence of the ratchet effect, when maximising utility in a one period framework and subjected to the ratchet price. The ratchet price thus makes the dynamic problem into a static, one period, utility maximisation problem.

It is easy to see the effects of such a reimbursement system in the present model. Returning to the physicians utility function, we see that such a change would make the utility function look like

$$u_D = u_D((1-\phi)(\tau-\beta-s), B(s)) \tag{14}$$

where $1-\phi$ ($\phi \in (0,1]$) is the ratchet price (or index) per cost unit. Maximising this with respect to s, the first order condition for the physician is then

$$1-\phi = \alpha B'(s) \tag{15}$$

where α is defined as above and there exists an interior solution.

The ratchet effect makes incentives less steep, thus appropriate implementation of ratchet models of regulation could be used as a route to finding an optimal quality level, without using explicit cost sharing. Significantly, although ratchet effects have traditionally been seen as a problem in the regulatory literature, this analysis suggests that they can also function as a policy instrument.

2.3 Adverse selection

It is interesting to know how the behaviour of hospitals will be effected if there are cost differences between hospitals. If we assume that hospitals are subjected to a break-even constraint, the incentives for the managers to decrease costs will be very strong as the hospital approaches this limit.

The intrinsic efficiency of the hospitals is assumed to be distributed $f(\beta)$ from $\underline{\beta}$ to $\overline{\beta}$, and the least efficient hospital is assumed to just break even. The model precludes a lower quality level than zero. This can be seen by noting that the marginal benefit of increasing quality approaches infinity as quality approaches zero, according to the Inada conditions assumed to hold for the benefit function of quality.

For the least efficient hospital in operation, the following condition holds

$$\tau - \overline{\beta} = 0 \tag{16}$$

As quality approaches zero, the marginal benefit of quality approaches infinity. If the hospital manager must have a rent level of zero to be willing to continue the operations, no greater cost reduction is conceivable. The maximisation problem for the physician is thus

$$\underset{s}{Max}\, u_D((1-\phi)(\tau-\beta-s), B(s)) \tag{17}$$

s. t.

$$\tau-\beta-s \geq 0$$

The Lagrangian maximisation problem is

$$\max_{s,\lambda} \mathcal{L} = u_D((1-\phi)(\tau-\beta-s), B(s)) + \lambda(\tau-\beta-s) \tag{18}$$

where λ is the langrange multiplier. The Kuhn-Tucker conditions are

$$-(1-\phi)\frac{\partial u_D}{\partial U_H} + \frac{\partial u_D}{\partial B} B'(s) - \lambda = 0 \tag{19}$$

$$\tau-\beta-s \geq 0$$

$$\lambda^* \geq 0$$

$$\lambda^*[\tau-\beta-s] = 0$$

If the ratchet effect is close to one, we see that λ will be larger than zero, the restriction binding and that quality thus will be

$$s = \tau-\beta. \tag{20}$$

The least efficient hospital will have s=0. There is a direct linear relationship between the intrinsic efficiency level of the hospital and its quality level.

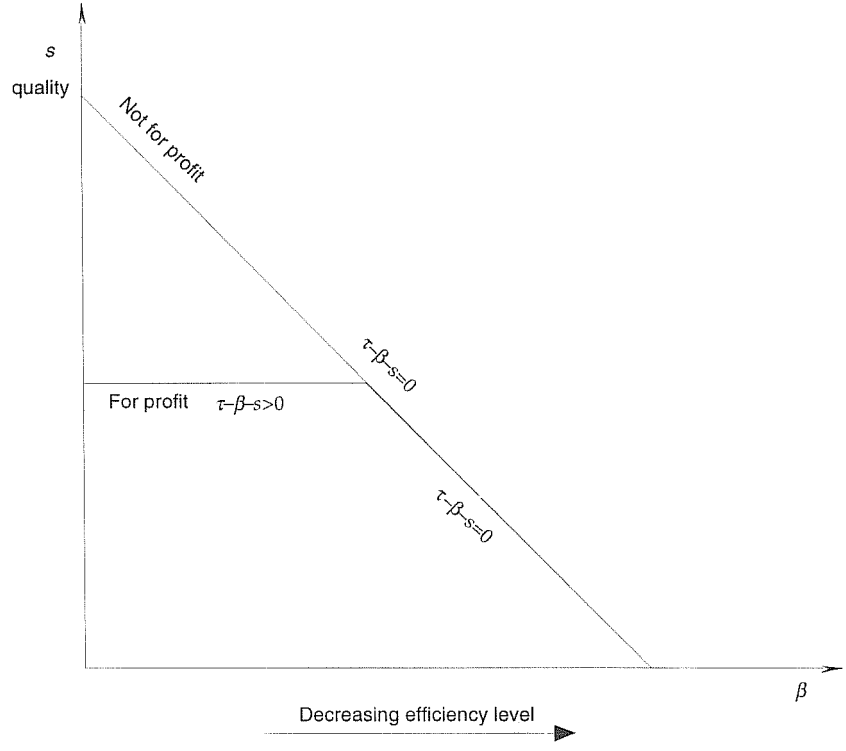
This can be contrasted with a for-profit hospital. Assuming that the main objective of the FP hospital is to make a financial surplus, we can rule out the situations where the restriction is binding. This means that the least efficient hospital will just break even. Above that limit, hospital managers are assumed to be given incentives linear in achieved costs.

When the ratchet effect is zero and the restriction is not binding, the quality level will be given by

$$B'(s) = \frac{\partial u_D}{\partial U_H} \Big/ \frac{\partial u_D}{\partial B} \qquad (21)$$

We illustrate this in Figure 2.

Figure 2 **The quality level of for-profit and not-for-profit hospitals**



For-profit hospitals have a segment where the quality level is constant. It could, as noted, be argued that the downward sloping segment of the for-profit hospital will never be relevant, since the existence of those

hospitals presume that physicians show a special constraint in their work. If there are intrinsic investments made by the for-profit hospital, the physicians would be likely to expropriate these investments by their choice of quality. This would make it unprofitable for FP hospitals to enter into the market. Only the leftmost part of the schedule would thus be at issue for the FP units.

2.4 Moral hazard

The cost function for the hospital is here defined as

$$C = \beta + s - e, \tag{22}$$

where e is the managers effort to reduce the cost of the hospital. The physicians choice of quality and the managers choice of effort is characterised as a Stackelberg equilibrium, where physicians make their choice of quality before the managers choose effort.

The managers utility can be expressed as

$$\begin{aligned} U_H &= (1-\phi)(\tau - C) - \psi(e) \;\forall\; \{\tau - C\} \geq 0 \\ U_H &= -\infty \;\forall\; \{\tau - C\} < 0 \end{aligned} \tag{23}$$

where ψ(e) is the managers disutility of effort ($\psi'>0$, $\psi''>0$). His rent is thus equal to the Langrangian

$$\mathcal{L} = (1-\phi)(\tau - \beta - s + e) - \psi(e) + \lambda(\tau - \beta - s + e). \tag{24}$$

Maximising this with respect to e and λ, and assuming that the physician chooses quality independently, gives us the Kuhn-Tucker conditions

$$\psi'(e) = (1-\phi) + \lambda \qquad (25)$$
$$\tau - \beta - s \geq 0$$
$$\lambda^* \geq 0$$
$$\lambda^*[\tau - \beta - s + e] = 0$$

The physicians utility is, as before

$$u_D = u_D(U_H, B(s)), \qquad (26)$$

where U_H is equal to the langrangian

$$U_H = (1-\phi)(\tau - \beta - s + e) - \psi(e) + \lambda(\tau - \beta - s + e). \qquad (27)$$

His maximisation problem is

$$\underset{s}{Max}\, u_D(U_H, B(s)), \qquad (28)$$

which yields the first order conditions

$$\frac{\partial u_D}{\partial U_H}\left[(1-\phi) + \lambda^*\right] = \frac{\partial u_D}{\partial B} B'(s), \qquad (29)$$

where λ^* is given by (25). This can be simplified as

$$\alpha B'(s) = (1-\phi) + \lambda^*. \qquad (30)$$

If the langrange multiplier is zero, this is the same first order condition as in the situation without moral hazard. We now assume that the hospital is not allowed to have financial surplus. This is equivalent to assuming that ϕ is one and consequently that λ is greater than zero and the restriction binding. The physicians will then choose quality to extract the hospitals resources and the limit will be where the marginal disutility for the manager of increasing effort is equal to the value of the loss of rent for the manager in the physicians utility function.

Rewriting the first order condition when λ is greater than zero and using the restriction, gives

$$\psi'(\beta + s - \tau) = \alpha B'(s). \tag{31}$$

If the physicians increase quality in this situation, the manager will have to increase effort to keep the hospital at a break even level. The equilibrium occurs where the negative marginal effect on the hospital manager from this forced increase in effort is equal to the marginal utility from quality in the physician's utility function. Low cost hospitals will exert little effort to reduce costs, while high cost hospitals will exert high effort. High levels of cost (the β parameter) will be accompanied by low levels of quality and high effort to reduce inefficiencies by the hospital managers. Low cost hospitals will have a high level of quality and a low effort from managers to reduce costs. A regulator might want to use the prospective payment as a way of controlling both the level of quality as well as giving incentives to reduce costs. These two targets has to be determined by weighting their respective adequacy and using the prospective payment as the only regulatory instrument. For a FP hospital, in contrast, there will be a limit to the investment in quality and slack, and the hospital will make a profit if the gross reimbursement is larger than the equilibrium cost level. The difference between the gross reimbursement and this cost level will then turn out as profit for the hospital. If the equilibrium cost level is higher than the reimbursement given, the for-profit hospital will behave like the not-for-profit hospitals. For reasons discussed earlierly, it is doubtful if the FP hospitals would enter a market where this restriction is expected to be binding. However, the discussion about that is beyond the scope of this paper.

3 Interpretation

When prospective reimbursements have been introduced in the health care system, yardstick competition between hospitals has been assumed to certify that production is carried out as efficiently as possible. Lack of profitability can then be seen as a sign that the services provided by a hospital is worth less for society than the cost of providing the services. However, this idea builds on a notion that all hospitals will be interested in making as large a surplus as possible. In fact, making a FP hospital enter the market presupposes some opportunities for profit making. NFP hospitals are often financed by the state or by donations. The entry decision does not rely on the opportunity to make profits.

While the FP/NFP status have implications for the entry decision, it also has implications for the effectiveness of different regulation models. The assertion that hospitals will react to different cost sharing arrangements builds on a notion that managers always have incentives to reduce the costs of the hospital.

The model presented in this paper show, however, that this is not true. There are no immediate incentives to go beyond the level of break even in cost reduction for NFP hospitals. The cost level of a NFP hospital will thus only reflect the size of the reimbursements to the hospital, and not its potential efficiency. Prospective payments, in a system of only NFP hospitals, will then be ineffective as a way of enhancing productivity. However, if there is some fraction of for-profit hospitals in the market, the costs of these hospitals can be used as a benchmark and a population of yardstick competing hospitals. The prospective payments calculated from this population can then be used to reimburse NFP units as well. The model does not predict any differences in

productivity between FP and NFP units, for a given level of quality. This result differs from models based on competitive forces, such as Ma (1994) or Hodgkin & McGuire (1994).

In countries, such as Sweden, where the fraction of FP units is negligible, reimbursement systems of the type used for Medicare in the United States would be impractical as a regulatory device. Regulators would have to resort to other forms of regulatory means, such as purchaser-provider models or global price caps for hospitals. These forms of regulatory systems take more administrative effort from regulators, since the regulators (often politicians) in practice will have to set reimbursement levels directly, or use auctioning systems, instead of relying on profit levels as signals to whether a hospital should remain in business or not. If the regulator's informational capacity is limited, and an equal quality across hospitals is desirable, a for-profit structure would be preferred, since the effect of marginal economic incentives then will be independent of the intrinsic efficiency levels of the regulated hospitals.

This model shows that prospective reimbursements, when applied to NFPs, could cause a diversion in the quality level of different hospitals when their costs are different. For FP units, on the other hand, the quality level is not affected by efficiency differences to the same degree. Their quality-level is determined by marginal economic incentives. It is also shown that when hospital managers can exert effort, the hospital will trade off quality against effort and the choice of quality will be accompanied by a specific choice of efficiency enhancement. A reduction in reimbursements will affect quality negatively and effort positively. High quality is thus always accompanied by a lot of slack in the organisation.

Another interesting aspect is that, given the institutional structure with personal responsibility for physicians and regardless of ownership, two hospitals with the same cost level will always have the same ratio between efficiency and quality. The choice of ownership status for hospitals is then only to be determined by the preferred regulatory model. To use marginal economic incentives, through cost sharing and prospective payments, is a natural method to regulate for-profit hospitals, while auctioning and purchaser-provider models are more suitable in the regulation of not-for-profit hospitals.

There are many differences between this model and earlier models of the behaviour of NFPs. This is not to say that they are wrong or that this model is the only correct one, but only to point out that the assumptions about the origin of incentives for quality maintenance makes a difference. Assumptions about an agency role for the physician make the concern for quality more of an internal question in the hospital. Laffont & Tirole (1993), Hodgkin & McGuire (1994) and Ma (1994) assume that demand in some way indirectly determines the managers choice of quality. Here it is shown that there is no need for assumptions about the patients demand for quality, to derive a model describing the choice of quality and efficiency by a hospital.

References

Akerlof G. A., 1970, The market for 'lemons': quality uncertainty and the market mechanism, *Quarterly Journal of Economics*, Vol. 84, No. 3, 488–500

Anderson R. K., D. House & M. B. Ormiston, 1981, A theory of physician behaviour with supplier induced demand, *Southern Economic Journal*, Vol. 48, No. 1, 124–133

Arrow K. A., 1951, An extension of the basic theorems of classical welfare economics, In: *Readings in Mathematical Economics*, Ed. P. Newman, John Hopkins Press, Baltimore

Arrow K. A., 1963, Uncertainty and the welfare economics of medical care, *American Economic Review*, Vol. 53, No. 5, 941–973

Ballard C., J. Shoven & J. Whalley, 1985, General equilibrium computations of the marginal welfare costs of taxes in the United States, *American Economic Review*, Vol. 75, No.1, 128–138

Barnum H., J. Kutzin & H. Saxenian, 1995, Incentives and provider payment methods, *Department of Human Capital and Operations Policy at The World Bank*, Working paper, No. 51

Blomqvist Å., 1991, The doctor as a double agent: Information asymmetry, health insurance, and medical care, *Journal of Health Economics*, Vol. 10, No. 4, 411–432

Chalkley M. & J. M. Malcomson, 1995, Contracting for health services with unmonitored quality, Discussion paper series No. 9510, *Department of Economics at the University of Southampton*

Coase R. A., 1960, The problem of social cost, *Journal of Law and Economics*, Vol. 3, Oktober, 1–44

Dada M., W. D. White, H. H. Stokes & P. Kurzeja, 1992, Prospective payment for psychiatric services, *Journal of Health Politics, Policy and Law*, Vol. 17, No. 3, 483–508

Debreu G., 1959, *The Theory of Value*, New York, Wiley

Easley D. & M. O'Hara, 1983, The economic role of the non–profit firm, *Bell Journal of Economics*, Vol. 14, No. 2, 531–538

Ellis P. S. & T. G. McGuire, 1986, Provider behaviour under prospective reimbursement, *Journal of Health Economics*, Vol. 5, No. 2, 129–151

Ellis P. S. & T. G. McGuire, 1990, Optimal payment systems for health services, *Journal of Health Economics*, Vol. 9, No. 4, 375–396

Enthoven A. C., 1980, *Health Plan: The only practical solution to the soaring costs of medical care*, Addison—Wesley, Reading

Fetter R. B., D. A. Brand & D. Gamache, 1991, *DRGs – Their design and development*, Health Administration Press, Ann Arbor, Michigan

Frank R. G. & J. R. Lave, 1984, The effect of benefit design on length of stay of Medicare psychiatric patients, *NIMH Conference on the Economics of Mental Health*

Frank R. G. & J. R. Lave, 1986, Per case prospective payment for psychiatric in-patients: an assessment and alternatives, *Journal of Health Politics, Policy and Law*, Vol. 11, No. 1, 83–96

Glaser W. A., 1987, *Paying the Hospital*, Jossey–Bass Publishers, London

Hodgkin D. & T. G. McGuire , 1994, Payment levels and hospital response to prospective payments, *Journal of Health Economics*, Vol. 13, No. 1, 1–29

Holmström B., 1996, The firm as a sub–economy, Department of Economics, *Massachusetts Institute of Technology*, Paper presented in Beijing on 1–2 of September 1996

Laffont J–J., 1987, Toward a normative theory of incentive contracts between government and private firms, *Economic Journal*, (conference issue), 17–31

Laffont J–J. & D. Martimort, 1997, The firm as a multi-contract organisation, *Journal of Economics and Management Strategy*, Vol. 6, No. 2, 201–234

Laffont J–J. & J. Tirole, 1986, Using cost observation to regulate firms, *Journal of Political Economy*, Vol. 94, No. 3, 614–641

Laffont J–J. & J. Tirole, 1989, Provision of quality and power of incentive schemes in regulated industries, In: *Equilibrium Theory and Applications*, Ed. W. A. Barnett, Cambridge University Press, Cambridge, 161–193

Laffont J. & J. Tirole, 1993, *A Theory of Incentives in Procurement and Regulation*, The MIT press, Cambridge

Lundbäck M., 1997, Imperfect agency and the regulation of hospitals, *The Geneva Papers on Risk and Insurance Theory*, Vol. 22, December, 151–168

Ma C. A., 1994, Health care payment systems: Cost and quality incentives, *Journal of Economics and Management Strategy*, Vol. 3, No. 1, 93–112

McClellan M., 1997, Hospital reimbursement incentives: An empirical analysis, *Journal of Economics and Management Strategy*, Vol. 6, No. 1, 91–128

MH – CASC, 1995, Progress report – Initial consumer groupings, Mental health classification and service costs project, *Mental Health Research Institute*, Parkville

Mirrlees J., 1974, Notes on welfare economics, information and uncertainty, In: *Essays in Economic Behaviour under Uncertainty*, Eds. M. Balch and D. McFadden, North–Holland

Mirrlees J., 1976, The optimal structure of incentives and authority within an organisation, *Bell Journal of Economics*, Vol. 7, No. 1, 105–131

Musgrave R. A., 1959, *The Theory of Public Finance*, McGraw–Hill, New–York, Toronto, London

Newhouse, 1996, Reimbursing health plans and health providers: Efficiency in production versus selection, *Journal of Economic Literature*, Vol. 34, September, 1236–1263

Oswald S. L., L. R. Gardinger & J–J. Jahera Jr, 1994, Ownership effects on operating strategies: Evidence of expense-preference behaviour in the hospital industry, *Managerial and Decision Economics*, Vol. 15, No. 3, 235–244

Pope G. C., 1990, Using hospital specific costs to improve the fairness of prospective reimbursement, *Journal of Health Economics*, Vol. 9, No. 3, 237–251

ProPAC, 1992, *Medicare and the American Health Care System*: Report to the Congress, Washington D. C.

Reinhart U. E., 1987, A clarification of theories and evidence on supplier induced demand for physicians' services, *Journal of Human Resources*, Vol. 22, No. 4, 611–620

Ross S., 1973, The economic theory of agency: The principal's problem, *American Economic Review*, Vol. 63, No. 2, 134–139

Rupp A., M. Steinwachs & D. S. Salkever, 1984, The effect of hospital payment method on the pattern and cost of mental health, *Hospital and Community Psychiatry*, Vol. 35, No. 5, 460–469

Saltman R., 1985, Organisational and behavioural issues in the design of standardised protocols, *International Journal of Health Planning and Management*, outcast

Schleifer A., 1985, A theory of yardstick competition, *Rand Journal of Economics*, Vol. 16, No. 3, 1–41

Siegel C., J. Kristine, E. Laska, M. Meisner & S. Lin, 1992, A risk-based prospective payment system that integrates patient, hospital and national cost, *Journal of Health Economics*, Vol. 11, No. 1, 1–41

Sharfstein S. S., 1991, Prospective cost allocations for the chronic schizophrenic patient, *Schizophrenia Bulletin*, Vol. 17, No. 3, 395–400

Shepard D. S., T. Vian & E. F. Kleinau, 1990, Health insurance in Zaire, Policy Research and External Affairs Working Paper, No. 489, Africa Technical Department, *World Bank*, Washington D. C.

Stigler G. J., 1966, *The Theory of Price*, Macmillan, New–York

Stiglitz J. E., 1974, Incentives and risk sharing in sharecropping, *Review of Economic Studies*, Vol. 41, April, 219–255

Stiglitz J. E., 1975, Incentives, risk and information: notes toward a theory of hierarchy, *Bell Journal of Economics*, Vol. 6, No. 2, 552–579

Stiglitz J. E., 1987, Principal and agent, In: *The New Palgrave: A Dictionary of Economics*, eds. J. Eatwell, M. Milgate & P. Newman, Macmillan Press Ltd, London

U. S. Government Publications, 1995, 1994 GREEN BOOK: Overview of entitlement programs, *Committee on Ways and Means: U. S. House of representatives*, Appendix D, U. S. Government printing office, Washington

Vistnes G., 1994, An empirical investigation of procurement contract structures, *Rand Journal of Economics*, Vol. 25, No. 2, 215–241

Weisbrod B. A., 1977, *The Voluntary Non-profit Sector*, Lexington Books, Lexington, Massachusetts, Toronto

Weisbrod B. A., 1992, Productivity and incentives in the medical care sector, *Scandinavian Journal of Economics*, Vol. 94, Supplement, 131–145

Weitzman M. L., 1980, The "ratchet principle" and performance incentives, *Bell Journal of Economics*, Vol. 11, No. 1, 302–308

Williamson O. E., 1963, Managerial discretion and business behaviour, *American Economic Review*, Vol. 53, December, 1032–1057

World Bank, 1993, The organisation delivery and financing of health care in Brazil, Report 12655–BR, Latin America and the Caribbean Country Department, *Human Resources Division*, Washington D. C.

Zweifel P. & F. Breyer, 1996, *Health Economics*, Oxford University Press, Oxford

Prior reports published by CEFOS:

Behovsbudgetering. Nya budgetprinciper i kommuner. Cecilia Bokenstrand. 1/1993

Självständighet eller statsbundenhet. Den kommunideologiska idédebatten 1962-1974. Urban Strandberg. 2/1995

Finansieringsanalysens dimensioner. Teori och kostnader. Pär Falkman och Stefan Pauli. 3/1995

Utveckling och tillämpning av ekonomistyrning i en decentraliserad kommunal organisation. Problem och möjligheter vid förändringsarbete - en fallstudie om ekonomistyrning. Lennart Jansson. 4/1995

Risk Management and Health Care. Per-Johan Horgby. 5/1995

Skolchefers arbete. Om chefskap och styrning inom skolsektorn. Anna Cregård. 6/1996

Distributive Justice and Cooperation in Asymmetric Social Dilemmas. Daniel Eek. 7/1996

Kontakter och förankring. En studie av kommunstyrelsens ordförande. Petter Wrenne. 8/1997

Nära nyheter. Studier om kommunaljournalistik. Kent Asp, Bengt Johansson och Lars-Åke Larsson. 9/1997.