Paying the medical specialist: the eternal puzzle: experiments in the Netherlands

Mot, E.S.

Citation for published version (APA):
Mot, E. S. (2002). Paying the medical specialist: the eternal puzzle: experiments in the Netherlands

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
2 Theories on the effect of payment systems

2.1 Introduction

In Chapter 1, the main research question of this study was given: What is the effect of the experimental changes in the payment system on the diagnosis and treatment of patients by medical specialists? In Chapter 4, the changes that were made in the payment system, will be described in detail. Roughly speaking, the system changed from a fee-for-service (FFS) system to a ‘lump-sum budget’ system in which payment was much less variable. The financial incentive changed from a very detailed incentive, connected to individual services, to a much more aggregate incentive. Within a budget year, the budget was fixed and there was no financial incentive whatsoever from additional production. From year to year, additional production might give rise to a higher budget, depending on the precise financial arrangement and on consultation between the three parties on this point. In this chapter, theories and empirical findings on payment systems from the literature are described and hypotheses are deduced about the effect of the change in payment systems on the behaviour of specialists.

The different parties working in the health care sector or studying it have widely differing expectations and beliefs about the degree to which the payment system influences the behaviour of specialists. There are many health economists who regard it as a matter of course that the payment system does have a definite influence on behaviour. All proponents of the Physician Induced Demand (PID)–hypothesis believe that a physician can influence the demand of patients to further his or her own interests. Important advocates of this hypothesis are Evans, Fuchs, and Dranove. In such cases, the demand differs from what the patient would have demanded if he had known as much as the doctor. This means that a change in the payment system that changes the doctor’s optimal behaviour from his own point of view, makes the doctor influence demand. Not only would he want to do so, but he would also be able to do so according to the hypothesis.

9 In the following I want to avoid his/her and he/she constructions. I choose to refer to the specialist as a man since practically all specialists I met during the evaluation study were men, and it is possible that women have a somewhat different utility function (for example regarding the weights of different elements).
There has been an extensive debate about this not-so-easy-to-prove hypothesis (see, e.g., Phelps, 1997, about the extent of the debate). We will go into this briefly in the empirical section (2.3) of this chapter. In a recent review-article, McGuire (2000) states that there are theoretical reasons to believe that demand-inducement exists. On the empirical front too, he concludes: "A large body of credible research establishes that physicians set quantities, and they do so partly in response to self-interest."\(^\text{10}\) This does not necessarily mean that the PID-hypothesis is correct, since other quantity-setting mechanisms may be available to the doctor (see Section 2.2 below). But it does mean that a change in the payment system may be expected to have an effect on treatment.

Medical specialists, who naturally look upon the sector from a different perspective than the researchers, often state that only medical considerations play a role. At the start of the experiment, many medical specialists who took part in the experiment claimed that no change in the number of services would be found, since they were not influenced by financial considerations. Hospital management and the health insurers on the other hand, in many cases expected that the number of services would decrease because of the abolishment of financial incentives regarding the services.

Considering these differences of opinion, it is interesting to see in more detail what the literature on payment systems has to offer on this point. In Section 2.2, different theories from the literature on the effect of payment systems are described. In Section 2.3, the empirical findings of other researchers regarding the effect of different payment systems are described. In Section 2.4, the most appropriate theory is selected for analysing the Dutch experiment with payment systems for specialists, and a number of hypotheses are formulated.

2.2 Different theories from the literature

2.2.1 Introduction

There are three important aspects that determine the theoretical position of health economists. These concern their opinions about: (1) asymmetry of information between patients and doctors; (2) uncertainty about the best treatment; and

(3) elements of the doctor's utility function. Firstly, it is important to determine to what extent the theories of health economists are influenced by information-asymmetry between patients and doctors and possible ways for patients to reduce this asymmetry. The existence of information-asymmetry is very important to the functioning of the market for health care. Without it, there would be no opportunity for doctors to pursue their personal goals by influencing the patient's demand for their services. This does not necessarily mean that without asymmetry doctors would have no opportunity to influence quantity and/or price of health care used. McGuire (2000) points out that doctors also have this possibility with fully-informed patients under the conditions of monopolistic competition (where doctors have market power) and non-retradability of services (see, e.g., Farley (1986)). When patients are fully informed, but the quality of care is non-contractible, physicians can also influence the use of health care by determining the quality level. However, without information-asymmetry the doctor can not change the demand curve of the patient and there is no question of demand-inducement in the sense in which the concept is normally used.

Phelps (1997) sums up the consequences of information-asymmetry in one sentence: "Put most simply, the doctor might be able to deceive the patient and make more money doing so" (p. 7). It may be added that, instead of making more money, the doctor could also use his information advantage to have more leisure. In some circumstances, it is in his interest to increase demand, but he may also prefer to reduce demand (e.g. when he has a fixed salary).

Secondly, there are different opinions not only with regard to the uncertainty that doctors themselves experience when deciding what is the best treatment but also with regard to the consequences of this uncertainty.

The uncertainty of doctors has two aspects. First, one doctor may be better than another, in the sense of having more medical knowledge or experience concerning a certain subject. This aspect may be a determinant of the doctor's reputation. However, this aspect is not directly relevant for the analysis of the influence of the payment system, as will be explained in Section 2.2.2 next. Secondly, the uncertainties surrounding medical care may be such that even very good doctors do not always know what is the optimal course of action. This aspect is very relevant for the analysis of the influence of the payment system. When there is a great deal of uncertainty, doctors presumably have much more opportunity to introduce their personal goals into their decision making than
when it is absolutely clear (for the doctors) what they have to do. This aspect will be described below in Section 2.2.3.

Thirdly, there are different theories about what it is that doctors actually want, in other words which elements appear in their utility function.11 When it seems that doctors in certain circumstances can influence patients’ demand to further their own interest, it is important to know what that interest is. And we can also ask ourselves to what extent doctors actually pay attention to their own interest. This is a question that is often analysed in the context of agency theory, where the doctor is seen as a more or less perfect agent for the patient. This matter will be discussed separately later in Section 2.2.4.

2.2.2 Information-asymmetry between doctors and patients

One view is that asymmetry of information between patients and doctors plays an important role. This view can, for example, be found in Arrow (1963), Weisbrod (1978), Evans (1984), Wennberg (1985), Donaldson and Gerard (1993) and Light (1997)). Weisbrod (1978) makes an important point about information-asymmetry even after the health care service has been ‘consumed’:

"What a buyer wants to know is the difference between his state of well-being with and without the commodity being considered. For ordinary goods, the buyer has little difficulty in evaluating the counter-factual - that is what the situation will be if the good is not obtained. Not so for the bulk of health care .... The noteworthy point is not simply that it is difficult for the consumer to judge quality before the purchase ... but that it is difficult even after...”

Information-asymmetry is in this approach not only very important, but it is also difficult for the patient to take action to diminish the information-asymmetry. Evans (1984) has a similar point of view. He makes an important distinction between acute and chronic illness:

11 In this thesis, a doctor is considered to be a person who wants to maximise his utility. Mcguire (2000), in his ‘state-of-the-art’ article presents several models in which doctors maximise profit. The concept of leisure is not explicitly mentioned. However, it could be expected that leisure is one of the things that doctors like to have (part of their utility function).
"The technology of health care is sufficiently complex that consumers purchasing care in arm’s-length transactions, responding only to relative prices, would make mistakes. Moreover such mistakes may be serious, and potentially irreversible - the wrong care or care of poor quality (or no care) at a critical point can have permanently damaging or fatal results. And the possibilities for learning from the experience of others, or even from one’s own, are distinctly limited. In cases of chronic illness, or of well-defined and fairly common episodic illnesses, one may be guided by past, or others’, experience. But the same presenting complaint or symptom may at different times or in different persons represent very different problems and require quite different responses." (p. 70)

Another relevant point that Evans (1984), amongst others, stresses is that the asymmetry of information for the patient does not so much concern the characteristics of a good or service, but rather the effect the service will have on the user (pp. 72-73). There would be no problem if the patient could buy health or ‘health status’ directly. But the patient can only buy health care. And the patient has much less information about the relationship between health care and health than the provider (both generally speaking and in the patient’s specific circumstances) (see also Arrow, 1963).

It is very difficult for acute patients to collect information in such a way as to largely remove the role of information-asymmetry from diagnosis and treatment. Some researchers place more emphasis on the possibility for patients to acquire information about a doctor and the influence of this possibility on the functioning of the market for health care. Pauly and Satterthwaite (1981) are typical representatives of this way of thinking. They developed what they called an ‘increasing monopoly model’ for primary care physicians’ services, in which consumer information plays an important role. They consider primary medical care a reputation good, because “each physician delivers a service that is differentiated in place, style and technical competence from the service provided by any other physician” (p. 489), and because they believe that people search for a doctor mostly by asking family members, friends and acquaintances for recommendations. They consider the market for primary care in metropolitan areas to be monopolistically competitive (an assumption that is often made about care markets). An important element in this theory is that doctors can set their own price, and that there is price variation between doctors. They also point out that health insurance covers only a small part of costs for the relevant primary
care physicians (general practitioners, pediatricians and internists). This means that the price is important for the patients. The theory of Pauly and Satterthwaite is based upon two assumptions:

1. When the number of doctors in a community increases, the amount of information available about every individual doctor decreases.
2. When it becomes more difficult to acquire information, the price elasticity of demand decreases.

The conclusion from these two assumptions is that when there are more doctors in the community, they can charge a higher price for their services (other things being equal). The theory of Pauly and Satterthwaite was developed to explain this phenomenon that is often observed but seems paradoxical. The idea behind the theory is that acquiring information about a certain doctor becomes more costly when there are more doctors, because the probability is smaller that relatives and friends have information about this physician. This explanation does not seem to be very solid. It is not clear why patients ask for information about Doctor X. They could also ask relatives and friends for information about their own doctors and how satisfied they are with them. Apart from that, an important question is what sort of information could be gathered from relatives and friends about doctors. Could they really judge the medical-technical quality of the doctor, or could they just know whether they liked the doctor's attitude to his patients? Probably it would be difficult for patients to know even afterwards if they had had the best possible care. Only when something very obviously went wrong, would they be able to tell other people to avoid a certain doctor. The explanation of Pauly and Satterthwaite for the observed relation between the number of doctors and the price is not very satisfactory. Other explanations are suggested, such as a decrease in travel time (resulting from a higher density of doctors) that causes an increase in demand. However, it is the view of the present author that supplier-induced demand is a more convincing explanation for this phenomenon.

12 In the Netherlands, the CTG (National Health Tariffs Authority) determines maximum prices for different medical services. During the experiment, and at present, there were, and are, no physicians who charge less than the maximum price. The costs for primary care are largely covered by insurance.
Another important question, also connected with the Pauly and Satterthwaite-explanation, is what type of information-asymmetry has the largest influence on the market for health care. Is this information about the quality of a certain doctor’s treatment as a whole, or is it information about the best treatment in a specific situation? The latter question is most relevant for the influence of the payment system upon the behaviour of doctors. When patients know more about the quality of specific doctors, they will try to go to those doctors who are ‘good’ and try to avoid the ‘bad’ doctors. This means that some doctors will have a more thriving practice than others, in so far as patients have a free choice of doctors. But the problem that is posed in this thesis is: What is the influence of the payment system on diagnosing and treatment by doctors, apart from the medical-technical ‘quality’ of a specific doctor which may be taken as given. Will a doctor with a given quality do more or less than the optimal standard of care prescribes in a certain situation, under the influence of financial incentives that are determined by the payment system?13

Above, it was explained that it is often very difficult for the patients to know what is the optimal treatment for their medical problems. It may be safely concluded that, at present, asymmetry of information between doctor and patient plays an important role in the curative sector.14

2.2.3 Does the doctor have complete information?

This section is concerned with the question: To what extent does the doctor have information about the best treatment in a specific situation? Common sense suggests that it is not possible to know everything. There is evidence from a large number of studies about differences in treatment between regions and countries for comparable populations (or controlled for age and sex composition

13 In theory, when a doctor does consistently more or less than is optimal according to objective medical standards, this can influence the doctor’s reputation with other doctors and perhaps with patients. However, when all or most doctors under a certain payment system do consistently more or less than is optimal according to objective medical standards, the reputation of the individual doctor will not be influenced.

14 Possibly in the future this asymmetry will diminish because of increased use of the Internet to answer medical questions. Even so, the Internet can only give general answers which may not be directly applicable to a specific patient and his or her specific complaints.
of the populations) and about differences in treatment between doctors in the same region.\textsuperscript{15}

Studies about differences between individual doctors in one region are especially interesting, since these differences can not be caused by differences between regions. In Phelps, Mooney, Mushlin et al. (1994), the results were controlled for differences in case-mix and severity of illness, age, gender etc. and all doctors operated within the same insurance plan. So, all differences found between doctors were probably ‘real’ differences and not the results of shortcomings in the data or the methodology. The study concerned primary-care doctors who served as gatekeepers for other health care. Because of this system, all medical costs were attributed to the primary-care doctor. In this study it was found that there were large differences between individual physicians. ‘On average, the lowest-cost 10 percent of doctors used about one-third the medical care resources as the top 10 percent of the doctors’ (Phelps (1997), p. 93). The difference between the decile with the lowest costs and that with the highest costs was about $1000, while the average expense for the specific population (under-65) was $879 per patient per year (see Figure 2.1).

\textsuperscript{15} Phelps (1997) in his overview of this evidence mentions, for example, Phelps and Parente (1990); Wennberg (1990); Chassin et al. (1986); Roos et al. (1986); and McPherson et al. (1982). Flierman (1991) analyses uncertainty in the case of GPs.
Figure 2.1, Difference from average annual cost per patient in dollars (by decile of practice costs) in primary care physician practices in Rochester, New York.


Studie s that compared regions usually used hospital admission rates for various procedures as the basis for analysis. Phelps concludes from the studies that there are many surgical and medical interventions for which a stable pattern of relatively high variation is found. That is to say, between studies there may be rather large differences in absolute variations for certain procedures, but the procedures with a relatively high variation are usually the same ones in the

\[16\] The intention is to analyse differences in the use of procedures in comparable circumstances. Patients are admitted to enable the physicians to perform the procedure. The analysis is not aimed at the question whether a clinical admission is necessary to perform a certain procedure.
different studies. Examples of surgical procedures with a high variation are removal of tonsils and adenoids and removal of haemorrhoids. The uncertainty about hospitalisation was found to be at least as great for a number of nonsurgical procedures. In Wennberg, McPherson and Caper (1984), very high variations were found "in such areas as urinary tract infections, chest pain, bronchitis, middle-ear and upper-respiratory infections (both adults and children), and pediatric pneumonia". The coefficients of variations were over 0.4. (Phelps, 1997, pp. 85-86). Another example of high variations was adult admissions for medical back problems (see Phelps and Parente, 1990). The coefficient of variation was 0.31.

Wennberg (1990) studied the use of a number of surgical procedures in the market areas of 16 major university hospitals and large community hospitals. Some examples of the ratios between the highest and the lowest rate of use are: 7.42 for total knee replacement, 5.6 for laparotomy, 4.48 for cardiac catheterisation, 3.62 for coronary bypass surgery, 3.12 for prostatectomy and 2.29 for open heart surgery.\footnote{Phelps (1997), p. 87-88.}

Phelps shows ample evidence that there are large differences in treatment that cannot be explained by differences in the composition of the population of patients. He attributes these differences to uncertainty about the best treatment, after having checked that the differences can not be explained by differences in relative costs between treatments.\footnote{It can be concluded that the differences in relative costs are not the cause of differences in treatment because, in a study designed to analyse this phenomenon (Phelps and Mooney, 1993), no significant negative correlations were found between therapies that could serve as substitutes (except in one case).} Other possible causes of the differences (apart from uncertainty and differences in relative costs), include:

1. differences in the goals of doctors;
2. differences in the preferences of patients;
3. institutional differences in the organisation of health care (e.g. the payment system or the legal system).

The last of these categories (3) cannot play a role in the study by Phelps, Mooney, Mushlin et al. (1994), since that study concerned one region and one insurance plan. It is also difficult to imagine why there would be large differences in preferences of patients between doctors in this one region, or why there would
be large differences in preferences between the doctors. We can conclude therefore that there are different ‘practice styles’ that may be attributed to uncertainty or different habits of doctors.

Evans (1984) expects that this lack of information will lead the physician in a certain direction. “And the combination of professional training with the perfectly natural human desire to ‘do good’ (or more important, to have done good) for one’s patients leads to an overestimate of the efficacy of interventions in general, relative to what can be scientifically substantiated” (p. 77). So Evans expects that doctors will do ‘too much’. 19

Flierman (1991) proposed a different method for measuring the professional uncertainty of doctors. He is of the opinion that uncertainty should be related directly to the morbidity of patients, instead of just correcting for age and gender. His method is to try to explain the number of certain diagnostic or curative services per doctor from the numbers of different relevant diagnoses or complaints per doctor. The resulting $R^2$ is his measure for professional certainty: when the performing of services is determined to a high degree by diagnoses made, the level of professional certainty is high. Flierman calculates a number of these $R^2$'s for services performed by Dutch GPs, from which it can be concluded that there is a lot of uncertainty for some services as regards the question of in what circumstances they should be performed by the GP. Examples are: making an ECG, ESR-measurement, bladder catheterisation and examining vaginal discharge. For these services the $R^2$ is 0.2 or below. Furthermore, Flierman's results indicate that a larger degree of uncertainty for certain services leads to a larger effect of changes in the payment system. 20 We will discuss the theoretical model he uses further on.

There is more than enough evidence to conclude that doctors – though better informed than their patients – also experience a considerable degree of uncertainty in diagnosing and treating patients for certain conditions.

19 This overestimate is a possible explanation for the observed phenomenon that doctors and their families (who are better informed than other patients) use more health care. Another possible explanation is that health care is cheaper for them.

20 This hypothesis could not really be tested strongly because of shortcomings in the data for an experiment in Denmark.
2.2.4 What does the doctor want?

In this section, the goals of the physician are described. These goals are very important to the influence of the payment system on the diagnosing and treatment of patients. Not only will these goals determine the doctor’s conduct, but through his incomplete agency role they will also influence the demand of the patient.

A number of different utility functions can be found in the literature. An important distinction is the one between ‘neoclassical standard’ utility functions for doctors in which only their own direct interests play a role (most importantly income and leisure) and ‘special doctor utility functions’ in which the patients’ interests also play a role in some form. The latter also contain elements like compliance with medical standards, ethical considerations, patients’ interests, social approval, professional satisfaction and a target income (see Figure 2.2). Health economists who believe that doctors only care about their own interests usually believe that there are good possibilities for patients to collect information about the doctor or the best treatment. If ‘complete selfishness’ for doctors were to be combined with an important influence of information-asymmetry there would be no limit to the amount of demand doctors could and would induce. So health economists who believe that information-asymmetry has a large influence usually also believe that there is some consideration, such as ethics or medical standards, in the doctor’s own mind that stops him from endlessly inducing demand. Evans (1984) was right when he wrote: “Removing healthy organs, or drilling healthy teeth, has a negative impact on the practitioner's overall satisfaction, even if it is profitable and the patients, believing the organs/teeth were diseased, are satisfied.” (p. 151).

In market-oriented models of doctors’ behaviour a ‘standard’ utility function is used in which net income and leisure play a role. The doctor wants to find the combination of net income and leisure that maximises his utility. This process determines his supply curve. When the price for the doctor’s services changes, an

---

21 A profit maximising function can also be used, but here we concentrate on utility functions.
income effect and a substitution effect will play a role.\textsuperscript{22} Rizzo and Blumenthal (1994) find that for male physicians there is a substantial income effect, but the substitution effect dominates. The total estimated wage elasticity for male physician labour supply is 0.23 (p. 449). As stated above, in this class of models it is usually assumed that there is an independent demand curve. The outcome of the process is determined by the meeting of supply and demand. In the often-used market form of monopolistic competition, the doctor does have market power. However, this is not unlimited. At some point his patients will leave him if he sells them too many or too expensive services. In this model, there is no place for a special role of doctors. They are just like all other suppliers of goods and services.

However, it is considered here that doctors are not just like all other suppliers. Generally speaking, we expect that ethics and professional standards play an important role for extensively trained and highly paid professionals who are concerned with something as important as health.\textsuperscript{23}

\textsuperscript{22} For example, when there is a price increase, performing services yields more money and this makes it more attractive compared with leisure. This is the substitution effect. But the price increase also makes it easier to reach a certain income level. From that point of view, the doctor could perform fewer services and still have the same income. This is the income effect.

\textsuperscript{23} Naturally, there are exceptions, e.g. doctors who harm their patients to earn more money.
McGuire (2000) gives four arguments why doctors can be expected to look beyond their own interests (p. 520). The first is that doctors are often self-employed and therefore they are free to pursue whatever is important for their utility. Secondly, since specialists are mostly relatively well off, it may well be that other objectives are relatively more important than moneymaking.\(^{24}\) Thirdly, the costs of giving priority to their own interests may be much higher for their clients than for those of other suppliers with an information advantage. Fourthly, there is a consensus in society that physicians have their professional autonomy in exchange for the notion that they act in their patients' best interest. In other words, they have to be good agents for their patients.

The present author considers that the third and the fourth argument are the most important\(^{25}\): doctors know that their services are very important for patients with serious diseases and they want to be good agents for their patients. However, their agency is not complete, since they also consider their own interests. And we

\(^{24}\) That specialists are relatively well off goes for Dutch society as well.

\(^{25}\) That doctors are self-employed can also make them feel it is very important to earn enough money. In discussions and negotiations about their income, Dutch specialists usually do not give the impression that the marginal utility of their income is falling.
should note that doctors do have many opportunities to further their own interests without harming their patients’ interests. We will go into that below.

Evans (1984) also believes that doctors can never be perfect professional agents.\textsuperscript{26} However, to stimulate them to be as complete agents as possible, society gives them a number of privileges. The doctors are protected from competitive market forces. They can earn a high income, even higher than their long period of education and long working weeks would bring them in a neoclassical world. There are restrictions to entrance which give doctors a form of monopoly power. So, being cushioned from market forces and having regard to medical ethics stimulate doctors to play their (incomplete) professional agency role. The degree of uncertainty also influences this process. Medical ethics keep the doctors from performing, just for economic reasons, services that are actively harmful to their patients, such as the above mentioned removing of healthy organs. However, their ethics might not always keep them from doing services whose usefulness is uncertain, but which do not seriously harm the patient at the same time. An example of this might be extra diagnostic testing compared with an objective standard of care. In such cases, the trade-off between ethics, income and leisure time might have a different outcome. This seems all the more plausible as physicians’ training probably encourages them to do everything possible to be on the safe side. This means that the more uncertainty there is about the correct diagnosing and treatment, the larger is the influence of the payment system. Differences in treatment between doctors may not only be caused by differences in subjective assessments; they may also be caused by differences in financial incentives in those situations where medical ethics do not give a 100% certain guideline.

Apart from net income, leisure and ethics, a fourth argument can play a role in the doctor’s utility function (see, e.g., Evans (1984), p. 151). This is a preferred practice style, included by Evans because it was observed that there was an

\textsuperscript{26} Evans makes a clear distinction between a professional agency relationship and a principal agent relation. In the latter case, the agent has to be given incentives to also pursue the principal’s interest. When he can ignore the principal’s interest without detection, he will do that. In the case of a professional agency relationship, the doctor will take the patient’s interest into account, even when there is no control, because it plays a role in his own utility function.
underuse of auxiliary personnel, or "intermediate-level health practitioners".\(^{27}\) Physicians could hire such personnel and substitute part of their own expensive time for cheaper time of auxiliary personnel. In this way, they could raise their profits. Evans states that there are some institutional impediments, but these do not offer a complete explanation for the underuse. An explanation may be that doctors are attracted to certain practice styles, for example they prefer doing things themselves to managing teams of people.\(^{28}\)

Another aspect which may be included in the utility function of doctors is the interest that society as a whole has in the cost-consciousness of physicians. Doctors have a large influence upon resource use in the health care sector. They decide not only about their own involvement in the patients' treatment but also about many other related production factors. Of course, society's interest does not automatically mean that cost-consciousness is important for the doctors. Evans considers it of importance, but does not include it in the utility function. He sees the interest of society as one of the reasons why the relationship between doctor and patient can never be a complete professional agency relationship. Delnoij (1994) believes that society's interest is part of the doctor's utility function because of his interest in social approval. In her view, the doctors' utility depends upon both physical well-being ('produced' by income and leisure) and social approval, which may be produced by patients' appreciation of the doctors' services or by the appreciation of society. Proxies for society are, for example, colleagues or third-party payers. However, she believes that: "since quite often there is a considerable amount of uncertainty a physician has to deal with when assessing the benefits of a particular medical procedure for a specific patient, and since apart from that, social control is often not so strict as to outweigh the disadvantages of acting against an individual patient's best interest, physicians will relatively infrequently choose to let societal interest prevail" (p. 28). The present author agrees with her on the probability that society's interest prevails. That is why cost consciousness is not included in the doctor's utility function. However, it is possible that social approval in other forms plays a role in the utility function. During the experiment social approval of the partners in the experiment may

\(^{27}\) This phenomenon was analysed by Reinhardt (1972, 1973, 1975).
have been gained by working towards the not-directly financially stimulated goals of the experiment, like cultural change and efficiency projects.

2.2.5 Conclusions on the basis of theory

Doctors are by no means always certain about the optimal way to treat patients. Still, they usually have a lot more information than their patients and this information-asymmetry can not easily be removed. This means that doctors can influence patients demand for care through their advice: there is no independent demand curve. Fortunately, it seems likely that doctors do not only pay attention to their own interests (income, leisure, pleasant work and perhaps social approval) but also to their patients' interest (ethics). Since these different elements of the doctor's utility function are traded off against each other, doctors are not complete professional agents for their patients. As incomplete agents, they will be more constrained by their ethics, the more certain is the optimal treatment and the more serious is the (possible) disease of the patients. In the latter case, the cost to the patient of the doctor not being ethical can be a lot higher.

In this theoretical model, it is to be expected that a change in the payment system does influence the behaviour of physicians, at least as long as there is no complete certainty. A change in the payment system changes the optimal work pattern of diagnosis and treatment for the doctor through the elements in the utility function net income and leisure time, unless there is an absolute constraint from ethical considerations. In many cases, there will be no such absolute constraint because there is a large grey area in the optimal treatment. Instead, a new optimum of income, leisure, ethics and pleasant work will be found. So there will be some change in the doctor's preferred way of working and, since patients tend to follow the doctor's advice to a large degree, this change in the doctor's preferences will manifest itself in the actual process of diagnosis and treatment. Therefore, a change in the payment system will have an effect on the diagnosing and treatment of patients from a theoretical point of view. Such a

Phelps (1997) suggests another possible explanation. The price for a physician's labour as seen from the outside is high, since it includes monopoly rents. For the physician as owner of the firm, the price of labour is lower, so it only appears that he uses too much of his own labour.
change is expected to have the strongest effect for those services where there is the least certainty about the optimal treatment.

2.3 Empirical findings

The question in this section is whether the expected influence of the payment system has been found in empirical research. There is a very large amount of empirical research on this and on related topics, most of which is concentrated on the possibility of ‘supplier-induced demand’. This may be defined as: “the amount of demand, induced by doctors, which exists beyond what would have occurred in a market in which consumers are fully informed”. The research into supplier-induced demand originated from the above-mentioned empirical observation that there was a positive correlation between the available supply of doctors and price, and in the hospital sector that “bed availability is the principal determinant of bed use”. Many researchers were of the opinion that these phenomena pointed to the large influence of doctors upon the demand of their patients. Others stated that there were ‘normal’ economic arguments to explain the observed correlations and therefore the standard analysis did not have to be abandoned. An example of such an argument is that a higher density of doctors lowers the time cost of patients, thereby increasing demand.

Not only is the concept of induced demand controversial, but so too are the methods used to investigate it empirically. Dranove and Wehner (1994) show that the standard approach used to investigate induced demand leads to the strange conclusion that demand for childbirth is induced by obstetricians. It seems, therefore, that there are methodological difficulties with the research into supplier-induced demand. So, while there are many indications that supplier-induced demand exists, most of them are not conclusive for methodological reasons or for data reasons.

One study is often cited in publications upon this subject, because many controls are built in to avoid biased results. This is the study of Hickson, Altmeier and Perrin (1987). They tried to ensure that the results of their study were not affected by differences in setting, types of doctors, and patient behaviour and

30 Evans (1984), p. 84.
31 DeVany, House and Saving (1983).
characteristics. The study was conducted in a pediatric trainee specialists' continuity clinic. In this setting, it was easy to assign the trainees randomly to two different reimbursement systems: salary and FFS. In this way, differences in setting were avoided. The parents of the patients were all confronted with the same payment conditions. They did not know how their own physician was being paid. There were no statistically significant differences in the characteristics of the two groups of patients that were assigned to the salaried and the FFS trainee specialists. One problem remained, i.e. the two groups of trainee specialists, while matching each other in all other characteristics measured, differed in their career plans. In the salaried group, more trainee specialists wanted to go into private practice (versus an academic career) compared with the FFS-group. So the researchers checked whether the results of the comparison of the two groups of trainee specialists were resistant to controlling for career choices. The results that still held were that in the FFS-group the continuity of care was better (a greater number of the visits were attended by the patient’s primary physician), there were less emergency-room visits per patient and more preventive visits were scheduled per patient. To find out whether the extra visits were excessive, a number of charts were examined and the documented visits were compared with the guidelines of the American Academy of Pediatrics. It was found that: "Patients assigned to fee-for-service physicians missed fewer recommended visits than patients of salaried physicians (3.0% v 9.4%) and made more visits in excess of the recommendations (18.0% v 4.7%)." The following conclusion can be drawn from the results: under FFS, doctors do 'too much' compared with the objective standard of care, and under salary they do 'too little'. This conclusion is not drawn so strongly by Hickson, Altmeier and Perrin. They consider it possible that the extra visits to well children were appropriate "because some families need more than the usual amount of well child care" (p. 349). One of their overall conclusions is:

"Our results add to the existing literature suggesting that physicians given appropriate motivation (fee-for-service reimbursement) and appropriate circumstances (less than full capacity of patients) can manipulate demand for care and consequently patient use of services" (pp. 349-350).

Summarising, this thorough study shows that the payment system does indeed influence the behaviour of doctors. And it shows that this influence is also

possible for situations where 'a correct standard of care' is available (in the form of the guidelines of a respected scientific organisation), or, in other words, where the doctor can know exactly what to do.\textsuperscript{33}

There is a body of research in another direction that also supports the conclusion that the payment system is of influence.\textsuperscript{14} This research concerns referral by doctors to facilities in which they have a financial interest. These may be joint ventures with doctors as financial partners in diagnostic imaging centres, radiation therapy facilities or physical therapy centres. Mitchell and others conducted a number of studies into this phenomenon.\textsuperscript{35} These studies yield consistent evidence that doctors' referral to facilities in which they have a financial interest does not lead to improved access to care or better quality (arguments that are sometimes used to defend or promote this practice). However, it does lead to increases in the amount of services used and increases in cost. In a number of cases, it was found that the prices for comparable procedures were higher in joint venture-facilities than in independent facilities. Phelps (1997) concludes from this research: "... these studies all show the potential importance of induced demand in settings in which the doctor both has the informational 'power' to recommend treatment to patients and a direct financial reward from doing so" (pp. 252-253).

So, both direct empirical research into the influence of the payment system on the treatment of patients and indirect empirical research into the influence of doctors having a financial interest in facilities on referral, show that financial incentives do indeed influence the outcome of the process.

2.4 Preferred theory and hypotheses

On the basis of theoretical considerations, the expectation was formed that a change in the payment system does influence the behaviour of physicians, at least as long as there is no complete certainty. This hypothesis is supported by

\textsuperscript{33} However, it is still possible that there is uncertainty connected with the individual circumstances of the patient.

\textsuperscript{34} Phelps (1997), pp. 251-253, gives an overview of this research.

\textsuperscript{35} Mitchell and Scott (1992a); Mitchell and Scott (1992b); Mitchell and Sunshine (1992); Hillman et al. (1990).
empirical findings. Therefore, the following model is selected as the theoretical basis for the empirical analysis undertaken in this thesis.

Information-asymmetry between doctor and patient plays an important role. As a consequence of this there is no independent demand curve. However, the doctor does not have complete information. He is confronted with uncertainty about the optimal treatment. Under these circumstances, he tries to maximise his utility function, elements of which are income, leisure, ethics, pleasant work, and possibly social approval. It is easier for him to concentrate on his own interests (mainly income and leisure) when the optimal treatment is uncertain and the condition of the patient does not appear to be too serious.

In this model, what can we expect from a change in payment system from (largely) FFS to lump-sum budgets that are fixed for a year? The main expectations are set out in Box 2.1 in the form of five hypotheses.
Chapter 2

Box 2.1 Expected effects of a change in the payment system for specialists

1. If, in the old situation, services were performed that were not strictly necessary from a medical point of view, the number of these services will decrease (this increases leisure without decreasing income or increasing ethical costs);

2. If, in the old situation, services were performed that were necessary but did not have to be performed by the specialist, more referrals will be made, for example to the GP (this increases leisure without decreasing income or increasing ethical costs);

3. Because income is less dependent upon production (to a degree depending upon the specific arrangements in a particular project), leisure will become more attractive compared with the old situation, unless there is a binding ethical constraint (the financial costs of leisure decrease while the ethical costs do not change);

4. It may become more attractive to generate production in terms of the units in which it is measured in the new system, but this depends upon the specific trade-off between income and leisure. Producing more in terms of production units contributes to reaching the arranged production level and may (depending upon the specific arrangements) lead to a higher income in the following year. However, producing more according to the new definition may or may not decrease leisure, depending upon how much efficiency can be improved, for example by no longer performing the above mentioned services (see 1) and 2) above).

5. It will become more attractive to decrease production in terms of the units that do not play a role in the new system, as long as it is not really medically necessary that these things are done by the specialist himself.

36 This may be the case when the doctor believes that waiting can have serious consequences for patients, for example women with suspected breast cancer.
37 This concerns not just the number and type of services (see footnote 37).
38 For example, when the number of repeat visits to the specialist at the outpatient department is not important for the measurement, having less repeat visits increases leisure without having to give up income. If some repeat visits are not absolutely necessary or patients can be referred back to the GP for control, the ethical costs do not increase. In that case, the optimal choice for the specialist is to have less repeat visits of patients.
It can not be said a priori whether production (measured according to the new definition) and leisure will increase or decrease. On the one hand -supposing there are no binding ethical constraints- leisure becomes more attractive and non-measured production becomes less attractive. On the other hand measured production may become more attractive. In that case non-measured production will decrease. Leisure may increase or decrease depending upon the relative attractiveness of measured production and the possibilities to use care more economically. One of the goals of the experiment was to promote efficiency. Perhaps it is possible for doctors to have more leisure and more measured production at the same time by working as economically as possible.

What do the expectations formulated in the hypotheses mean for the extent to which the use of care is suitable? The first hypothesis of no longer performing unnecessary services concerns unequivocally an improvement in efficiency and cost savings. So this promotes suitable use of care. The second hypothesis of more referrals promotes economical use of care. The patient is treated by the most appropriate provider when the referrals are well-considered. For example, when the specialist no longer performs services that can easily be done by the GP, this gives the specialist more time for specialized work. The third hypothesis about leisure becoming more attractive promotes suitable use in as far as a more rested specialist can give better care. However, it can hurt the quality of care because of increased waiting times or it can lead to higher costs to keep waiting times at the same level. So this expected effect of the new payment system appears predominantly unfavourable for suitable use. The fourth hypothesis about possibly producing more in terms of the new production parameters is favourable in the sense of keeping waiting times lower and being an incentive to produce in an economical way (both provided the new production parameters are wisely chosen). Cost control regarding the budgets of specialists is probably not maximal in this situation, since the incentive for the extra production is the possibility to increase income. Cost control in the hospital may be promoted by using care more economically. The fifth hypothesis of decreasing production in terms of units that are not measured anymore does not harm the quality of care as long as there is no need for the specialist to perform these activities himself. Quality of care can, in principle, be improved by lower waiting times because of freed time. However, the greater attractiveness of leisure may undo this advantage. The use of care becomes more economical when activities are no
longer carried out or carried out at a more appropriate level. The consequences for cost control are not clear a priori.

In Chapter 4, the arrangements and the financial incentives within the different local projects that took part in the experiment are discussed in detail. Following this discussion, more detailed hypotheses will be formulated. Our empirical analysis will show whether or not these hypotheses are confirmed.