CHANGING THE PAYMENT SYSTEM OF GENERAL PRACTITIONERS

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Changing the payment system of general practitioners / Henk A. Fliermann. - Utrecht : Nederlands instituut voor onderzoek van de eerstelijnsgezondheidszorg NIVEL
Proefschrift Rijksuniversiteit Utrecht. - Met lit. opg.
ISBN 90-6905-161-3
Trefw.: huisartsen ; honorering / omslagstelsel ; gezondheidszorg.

Wordprocessing and lay-out	Marcelle van der Meulen, Bernadette Kamphuys
Cover	Mieke Cornelius
Correction and translation	Trevor S. Preston
Printing	Boekbinderij Post

CHANGING THE PAYMENT SYSTEM OF GENERAL PRACTITIONERS

Verandering van honorering van huisartsen
(met een samenvatting in het Nederlands)

Proefschrift ter verkrijging van de graad van doctor
aan de Rijksuniversiteit te Utrecht
op gezag van de Rector Magnificus, Prof. Dr. J.A. van Ginkel
ingevolge het besluit van het College van Dekanen
in het openbaar te verdedigen
op 29 november 1991 des voormiddags te 10.30 uur

door

Hendrik Antheunis Fliermann
geboren op 24 mei 1948 te Middelburg

Promotor:	Prof. Dr. R. Wippler, Faculteit der Sociale Wetenschappen van de Rijksuniversiteit te Utrecht
Co-promotor:	Dr. P.P. Groenewegen, Nederlands Instituut voor Onderzoek van de Eerstelijnsgezondheidszorg NIVEL

Dit proefschrift werd mede mogelijk gemaakt met financiële steun van het Ministerie van Welzijn, Volksgezondheid en Cultuur en de Commissie Programma Evaluatie.

Acknowledgements

Jouke van der Zee (NIVEL) initiated the study reported here. Peter P. Groenewegen (NIVEL) made its design, insisted on its data collection and guided its analysis.
Allan Krasnik, Mogens T. Damsgaard, Gavin Mooney (Institute of Social Medicine, University of Copenhagen), Poul A. Pedersen and Peter van Scholten (Central Research Unit for General Practice, Copenhagen) organized and executed the study's data collection, as well as providing earlier drafts of the analysis with stimulating criticism. Together we reported the first results of the study (Krasnik et al., 1990).
Reinhard Wippler (Department of Sociology, University of Utrecht) inspired the scientific choices in the study's analysis.
Peter P. Groenewegen is co-author of Chapters 2 through 7. Koos van der Velden (NIVEL) is co-author of Chapter 6. Loek J. Stokx (NIVEL) is co-author of Chapter 8. The defender of this thesis is principal author of all chapters.

Contents

1. Policy and scientific context and design of the study

1.1 Introduction

In many countries the health care systems are under debate (Ham et al., 1990). The Netherlands is no exception.
In the Netherlands, a change of health care system is being prepared by the central government. The existing system is typified as one in which for most patients health care demand is not constrained by financial considerations, and rising public health care costs are counteracted by attempts to control quantity of supply. The proposed change is one towards a competitive health care market subject to quality controls (op.cit.: 44-46; Ministerie, 1990: 50-62).
Such a market's institutional conditions are best designed if one knows in advance what effects they will have on its outcomes. In this opening section we will argue that, compared to effects of changes on the demand side, evidence on effects of changes in institutional supply side conditions is scarce. Effects of changes on the supply side are the subject of this study.

1.1.1 Demand side

In the Netherlands, health care is freely accessible to those whose health care costs are 'publicly insured'. Thus insured are employed persons whose incomes are below a level set by law, and some categories of old age pensioners and social security dependents. Their access to health care is limited only in that they have to register with a general practitioner (GP), and need to be referred by their GP to visit a medical specialist. They make up some 65% of the population.
To people whose incomes are above the level set by law, public insurance is closed. They are free to take out private health insurance. Such a policy may or may not cover reimbursement of GPs' bills. If it does, registration with a GP is not a prerequisite. For reimbursement of bills from medical specialists and hospitals, referral by a GP is not a strict prerequisite either.
With public insurance, premiums are income-related. With private insurance they are nominal, and often have deductibles. In the new health care system, a mix of both types of insurance should apply to the whole population. In that mix, premiums for some types of health care should be income-related, and those for other types should be nominal, allowing for deductibles. Effects to be expected from the introduction of such a mix are well established.
In the Netherlands it was found that controlling for health status, privately insured persons without insurance coverage of GPs' services visit their GP less often than those with such coverage (Van Vliet & Van de Ven, 1986; Van der Zee, 1990). This finding is consistent with results of a controlled trial of cost sharing in health insurance in the United States, where

more cost sharing by the insured resulted in smaller probabilities of visiting a physician (Newhouse et.al., 1981:1505). When serious and minor symptoms are distinguished, this holds only for persons with minor symptoms (Shapiro, Ware & Sherbourne, 1986). The constraint on demand for ('elective') care of minor symptoms that is exerted by money costs to patients seems to be well established.
Thus if premiums for elective care of minor symptoms were nominal with deductibles, and if premiums for necessary care of serious symptoms were income-related, health care demand would be constrained while quality of care would be maintained. The problem that remains to be resolved is how to design such demand-side conditions of a health care market. No problem with absence of evidence exists, as it does for the supply side.

1.1.2 Supply side

In the health care market whose institutional conditions are being prepared by the Dutch government, former public and private insurance carriers alike should contract with health care providers.
Among health care providers, GPs so far contract only with public insurance carriers, or 'sick funds'. Up to now, sick funds have paid GPs capitation fees for their publicly insured patients. Privately insured patients pay GPs fees for services, reimbursement of which depends on the insurance policies chosen.
Anticipating the health care system change, the Dutch Association of General Practitioners (LHV) advocates the introduction of a uniform payment system for GPs' care of the whole population. That payment system would be a mix of capitation fees and fees for services. The effect of such a payment change on health care provision is far less well established than the effect of cost sharing on demand. Two reasons for that can be pointed to.
For one, payment changes are generally put through on a nation-wide scale, as is now proposed by the Dutch GPs' association. In such a case no suitable control group exists for the evaluation of the change. Next, cross-sectional analyses are often a bad substitute because differences in the payment of providers coincide with differences in health status or insurance of patients. Again, the publicly and privately insured sections of the Dutch population are an example. In such circumstances, one has no evidence of the effects to be expected from a payment change.
However, a payment change similar to the one that is now proposed in the Netherlands has recently been implemented in Copenhagen, Denmark. There a suitable control group was formed by the surrounding area, where the GPs' payment system remained unchanged. Also, circumstances in Denmark and the Netherlands seemed comparable enough to allow for generalization of Copenhagen effects to the Netherlands, if remaining institutional and distributional differences are taken into account.
In this study, this situation will be used to estimate effects to be expected in the Netherlands. In Section 1.2 the policy context of the study is elaborated. Its scientific context is discussed in Section 1.3. In Section 1.4 the design of the study is explained. The contents of the book are sketched in Section 1.5.

1.2 Policy context

Up to now, Dutch GPs have been paid capitation fees for publicly insured patients on their list. As Glaser (1970: 254, 271) noted, historically the principal arguments for capitation were its administrative simplicity and its encouragement of preventive medicine.
As for preventive medicine, the argument went that a GP would minimize work to be done for a fixed fee per registered patient by keeping his patients healthy. Yet it was soon recognized that the argument did not hold, since generally GPs simply do not see all of their registered patients, who should be kept healthy to minimize work. Generally they see only the patients who present them with a complaint.
Moreover, a GP might minimize work to be done in return for capitation by underprovision of necessary care. Glaser (op. cit.: 286) argues that this is counteracted by medical ethics, as observed by individual GPs and as enforced by their professional organizations.
Finally, a GP may minimize work to be done by early referral to specialists. Glaser (op.cit.: 287) notes that this is avoided only if there is firm agreement on the division between the basic care to be provided by GPs and the advanced tasks to be performed by specialists.

To strengthen the GPs' profession, the Dutch GPs' association has agreed on a 'basic set of tasks' (LHV, 1983) that every Dutch GP should be prepared to perform for his patients. In that way, the profession tried to rule out by explicit ethics both underprovision of necessary care by leaving these tasks undone, and early referral by leaving them to specialists. Still, in doing so negative incentives attributed to capitation are counteracted at the most.
Next, in a working paper published in 1987 (LHV, 1987), the association advocated the payment change considered above to one payment system for both publicly and privately insured patients, consisting of a mix of capitation fees and fees for services. The introduction of one payment system was advocated because of administrative simplicity. The introduction of fees for services to publicly insured patients was advocated because "capitation fees contain no monetary incentives for self-activation". On the other hand, the introduction of capitation fees for privately insured patients was advocated because pure fee-for-service "stimulates quantity instead of quality" of services rendered. Fees for services included in the basic set of tasks were proposed to be deducted from capitation.
Finally, in 1990 the Dutch general practitioners' association and the association of Dutch sick funds agreed in principle on the introduction of fees for an initial set of seventeen services not to be deducted from capitation*. Hence it was implicitly agreed that these services would not be included in the basic set of tasks. For all of these services it was argued in the agreement that their performance serves to replace a referral to a medical specialist. By that time, for some of these services standards on how and when to perform them had

---

* These services are: electrocardiogram, check-up on type II diabetes mellitus, examining vision, examining vaginal discharge, audiometry, removing foreign bodies from eye, or from throat, nose or ear, incising an abscess, stitching an open wound, bladder catheterization, bandaging a sprained ankle, excising a sebaceous cyst, excising a nevus or mole, excising the wedge of an ingrown toenail, puncture or injection with bursitis, adjusting or removing an IUD, or a pessary.

been published by the Dutch Society of General Practitioners (NHG). For these services, the condition was made in the agreement that they should be performed according to these standards.
Since 1989, in a continuing endeavour to strengthen the profession, standards on how and when to perform medical services are being developed and published by the Dutch GPs' society (Tielens, 1989). For the future, the GPs' association plans to make registration and five-year re-registration as a GP conditional on tested ability to perform services included in the 'basic set of tasks' according to such standards (LHV, 1987: 34-35; LHV, 1990; Verdenius, Brands & Oudkerk, 1990).

Thus the profession advocates the abandonment of pure capitation because it is held to contain no money incentives for self-activation. Thereby the profession recognizes that under capitation, despite explicit ethics, GPs may minimize work to be done through underprovision of care or early referral. It recognizes that medical ethics may be counteracted by the institutional features of the payment system.
That recognition is also implicit in the profession's acknowledgement that pure fee-for-service stimulates quantity instead of quality. And again, it is implicit in the profession's assent to the condition that services paid for should be performed according to standards.
Still, it is claimed that if for earning his income a GP is not fully dependent on fees for services, and if his performance of services is monitored by standards, the introduction of fees for services will have positive effects. Quantity of services rendered to publicly insured patients will rise while quality is maintained. As a result, referrals will decline, it is claimed.

In policy evaluation, the validity of such a claim should be investigated. A set of questions should be answered. Are services rendered more often when fees for them are introduced? Are referrals avoided by services performed? Will quality be maintained in the process? Are these relationships dependent on monitoring by standards? Are they dependent on the part of income made up by fees for services? And are they dependent on other institutional and distributional characteristics of the setting of the payment change?
Here an a priori evaluation of the payment change proposed in the Netherlands will be reported that is performed by generalizing effects assessed in Copenhagen, Denmark. In assessing effects in Copenhagen, Denmark and in generalizing them to the Netherlands, these questions should be answered and the answers should be taken into account.
To answer the questions asked, we will first analyze the scientific context of the study.

1.3 Scientific context

The initiatives and arguments of GPs' organizations described above can be understood as reflecting their position of 'professionals as incomplete agents' (Evans, 1984: ch.4; Arrow, 1963).
A complete professional agency relation of doctor to patient would imply that the doctor

acts only and fully in the best interests of the patient, and that the patient expects him to do so. It presupposes that the doctor would know for certain what health care is in the patient's best interests, given his health status. If so, professional ethics would dictate that the doctor sees to it that such care is provided.
The pursuit of such a complete professional agency relation can be understood as a social response to three peculiarities of health care as a commodity: the uncertainty of illness incidence and resulting demand for health care, the assault on personal integrity implied by illness, and the ignorance of the patient, compared to the doctor, as to the instrumental value of the health care provided to his health status.
The incompleteness of the professional agency relation follows from the coincidence of two conditions.
For one, acting as a professional agent will produce professional satisfaction in a doctor, which will be a utility to him as a result of his socialization as a doctor. But meanwhile, in most instances a doctor sees that care is provided by providing it himself. If so, his income and leisure time are affected. And these are utilities to him as well.
As for the other condition, a doctor does not know for certain what care is in the patient's best interests, given his health status. "There remains, particularly for diagnostic and monitoring activities, a broad zone of uncertainty in which optimal treatment and the limits of efficacy have not been scientifically established. In this zone, the provider can exercise considerable discretion before encountering ethical constraints" (Evans, 1984: 89).

When doctors are modelled as incomplete professional agents, their utility function contains three arguments, one of which constrains the trade-off between the other two. The constraining argument is compliance with medical standards on when and how to perform the services contained in one's tasks, or compliance with medical ethics. The arguments constrained are income and leisure, or free time and money to spend. Time and money operate more strongly as arguments in the utility function if certainty about task definitions and standards is less. So doctors only exchange time for money, and vice versa, within the limits of the patients' interests. But the distance between these limits is larger if there is less certainty.
Now the strengthening of the profession as described in the previous section can be understood as changing the parameters of the professionals' utility function by increasing their certainty about task definitions and standards, thereby increasing the profession's demonstrable societal value.
As argued above, the utility function just postulated is produced by an analysis of the peculiarities of the health care commodity, and of the structure of the doctor-to-patient relation as a social response to these peculiarities. As a strategy, this seems superior to postulating the trade-off of income and leisure as the sole utility arguments to begin with, and then adding other arguments (patients' well-being, doctors' social esteem) ad hoc, to reconcile the model with findings persistent in the literature (e.g.: Richardson, 1981: 206-210). It results in a utility function that is more informative on the relative power (constraining vs. constrained) of the arguments included. And it calls for assessment of doctors' certainty

about task definitions and standards, as a prerequisite for deriving testable predictions. In this way it guides health services research.
As a spin-off, it invites one to look for other commodities sharing the peculiarities mentioned, and other incomplete professional agency relations developed as a response to them, resulting in a similar utility function to be postulated for other professionals. The lawyer-to-client relation in legal defense would be a candidate (Arrow, 1963: 948-949).
In either case, it seems appropriate to recall that, generally, professionals have a quite reasonable income anyhow. This could well be the precondition of the constraining power of ethics. A poor lawyer is a bad lawyer, and a poor doctor is a bad doctor, as the saying goes.

In rational choice theory, individuals are assumed to maximize utility, given their utility function, and given the resources and restrictions that make up the individuals' situation (Hechter, Opp & Wippler, 1990).
Here GPs are the individuals under study. Their utility function was postulated above. Their situation is modelled as consisting of two sets: the set of complaints presented to them, and their set of opportunities for looking after diagnosing and curing. Uncertainty is modelled as a characteristic of pairs of elements from both sets. Thus uncertainty is taken to be a characteristic of GPs' situation, and not a personality trait.
Opportunities for looking after diagnosing and curing, and for maximizing utility in doing so, are taken to be enhanced by resources and limited by restrictions. Both can be institutional (e.g., regulations concerning referral to medical specialists) or distributional (e.g., numbers of specialists available). This is the theoretical framework of our study.

Its subject is the effects of the introduction of a mix of capitation and fees for services to patients formerly covered by capitation only. That change in payment has been implemented in Copenhagen, Denmark, and is to be implemented in the Netherlands.
Effects to be investigated relate to the amounts of services rendered, and of referrals to medical specialists avoided in doing so. In Denmark as in the Netherlands, a GP is expected to refer a registered patient to a specialist when advanced services have to be rendered that a GP is not trained to perform. And in Denmark, as in the Netherlands, no money costs to either GP or patient are involved in such a referral. So far, institutional settings are equal.
What the sizes of changes in services and referrals in such a setting will be are empirical questions. But what the relation of these sizes will be under various conditions can be argued theoretically. We will now discuss what is implied by utility maximization in the settings mentioned, if within the utility function postulated above, time and money are fully or partly constrained by task definitions and standards. We will do so by presenting a set of four line-ups of conditions. This set is stretched in that it includes limiting cases in which doctors would be fully certain about task definitions, standards or both. Our reasoning about each line-up will be based on some simplifying assumptions too. We will assume that doctors are fully equal in the abilities they master and in the equipment at their disposal. Also, we will assume that there are no competing claims of GPs versus specialists on who

should render a service.

Line-up 1. Suppose that GPs are fully certain about both task definitions and standards, so that there is full agreement on what services are to be rendered by GPs, and on when and how they are to be performed. Then no utility maximization would take place, since the activities of GPs could be fully determined by the restrictions implied by patients' health conditions. If so, introducing fees for these services would not cause any change in numbers of services rendered or referrals avoided: all is fixed with patients' state of health.
Hence if there were full certainty about task definitions and standards, the payment system would have no effect on services and referrals.

Line-up 2. Now suppose that GPs are not fully certain about task definitions, but are so regarding standards. Because of certainty about standards, the patients' states of health inducing the performance of a service are fixed. Because of uncertainty about task definitions, a GP has a choice between either rendering the service himself or referring a patient to a specialist for that. Then time and money become relevant. Under capitation, a GP loses time only by performing the service. If a fee is paid for the service, he gets money in return. Thus under capitation he maximizes his utility by referring the patient. And after introduction of a fee for the service, he maximizes his utility by performing the service himself -- if the money gained is worth more to him than the time lost. If so, the introduction of a fee for the service will cause an increase in performances of the service by GPs. And because states of health inducing the service are fixed, it will cause an equally large decrease in referrals to a specialist for performing the service. Thus if there is no complete certainty about task definitions, shifts between services performed by GPs and specialists are possible.

Line-up 3. Next, suppose that the opposite conditions hold: there is full certainty about task definitions, but not regarding standards. Because of certainty about task definitions, a GP has no choice between rendering a service or referring a patient. Because of uncertainty about standards, given a patient's state of health he has a choice of whether or not to perform a service. Then under capitation he maximizes utility by deciding that the service is not due. When a fee for the service has been introduced, he does so by deciding that it is -- again, if money is worth more than time. Then the introduction of a fee for the service will cause an increase in performances of the service without any decrease in referrals.
With uncertainty about standards, the division of labor between GPs and specialists may therefore be too fixed for a payment change to produce any shifts between two.

Line-up 4. Finally, suppose that there is full certainty about neither task definitions nor standards. Then, given a patient's state of health, a GP has a choice of whether or not to think that a service is due. And if he thinks that it is, he has a choice between either rendering it or referring the patient to a specialist to perform it. Now under capitation he maximizes utility by either deciding that the service is not due, or deciding that it is and referring the patient. In both cases he does not lose time. After introduction of a fee he maximizes utility by

deciding that the service is due, and performing it. Now the introduction of a fee will cause an increase in performances of the service that is greater than the decrease in referrals: added to the instances that the service substitutes for a referral are the instances that GPs newly decide that it is due.

Now it seems safe to assume that among both Copenhagen and Dutch GPs certainty is incomplete about both task definitions and standards. Hence both in Copenhagen and the Netherlands an increase in services rendered and a smaller decrease in referrals can be expected to be caused by the introduction of fees for services.
Yet the degree to which the decrease in referrals can be expected to fall short of the increase in services depends on the relative degrees of certainty about both task definitions and standards. If in the Netherlands there is more certainty about standards and less regarding task definitions than in Denmark, Dutch GPs' conditions are more like those described in line-up 2 above. Then among Dutch GPs the decrease in referrals will to a lesser degree fall short of the increase in services.
From the policy context described in the previous section, this condition could well hold good. As for task definitions, only a 'basic set of tasks' was agreed among Dutch GPs, allowing for additional tasks to be performed by GPs inclined to do so -- leaving aside competing claims by specialists. And the services that fees were agreed to be introduced for are all implicitly agreed to be outside of this basic set. As for standards, if possible the condition was agreed on that the services should be performed according to written standards. In Copenhagen no such condition was implied in the agreement concerned.

In Holland and Denmark institutional rules concerning the referral by GPs of patients on their lists are equal. So far we have argued from that. What matters, however, are the opportunities to maximize utility that are offered within these settings. One distributional characteristic obviously relevant to that is the specialist-to-population ratio.
If specialist-to-population ratios are lower, so that there are fewer specialists per capita, then there are fewer opportunities to refer a patient to a specialist to have him perform a service within a given time. Now a standard with incomplete certainty may still prescribe that, given some patient's state of health, a service should be performed within a given time. A standard with more certainty will do so more often.
Therefore, lower specialist-to-population ratios more often rule out the choice between rendering the service or referring the patient: the option of referring a patient is more often excluded by waiting times. This affects the numerical relation of the instances discussed in line-up 4 above. Of the services performed after the introduction of fees, a smaller part will substitute for a referral that was excluded to begin with. Thus in a larger part of the instances GPs newly decide that the service is due.
With lower specialist-to-population ratios, services with less certain standards will take a larger part of the increase in services performed, and the decrease in referrals will to a higher degree fall short of the increase in services.
Now specialist-to-population ratios are in fact lower in the Netherlands. Thus from this dis-

tributional characteristic of the institutional setting a prediction follows that is opposite to the one based on relative certainty of task definitions and standards generally.

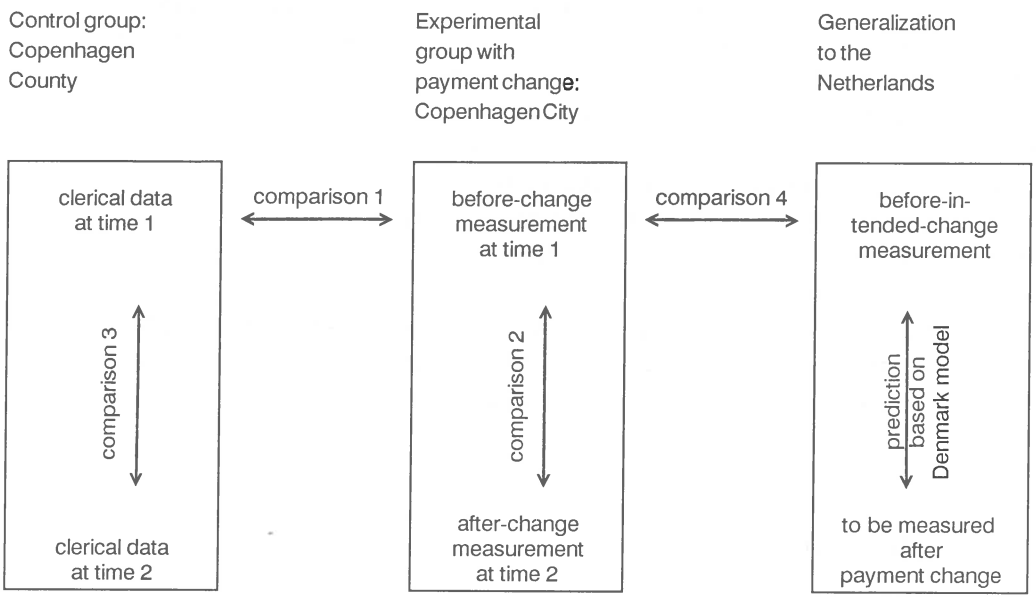
We ended the previous section asking a series of questions. One was: Are services rendered more often when fees for them are introduced? A priori we may answer: Yes, but the more so with less certainty about task definitions and standards. Another question was: Are referrals avoided by services performed? Again, a priori we may answer: Not fully, but the more so with less certainty about task definitions and more certainty regarding standards. A third question was: Are these relationships dependent on distributional characteristics of settings? Here one a priori answer is: Referrals are to a lesser degree avoided by services newly performed when specialist-to-population ratios are lower.
In assessing the effects of the payment change and in generalizing them to the Netherlands, ideally all of these a priori answers should be tested and results of that should be taken into account. Actually this will be done only partially here. For example, it is fully tested whether the increase in performances of a service is larger when certainty about when to perform it is smaller, and the result is fully taken into account in generalizing to the Netherlands. Next, the effect of relative referring opportunities is tested only by implication, and the result is taken into account only as a qualification of our predictions for the Netherlands. And finally, certainty about task definitions is left aside completely for lack of data.
We will now present a full description of the design followed in assessing and generalizing effects.

1.4 Design of the study

The study was designed as a natural experiment. The experimental group was made up by a sample of GPs in Copenhagen City, who recorded the contents of their consultations for three periods of one week. The contents of the consultations recorded relate to complaints presented and diagnoses made, services performed, and referrals made. The one-week recording periods were six months before, six months after and one year after the payment change in Copenhagen City. The payment change consisted in a shift from almost full capitation to a roughly equal mix of capitation fees and fees for services.
The control group in the natural experiment was made up by the population of GPs in the (sub)urban area of Copenhagen County, surrounding Copenhagen City. Data on complaints presented to them and on diagnoses made by them have not been collected. Data on their numbers of consultations, services and referrals are a clerical by-product of billing procedures under the system of payment of GPs and specialists. The GPs' payment system introduced in Copenhagen City was operative in Copenhagen County all the time.
Effects of the payment change in Copenhagen City are assessed by comparing changes in Copenhagen City with those in Copenhagen County. Effects are generalized by relating the data collected before change in Copenhagen City to data collected in the Netherlands, where an almost full capitation system is still operative for publicly insured patients. The

Dutch data are from the Dutch National Study on morbidity and interventions in general practice (Bensing, Foets, Van der Velden & Van der Zee, 1990). In that study, a sample of Dutch GPs recorded the contents of their consultations for a period of three months. Again, the contents recorded were on complaints, diagnoses, services and referrals.
The design of the study and comparisons made are summarized in figure 1.1.

Figure 1.1: Design of the study



In the opening section it was argued that cross-sectional comparisons are often a poor substitute for natural experiments. Differences in payment of providers often coincide with differences in health status or insurance of patients. In the cross-sectional comparison 1 in the figure, Copenhagen City before change is compared to Copenhagen County. In that comparison, payment systems of providers differ while insurance systems do not: health authorities pay doctors differently for their services to patients equally insured. Since Copenhagen City and County are one (sub)urban area, health status of the population may not differ much either. Comparison 1 is made to compare cross-sectional and natural experiment results. Natural experiment results come from comparing the changes assessed in comparison 2 with the ones assessed in comparison 3.
In both instances, numbers of services and referrals are compared leaving aside complaints presented and diagnoses made, since clerical data on complaints and diagnoses have not been collected. Next, concerning services comparison 2 is elaborated, relating the change in performance of a service to the certainty among GPs about when to perform it. That certainty is assessed as a characteristic of the relation between performances of the

service and complaints presented or diagnoses made. Thus, certainty about standards is included in the design. Certainty about task definitions is left aside.
To generalize effects assessed in Copenhagen City to the Netherlands, relevant institutional and distributional differences should be taken into account. Effects of these on numbers of services and referrals are investigated cross-sectionally in comparison 4.
Certainty about standards as a characteristic of medical culture may vary with local settings as well. Thus certainty among Copenhagen City GPs cannot be attributed to Dutch GPs a priori. Certainty among Dutch GPs is assessed from their performance of services relative to complaints and diagnoses, as measured before the intended change in the Netherlands. Finally, relevant characteristics of settings are taken into account when the effects as assessed in comparison 2 are generalized into the effects predicted to occur in the Netherlands.

1.5 Contents of the book

The payment of GPs by a roughly equal mix of capitation and fee-for-service was introduced in Denmark in 1961, except for Copenhagen City. In Copenhagen City it was not introduced until October 1987. In **Chapter 2** this mixed payment system is described in detail, as is its broader institutional setting.
In **Chapter 3**, GPs' services and referrals in Copenhagen City and Copenhagen County in March 1987 are compared. In **Chapter 4**, changes in Copenhagen City from March 1987 to March 1988 and to November 1988 are compared to those in Copenhagen County. From that comparison we will conclude on the effects of introducing fees for services generally in the Danish institutional setting.
Such effects are expected to occur here, according to a model of physician behavior that includes time and money as utility arguments. According to that model these arguments are constrained by certainty about task definitions and standards. Certainty about whether a service is included in GPs' tasks was not measured. Certainty about a standard includes certainty regarding when to perform a service, depending on patients' state of health. These states of health were measured by complaints and diagnoses. In terms of these, in **Chapter 5** a measure of certainty among GPs concerning a service is proposed. Using that measure, the prediction is tested that services about which GPs are more certain are indeed performed more often when fees are introduced, but to a smaller degree than services about which GPs are less certain.
To generalize effects to the Netherlands, institutional and distributional differences should be taken into account, as well as the nature of the complaints presented. Now generalizing changes in the performance of services to the Netherlands per morbidity category and then aggregating over categories again would be a cumbersome enterprise indeed. Therefore in **Chapter 6** it is checked whether one can at least neglect morbidity in comparing the performance of services in Copenhagen City before change and urban Holland. Fortunately, one can. It follows that differences in numbers of services can be attributed to institution-

al and distributional differences, leaving aside which complaint provokes a service.
Next, an inventory of institutional and distributional differences is produced by applying a heuristic. Assuming that the inventory is exhaustive, it is concluded by implication that relative referring opportunities are decisive for relative numbers of both services and referrals, before fees are introduced. In Section 1.3 these relative referring opportunities were represented by specialist-to-population ratios. We argued that a lower ratio will limit the referral reducing effect of introducing fees for services.
With that, the ground is laid for generalizing effects on services and referrals, as assessed in Copenhagen City, to the Netherlands. In **Chapter 7** this is done by estimating regression equations from Copenhagen City data, and by inserting Dutch values of independent variables into these. Predictions on services and referrals in the Netherlands follow. Dutch values for GPs' certainty about when to perform a service are assessed in **Chapter 8**. Thus Chapter 8 serves as an appendix to Chapter 7.
In conclusion, in **Chapter 9** we relate our findings to the scientific and policy contexts described above. In doing so we return to the questions asked before. In particular, we hypothesize on two questions left unanswered in the study: Will quality of care be maintained when fees for services are introduced? Does this depend on monitoring by standards?

2. The Danish example: a mixed system of fee-for-service and capitation payment*

2.1 Introduction

The payment system for general practitioners has been the subject of discussion in the Netherlands for some years now. Thus in 1984 a discussion was conducted in the Dutch general practitioners' journal 'Huisarts en Wetenschap' between Derksen as proponent of the capitation system in operation for publicly insured patients and Van der Wal as proponent of a differentiation of the capitation system (Derksen, 1984; Van der Wal, 1984). The latter proposed a system in which the amount of the capitation fee reflects the amount of work involved in caring for groups of patients.
At an earlier date the Dutch GPs' association suggested that payment for publicly and privately insured patients be put on a uniform basis. This payment system should be capitation based, possibly differentiated by the amount of work entailed by different groups of patients (LHV, 1981). In recent years the stress in the discussion has shifted towards greater attention to payment for separate services, in addition to capitation payment: the mixed-payment system (Commissie, 1987; LHV, 1987).
As an example of a mixed system, reference is often made to the payment of GPs in Denmark. There the GPs acquire about half of their income from a fixed amount per patient per quarter, and about half from payments for services that are billed monthly. As regards health care as a whole, the similarities between Denmark and the Netherlands seem greater than the differences. As regards the position of the GP, the principal differences are in the broader field of activity of the Danish GP and in his broader and longer training.
The purpose of this chapter is to give the reader more insight into the way in which and the conditions under which the mixed system functions in Denmark. To that end, first we offer a short description of the country and its health care system.

2.2 Denmark and its health care system

Denmark has a population of some 5.5 million, of whom about 40 percent live in and around Copenhagen. The rest of the country is much more thinly populated; the average population density is 121 inhabitants per sq.km. By way of comparison: the population density of the Netherlands is 421 inhabitants per sq.km.
The percentage of inhabitants aged 65 years and older is approximately 15, as compared

---

* The text of this chapter was published in Dutch in Huisarts en Wetenschap 1989; 32(6): 212-218.

to over 12 percent in the Netherlands. At birth, the life expectancy of the Danish population is practically identical with that in the Netherlands, whereas the (unstandardized) mortality is somewhat higher.
The administrative organization of the country has three levels. In addition to the central government there are the 'amten' or counties, which may be compared to the provinces in the Netherlands. There are fourteen 'amten' and two large municipalities that in practice have the same function (Copenhagen and Frederiksberg). The lowest administrative level is the 'kommune' or municipality.

Danish health care does not differ much from the Dutch system. This also holds for the position of the GP (Bentzen, 1983; Scherjon, 1983; Groenewegen & Willemse, 1987).
Since 1961 medical insurance has been of the nature of national insurance, implemented by the county authorities and financed by income-dependent taxation. There are two categories of insured persons; since 1973 one has been free to choose the category to which one wishes to belong. For 'group 1' insured persons medical care is free of charge, but the freedom of choice is limited. One has to register with a GP, and specialist care is accessible only after referral. In principle one can change one's GP once a year. About 93% of the population belong to group 1. 'Group 2' insured persons are not registered with a GP. They can call on specialist assistance directly, but for medical care outside the hospital they have to pay about half themselves.

Nearly 3000 GPs practice in Denmark; the number of inhabitants per GP (roughly the average practice size) is about 1750. In the Netherlands the average practice size is some 500 patients higher. Approximately 40 percent of Danish GPs work in a single-handed practice (and some 60% in the Netherlands).
Specialist care is only to a small extent provided by specialists in private practice. The physicians employed in hospitals are all on the salaried staff.
The number of hospital beds is larger than in the Netherlands: 6.9 per 1000 inhabitants in 1986, as against 4.7 in the Netherlands. With a few exceptions (such as the university hospital of Copenhagen), hospital care is the responsibility of the 'amten'. A planning procedure exists for the hospital sector, in which the counties make a plan for the facilities in their area. That plan contains a description of the existing structure, the developments to be expected as regards the content of the care, and the need for specialized departments, personnel and investments. The central government can issue directives for the county plans. In recent years directives have accordingly been given to the counties to concentrate on coordination of hospitals within a county and on shifting from clinical to ambulatory care. For expansion of the medical staff of hospitals the counties must always seek the advice of a national committee. This advice is always followed in practice.

Hospital use differs from that in the Netherlands. The number of admissions per 1000 inhabitants in 1986 was considerably higher in Denmark (203 against 107 in the Netherlands). This was accompanied, however, by a shorter average length of stay (10.2

days against 12.3 in the Netherlands).
The training of Danish GPs has been considerably lengthened in stages over the last two decades. Since 1983 it has been necessary after graduation first to follow an 18-month course: 12 months of internship in prescribed hospital wards and 6 months in a hospital ward at choice or in a general practice. To be permitted to work as a GP, one must then do a further 60 months of in-service training, including at least 12 months in a general practice.
The GPs are independent professionals, just like the physiotherapists and the dentists. As regards policy and organization, these independent professionals are the responsibility of the counties. The administrative settlement of their accounts takes place at county level; the establishment policy of GPs is also regulated at that level. The municipalities are responsible for preventive public health care, home nursing and social work as parts of primary care. The providers of care in these fields are employed by a municipal service for public health and social work. Within primary care there is in practice a rather pronounced separation between preventive and curative kinds of care. The GP is the linking element because he has contacts with both kinds of care and is active in both.
The tasks of the GP include perinatal care (if necessary in cooperation with a midwife) and youth health care (in cooperation with a nurse specially trained for this branch of care), conferring with the municipal service for public health and social work (for instance for advice on the indication for admission to nursing homes), and the writing of certificates. In thinly populated areas GPs are also active as school doctors. And, finally, the home nurses work under the supervision of the GPs.

2.3 Payment system and care

In his comparative study of the manner of payment of general practitioners in 16 countries (which did not include Denmark), Glaser (1970) arrives at the following generalizations. Where GPs in their training and activities differ less from specialists, they are more oriented towards technical care than personal care. And technical care can be better 'itemized', i.e. divided into clearly defined subactions, than personal care. As a result, this orientation forms a favorable breeding ground for a system of payment per subaction, per service. Conversely, this kind of payment system continues the orientation towards technical care.
In the Dutch discussion too this contrast plays a part. "Doctors suffer from a tremendous overrating of laboratory medicine", and "The patient too does not think much of the art of 'masterful inactivity', of waiting and observing", but "It's the smears and blood samples and so on that make things so expensive", are some remarks by A.J. Dunning in an interview some time ago (Elseviers Weekblad, 7 January 1989). When assessing the functioning of the Danish mixed payment system, it therefore seems important whether its functioning also reflects a mainly technical orientation.

Before taxes, the average income of the GPs in 1985 was Kr. 733,670, in Danish crowns*. In a national agreement it has been laid down that all income elements are adjusted to the price level every six months.
The amounts per income element are so set that the costs of running an average practice will amount to 25% of the total income. In the regular renegotiations it is investigated whether the various standard percentages have been attained; if not, the amounts per income element are adjusted accordingly.
The total income consists of a capitation part and a fee-for-service part. The capitation part is made up by capitation fees of Kr. 42 (in 1985) per registered 'group 1' insured person per quarter. The service part is formed by fees per service. Every doctor-patient contact is called a 'basic service' and is paid for separately. Some actions that fall within a doctor-patient contact or proceed from it are, moreover, subject to extra payment. These are called 'additional services'.
The amounts per capitation and per basic service are so set that the income elements from capitations and basic services in an average practice will be of the same size. The amounts per additional service are so set that the income element from additional services is 7.3 percent of the income element from capitations.

2.4 The agreement

The agreement in which these norms are laid down and worked out is concluded between the Danish association of GPs (Praktiserende Laegers Organisation, PLO) and the Medical Insurance Negotiating Committee, on behalf of the authorities. Application of the agreement is then a task for the county authorities. These cooperate for this with the PLO in county Cooperative Committees.
The tasks of these county Cooperative Committees include assessing the returns by individual GPs of services to be remunerated. For this purpose the averages per patient in each practice are compared with the county averages. In this comparison two norms are used. On the one hand the total expenditure per patient may exceed the county average by no more than 25 percent, and on the other the number of basic services (i.e. doctor-patient contacts) may exceed the county average by no more than 40 percent. If these limits are exceeded, the cooperative committees institute an investigation. If this investigation does not lead to a satisfactory result, the Cooperative Committee can impose a punitive deduction on a GP.
Thus the statistical surveys of the Medical Insurance Negotiating Committee are used in the application of the current agreement. These surveys also play a part in the renegotiations of the agreement, which were carried on every two years until recently. It may for instance prove that certain norm percentages are not attained. It may also prove that a service or a

* 1985 is the most recent year for which financial data are available (Van der Zee, 1985; SFU, 1987). Mid 1985, exchange rates for Danish crowns in guilders, pounds and dollars were Kr. 1.00 = f 0.31 = £ 0.07 = $ 0.09.

group of services is performed more frequently or less frequently than the authorities deem desirable. That may then lead to a change in fees.
The principal points of departure of the parties in these negotiations are the following:

- GPs' association and the authorities agree that the GP income to be expected from services must be equal to the income from capitations. It is assumed that as a result the advantages of the two payment systems are combined: an active primary care thanks to fees for services, and broad attention to the patient thanks to capitation fees.
- With regard to the distribution of the income from services over the fees per separate service, there is a difference of opinion: the authorities want to see the relevance of a service to public health and social well-being reflected in this; the PLO wants it to reflect the time taken up by a service, its complexity and its cost.

2.5 Services

Of the services that we describe in this section, we in all cases state the fees (or their extremes) and relative frequencies. The fees are the ones agreed to be paid per October 1987, and are in Danish crowns*. The frequencies relate to 1985, and are based on the total of 23,294,990 basic services performed in the area to which the national agreement applied. The frequencies are in numbers per person or per 1000 contacts, expressed to the nearest whole number. This precision meets our descriptive purposes. In the numbers per 1000 contacts, consultations in office and house calls are in the denominator.

2.5.1 Basic services

A distinction is made between four basic services. These are listed in Table 2.1.

Table 2.1: Basic services

Service	Fee	Number per insured person [a]
Consultation in office	Kr. 37.50	3
Telephone consultation	Kr. 21.00	2
Repeat prescription	Kr. 17.05	1
House call	Kr. 65.85	1
Total		7

[a] 'Insured persons' are persons of 16 years and over; children up to 16 years are insured together with their parents/guardians. 1985 figures.

* In October 1987, exchange rates for Danish crowns in guilders, pounds and dollars were Kr. 1.00 = f 0.29 = £ 0.09 = $ 0.14.

The amounts for practice consultation, consultation by telephone and house call include writing a prescription or a referral. The fees for house calls are differentiated by the distance to be covered; the amount stated here applies to a distance of 0-4 km.
All amounts stated apply to contacts 'during office hours'; outside office hours a triple fee is charged on weekdays from 4 to 8 p.m. and on Saturdays from 8 a.m to 8 p.m.; on Mondays to Saturdays from 8 p.m. to midnight and on Sundays and public holidays from 8 a.m. to midnight 3.5 times the fee is charged; from midnight to 8 a.m. a fourfold fee always applies.

2.5.2 Additional services

The additional services are divided into six groups. These are reported in Table 2.2.

Table 2.2: Groups of additional services

Group of services	Number per 1000 contacts
Diagnostic services	317
Perinatal care	83
Curative services	66
'Contraceptive' care	35
Socio-medical cooperation	6
Certificates	1
Total	508

There are 43 separately remunerated diagnostic services. These also include preparations for diagnostic examinations in general practice or elsewhere. One of those preparations is the service with the highest fee of this whole group: for a biopsy for microscopic examination by the pathologist the fee is Kr. 131.25. However, this service is performed only once per 1000 consultations and house calls. Urine tests with sticks have conversely a very low fee, as Table 2.3 shows.

Table 2.3: Diagnostic services

Service	Fee	Number per 1000 contacts
Urine test with sticks	Kr. 8.30	48
Taking a blood sample	Kr. 32.80	39
Measuring haemoglobin	Kr. 16.65	35
Urine microscopy	Kr. 24.95	26
Cervical smear	Kr. 32.80	25
Other	Kr. 8.30 - 131.25	144
Total		317

Next, perinatal care comprises examination of expectant and young mothers, and examination and vaccination of young children, as shown in Table 2.4.

Table 2.4: Perinatal care

Service	Fee	Number per 1000 contacts
Vaccinations of young children	Kr. 28.80	36
Examinations of young children	Kr. 127.30	31
Examinations of expectant/young mothers	Kr. 62.10 - 189.05	16
Total		83

The vaccinations of young children comprise three vaccinations against whooping cough, three combined vaccinations against diphtheria, tetanus and polio, three oral vaccinations against polio separately, and three combined vaccinations against measles, mumps and rubella. The first of these takes place when a child is five weeks old, and the last when it is eleven or twelve years old. One sees that Denmark differs in this from the combination of diphtheria, whooping cough, tetanus and polio vaccinations that are usual in the Netherlands.
Young children are given a total of eight examinations: the first when the child is five weeks old, the last when it is five years old. Expectant and young mothers in Denmark receive five examinations: the first of these takes place as soon as possible after the start of pregnancy, and the last nine weeks after delivery. These examinations include the taking of blood and urine samples and making a urine culture. The successive examinations of expectant and young mothers differ in content and thus in fee; the amounts stated are the extremes.
Next, there are 27 curative services that are separately paid for. In Table 2.5 the five curative services with highest frequencies are listed one by one.

Table 2.5: Curative services

Service	Fee	Number per 1000 contacts
Removing warts	Kr. 98.45	14
Removing ear wax	Kr. 65.65	13
Applying an immobilizing bandage	Kr. 65.65	9
First treatment of large wound	Kr. 131.25	8
Removing foreign bodies from eye/ear/nose/throat	Kr. 65.65	4
Other	Kr. 65.65 - 853.65	18
Total		66

Of these curative services, assistance with delivery is remunerated with the highest amount: with manual removal of the placenta Kr. 853.65, and without that operation Kr. 525.50. However, assistance with delivery by the GP hardly occurs any more (0.02 times per 1000 consultations and house calls). The next highest fee for a curative service is Kr. 262.55. That is the payment for treatment of for instance a major fracture or dislocation. But these more major curative services are also relatively rare: they are together performed only once per 1000 consultations and house calls.
The remunerated services that we here jointly call 'contraceptive care' total five in number; we list them in Table 2.6.

Table 2.6: 'Contraceptive' care

Service	Fee	Number per 1000 contacts [a]
Check-up on the use of contraceptives	Kr. 32.80	22
First instruction in the use of contraceptives	Kr. 65.65	8
Introduction of an intra-uterine device	Kr. 131.25	3
Examination for abortion or sterilization	Kr. 78.15	2
Counseling on abortion or sterilization	Kr. 16.40	-
Total		35

[a] In this and the following tables, a dash represents a number per 1000 contacts smaller than 0.5.

'Socio-medical cooperation' is the collective term for conferring with a municipal service for public health and social work. The services included are reported in Table 2.7. The fees for conferring at the municipal service are differentiated by the distance to be covered; the amount stated here applies to a distance of 0-4 km.

Table 2.7: Socio-medical cooperation

Service	Fee	Number per 1000 contacts
Telephonic conferring from municipal service	Kr. 21.00	4
Conferring with municipal officials at practice office, per first patient, for 15 minutes maximum	Kr. 30.20	1
Conferring at municipal service, with the same provisions	Kr. 44.65	-
Conferring per successive patient discussed	Kr. 30.20	-
Extra per patient discussed for more than 15 min.	Kr. 65.65	-
Total		6

The last group of remunerated services is writing certificates. These services are listed in Table 2.8.
As are perinatal care and socio-medical cooperation, the writing of certificates is an additional regular task of the Danish GP, compared to his Dutch colleague. However, per 1000 consultations and house calls, writing a certificate does not occur more than once. Relatively rare as this group of services is, they are remunerated with relatively high amounts. Those amounts are higher according as a certificate requires more extensive diagnosis.

Table 2.8: Certificates

Service	Fee	Number per 1000 contacts
Death certificate	Kr. 78.15	1
Certificate for short-stay intramural psycho-social care	Kr. 144.00	-
Certificate for intramural care and treatment of addiction	Kr. 118.00	-
Certificate for first admission to a psychiatric institution	Kr. 288.00	-
Certificate for re-admission	Kr. 107.00	-
Certificate for compulsory admission	Kr. 158.00	-
Total		1

The first three groups of additional services (diagnostic services, perinatal care, curative services) contain more technical actions, while the last three ('contraceptive' care, socio-medical cooperation, certificates) contain more person-oriented actions ('instruction', 'counseling', 'conferring with social work', 'psychosocial care', 'care of addiction'). Further, we see that the additional services in the first three groups are performed more often than those in the last three. Technical actions seem to be performed more often than person-oriented ones.
In the functioning of the mixed Danish payment system a mainly technical orientation does in fact seem to show, as we expected from Glaser's generalizations.

2.6 Advantages and disadvantages

Change in the payment system for GPs has broad support in the Netherlands at present. Both parties involved in the change think that they can achieve their aims with it. However, the pursuit of goals by other parties may preclude that, or render it invisible.

2.6.1 Authorities

Viewed from the angle of the authorities, the goal of a change in the payment system is to bring about shifts from specialist to primary care. Of the groups of services in the previous section the diagnostic and curative services seem to offer the greatest possibilities in that respect. However, this should be put into perspective. If we compare those two groups of

services on their relative frequency in Denmark, we see an almost fivefold preponderance of diagnostic over curative services: 317 against 66, per 1000 consultations and house calls.
Here a more complete quotation from the interview with Dunning cited above is called for: "Doctors suffer from a tremendous overrating of laboratory medicine, while it has long been proved that all those examinations do not contribute to decision-making and merely exclude other possibilities. It is rather like shooting with pellets while a good doctor has to use his eyes and ears and head to be able to fire a well-aimed bullet with the right weapon". Perhaps that explains the fivefold preponderance just mentioned.
Superfluous diagnostics is undoubtedly less expensive in primary care than in specialist care. However, if that shift occurs through a change in payment system as a result of which a technical orientation is reinforced that in turn strengthens the tendency towards superfluous diagnostics, cost control has moved far out of sight. Nevertheless, it was the aim of the whole shift.

2.6.2 General practitioners

Dutch GPs as a professional group pursue two goals with a change in the payment system. One goal is standardization, by which the differences between publicly and privately insured patients disappear and the administrative rules become the same for all patients. The other goal is the establishment of a more direct relation between the amount of work and earnings. The GP who does more (of the separately remunerated services) is directly paid for that. We might add: also if at the same time he performs fewer services that are not separately remunerated.
Here we touch on one of the essential points of the 'theory' behind the mixed system. That is: if GPs are paid extra for certain services, they will perform those services more often. But, following the same reasoning, will services that are not additionally remunerated in such a situation not be performed less often?
Viewed from the quality of the care, this is a relevant question. The danger exists that some services will be performed too often and others too infrequently. In that case it will be precisely the services that are more difficult to itemize (and for which it is more difficult to reach agreement on appropriate payment) that will be performed too infrequently. While it is precisely those services that can convert that burst of pellets mentioned above into a well-aimed shot.
Perhaps this can be avoided if the introduction of a new payment system is linked with the development and introduction of protocols and standards and with testing and (refresher) training. But then it may be wondered whether this could not lead to the desired aims of substitution (authorities) and recognition and motivation (professional group) even without a change in payment.

2.6.3 Patients and specialist care

As for specialist care, it is relevant to note that in Denmark the hospital specialists are in paid employment. As a result they have a much more indirect interest in the turnover than

Dutch specialists, who are paid per service or per case of disease. In the Netherlands, while referral rates of publicly insured patients remained the same, specialist production nevertheless increased through a larger number of chargeable services per referred patient.
As regards the patients, the Danish situation is comparable with the situation of publicly insured patients in the Netherlands. Services are supplied in kind with few if any copayments. Now if a change in payment system is accompanied by changes in conditions for patients, it becomes difficult to trace the effect of the payment change. A temporary switch from delivery in kind to restitution (the patient pays the GP and charges the amount to the insurance agency) had direct consequences in Denmark for the size and nature of the complaints presented at the first-aid departments of hospitals. When a conflict broke out between the GPs and the authorities at the end of 1984, the patients had to pay the fee direct to the GP, and they could charge the amount later to the insurance agency. In this period the number of visits to first-aid departments of hospitals clearly increased, whereas the nature of the complaints presented there shifted to a larger proportion of acute but minor diseases (Krogsgaard et al., 1985; Wolthers & Stellfeld, 1986).

2.6.4 Routine data collection and costs

Quantitative testing of production takes place in the Danish system in a way that corresponds to the activities of advisory physicians with the Dutch health insurance agencies. An advantage of a partial fee-for-service system is that data are collected on a larger number of aspects of general practice on a routine basis. For all separately remunerated services are also administered. In the present Dutch payment system for publicly insured patients, only data on referring and prescription of medicine can be used for testing. This means at the same time that only few data are available for health care research.
A partial fee-for-service system has the advantage of more data on the activities of GPs. It had the disadvantage of the costs for the administrative settlement of claims and the check on these. We have been unable to acquire data on the size of these costs in Danish health care. In his literature study on the payment of GPs, Janssen (1988) states that good use of personal computers could keep the registration side of the costs within acceptable limits. However, that is only part of the costs. Capitation payment has in any case the advantage of administrative simplicity.
Finally, separate payment of a number of services could also have consequences for the cost budget of a general practice. At a higher frequency of services the variable costs increase, whilst for the performance of some services -- for instance making ECGs -- non-recurrent investments are also necessary. The required investments could lead to joint practices becoming more attractive to GPs.

2.7 Conclusion

This discussion does not lead to a conclusion that clearly works out to the advantage of one system or the other. On the one hand, there are uncertainties about the occurrence of

the desired aims (i.e. shifts from specialist to primary care) and on the administrative costs that a change will bring in its train. On the other hand, to maintain quality of care it seems necessary to incorporate a number of safety valves by training, refresher training and testing.

3. Capitation and fee-for-service in Danish general practice*

3.1 Introduction

Recently Sandier compared health services utilization and general practitioners' earnings in a number of OECD countries. In these countries different methods of paying general practitioners (GPs) are in use. One conclusion drawn by Sandier is that the effect of the method of GPs' payment on utilization seems to be outweighed by other factors. These are the coverage of expenditure by health insurance agencies and their powers of intervention, morbidity in the population, medical technology, and the country's economic level. Another conclusion she draws is that, if these other conditions are equal, fee-for-service payment seems to allow a GP more room to increase his income than capitation payment (Sandier, 1989: 45, 42). In other words, only if these other conditions are equal will fee-for-service payment display the stimulating effect on the performance of services that is attributed to it in discussions of supplier-induced demand (Evans, 1974).
Most of the comparisons made by Sandier are between countries. Two exceptions are Denmark and the Netherlands. In both countries, two health care systems exist alongside each other. For both countries, a comparison between the two systems seems to support the conclusion "that the level of consumption is influenced more by the fact that it is free of charge to the patient than by the method whereby physicians are paid. In these two countries, the number of general practitioner services per insured person using a physician paid a capitation and receiving their treatment free is higher than that for persons covered by the other system, who are generally in better health, consult physicians on a fee-for-service basis, have to pay the physician themselves, and bear part of the cost" (Sandier, 1989: 37). Hence this is an instance in which the effects of coverage of expenditure by health insurance agencies and morbidity seem to outweigh the effect of the method of GPs' payment.

Here one more comparison is added to the ones Sandier reports. It takes advantage of the fact that until October 1987 in Denmark the part of GPs' income made up by capitation for patients receiving their treatment free differed between Copenhagen City and the rest of the country. In Copenhagen City it made up the major part of GPs' income, while in the rest of Denmark a lower capitation made up half the income of an average GP, by prospective estimation. In both cases the rest of GPs' income was earned by fees for services. For the group of patients receiving their services in kind (about 95% of the population), these fees

---

* The text of this chapter has been published in International Journal of Health Sciences 1991; 2 (1):23-28.

were paid directly by health insurance agencies to GPs*.
The set of services that fees were paid for was smaller in Copenhagen City than in the rest of Denmark. In Copenhagen City fees were paid only for perinatal services (preventive examinations and vaccinations), contraceptive services (instruction in the use of contraceptives, among other things), and the writing of certificates (of death, and for admission to mental health institutions). In the rest of Denmark fees were also paid for each consultation, for a set of 43 diagnostic services (such as urine analysis), and a set of 27 curative services (such as removing warts), and for conferring with a municipal service for public health and social work.
Of the two conditions of reimbursement mentioned above, the one without out-of-pocket payment by the patient (called group 1), covers by far the largest part of the population. If services per person under this condition of reimbursement in Copenhagen City and the rest of Denmark are compared, the effects of differences in health insurance coverage are ruled out. To some degree, so are the effects of differences in morbidity. However, morbidity varies not only with insurance status, but also with urbanization: more urbanized areas often have higher morbidity (e.g.: Van Sonsbeek, 1989). And Copenhagen City is much more urbanized than the rest of Denmark. This difference in urbanization is reduced when one compares Copenhagen City with the rest of Copenhagen County: Copenhagen County is the larger (sub)urban area of which Copenhagen City is the central part**. If services per 'group 1' insured person in Copenhagen City and the rest of Copenhagen County are compared, the more obvious conditions interfering with the effect of fee-for-service payment are equal. In this comparison, the difference between the two payment systems should have clear effects in numbers of services per person. What effects can be expected then?

1. In the county but not in the city, a fee was paid for each consultation. The effect of that would be that once a patient consults his GP, a county GP would be expected to arrange a return consultation more often. Thus if numbers of initial consultations per person were equal in city and county, **total numbers of consultations per person should be larger in the county**.
2. In the county but not in the city, fees were paid for separate diagnostic and curative services, and for conferring with a municipal service. Such conferring does not take place very often, as we saw in the previous chapter; it is not considered further here. For diagnostic and curative services the expected effect is unconditional: **numbers per person of diagnostic and curative services should be larger in the county**.
3. Both in the county and in the city, fees were paid for perinatal services, contraceptive

---

* Danish health care is outlined in Raffel (1984) and in Krasnik et al. (1982). Details of the reimbursement of GPs may be found in the agreement between GPs' professional organization and health authorities (PLO, 1988).

** Among the other municipalities in Copenhagen County, in Taarnby and Dragør the payment system for GPs was the same as in Copenhagen City until 1987. Here Taarnby and Dragør will be left out of comparisons.

services and certificates. Certificates are seldom written; again, they are not considered further here. For perinatal and contraceptive services, from the fact that in both the city and the county fees were paid for them, **numbers per person of perinatal and contraceptive services should be equal in city and county**.

4. 'Group 1' insured Danes have free access to their GP, but can visit a medical specialist in hospital or private practice only after referral by their GP. Neither GP nor patient runs any financial risk from such a referral. Now it could be that return consultations with a GP and diagnostic and curative services by a GP make such a referral redundant. If they do, to a city GP referring a patient only implies a gain of time, while to a county GP it implies a loss of income as well. Thus if they do, and if numbers of consultations, diagnostic and curative services were larger in the county, **numbers per person of referrals should be smaller in the county**.

Here we shall compare consultations, services and referrals per 'group 1' insured person in Copenhagen City and the rest of Copenhagen County. We shall check whether numbers per person were as expected, before October 1987. Until then, the differences in GPs' payment described were in existence. Since October 1987, Copenhagen City GPs have been paid the same way as other Danish GPs, with capitation making up half their income by prospective estimation.

3.2 Data and method

3.2.1 Sources

The data used in our comparison are from a number of different sources. The data on Copenhagen County are of administrative origin. Those on Copenhagen City were collected to investigate the effects of the change in GPs' payment there. Our data are on the situation six months before that change.

For Copenhagen County, the data on consultations and services by GPs and on referrals to specialists in private practice are from Copenhagen County health authorities, while data on referrals to hospitals are from hospital administrations. The health authorities data stem from billing by GPs and private specialists respectively.

For Copenhagen City GPs at the time of our comparison no billing was in order yet for consultations and diagnostic and curative services. A sample of Copenhagen City GPs voluntarily recorded all consultations for one week each*. The recording sheet covered type of consultation, perinatal and contraceptive services, a sample of diagnostic and curative services, and referrals.

---

* A comparison of available characteristics of participating and non-participating GPs showed that they did not differ in gender, but that participating GPs were more recently graduated.

3.2.2 Time span

The Copenhagen City GPs recorded their consultations for one week out of weeks 9 through 11 of 1987. For Copenhagen County, administrative data are used that cover time spans as similar as possible.
Hospital administrations produce data on periods of one week; data on referrals to hospitals are on week 10 of 1987. Health authorities produce data on periods of four weeks; data on consultations and services by GPs are on weeks 9 through 12 of 1987. Data on referrals to private specialists are on the preceding period, of weeks 5 through 8: health authorities judged the billing data of private specialists on weeks 9 through 12 to be unreliable*.
All data have been converted into average numbers per week.

3.2.3 Level of aggregation

For Copenhagen City data are available at the level of consultations.
For Copenhagen County, data as produced by health authorities and hospital administrations are available only at a higher level of aggregation. Data on consultations and services by GPs as produced by health authorities are available at the level of groups of GPs in 'practice areas'. So are data on referrals to hospitals as produced by hospital administrations. Data on referrals to private specialists as produced by health authorities are available at the level of municipalities. Therefore, data on all variables are available only at the level of regions that include both entire practice areas and entire municipalities.
Copenhagen County includes 11 GPs' practice areas and 16 municipalities**. From these, 8 regions can be constructed. These regions include a minimum of 10, a maximum of 82, and an average of 41 GPs.

3.2.4 Coverage

For Copenhagen County, administrative data are available on all consultations, all services that fees are paid for, and all referrals by the 326 GPs practising at the time of investigation.
For Copenhagen City our data are on a sample of GPs, who recorded only the diagnostic and curative services included in the recording sheet.
In Copenhagen City 283 GPs were practising at the time. All of them were invited to take part in the investigation mentioned. Finally, 87 of them participated.
As mentioned, outside Copenhagen City fees were paid for 43 diagnostic and 27 curative services. From these, a sample of 13 diagnostic and 8 curative services were included in the Copenhagen City GPs' recording sheet. These sample services were billed frequently in the county, and it was judged that the average Copenhagen City GP would have the necessary equipment to perform them.

---

* In weeks 9 through 12 of 1987 the numbers suddenly doubled, after which a conflict broke out between specialists and health authorities.

** Taarnby and Dragør are excluded here.

3.2.5 Biases

The differences in data described may produce a number of biases.

The county data stem mainly from billing and the city data from voluntary participation. Because of that, the city data may be a relative underestimation of real numbers.

Conversely, the city data may be a relative overestimation of real numbers, for two reasons. The city data are on real working weeks of participating GPs; the county data are on fixed administrative periods, which may have included winter holiday weeks of individual GPs. Also, our city sample showed an overrepresentation of recently graduated GPs, who may be more likely to perform a medical service if that service has only recently come into use.

The time spans covered by our data all center around the middle of March, except for referrals to private specialists in the county. For these the time span centers around the middle of February. If anything, chronic disorders turning acute and calling for a referral could be more probable in the more wintery month of February. Yet only a minor bias in numbers of referrals is to be expected from that.

3.2.6 Method

We shall compare numbers per person of consultations, services and referrals in city and county, to see whether the expected effects of fee-for-service payment do show. These effects should show in the behavior of GPs. Hence our units of analysis should be GPs: average numbers per person (per registered patient) should be assessed per GP, and differences between city and county GPs should be tested.

Our county data, however, are on the level of regions. Still a statistical test of differences could be applied if equal variance between city and county GPs were assumed. But making that assumption to test for cross-sectional differences so as to decide on effects of payment systems seems to be introducing rigour where it is not decisive after all. Decisive evidence on effects of payment systems can come only from natural experiments when payment systems are introduced.

We shall compare averages in city and county: averages over GPs in Copenhagen City, and averages over regions weighted by numbers of GPs in Copenhagen County. Nevertheless, we need a standard to decide whether these average numbers are different or equal, and expectations confirmed or rejected.

Such a standard is implicit in other reports on comparative health services research. Usually, the importance of differences in rates is expressed by describing multiples: "There are twice as many surgeons in proportion to population in the United States as in England and Wales, and they perform twice as many operations" (Bunker, 1970); "Surgical rates have been found to vary up to tenfold for different surgical procedures across market areas" (Wolff, 1989). Here we turn this usage into a rule. We consider numbers in Copenhagen City and Copenhagen County to be different if one is at least a multiple of, and thus at least twice as large as the other.

3.3 Results

In Table 3.1 numbers of consultations, services and referrals in Copenhagen City and Copenhagen County are compared. Three types of consultations are listed: consultations in practice and by telephone, and house calls. As for services, sum totals of diagnostic, curative, perinatal and contraceptive services are reported. As for diagnostic and curative services, sum totals relate to the samples of services recorded.
For all three types of consultations listed, numbers per person were less then twice as large in the county as they were in the city. Thus we consider them to be equal. By the same rule, numbers of diagnostic and curative services were larger in the county, while numbers of perinatal and contraceptive services were equal. For referrals both to specialists in private practice and to hospital, numbers of referrals in city and county were equal as well.

Table 3.1: Numbers per week per 1000 'group 1' insured patients on the GPs' lists in March 1987 in Copenhagen City and Copenhagen County

	City	County
consultations in practice	52.0	56.6
consultations by telephone	24.1	35.7
house calls	3.2	2.9
diagnostic services	4.0	12.0
curative services	.9	2.9
perinatal services	4.9	4.0
contraceptive services	2.2	2.5
referrals to private specialists	10.8	14.6
referrals to hospital	2.4	1.4

Hence, numbers of diagnostic and curative services were larger in the county when sum totals of sampled services are compared. In Table 3.2 numbers are compared for each service in our sample.
For all but three of the twenty-one sampled services listed, numbers were at least twice as large in the county, and are therefore considered to be larger there. By the same rule, numbers were equal for taking a blood sample, taking a cervical smear and inoculating for culture.

3.4 Discussion

According to our results, numbers per person of consultations were equal in Copenhagen City and Copenhagen County. We expected them to be larger in the county, if numbers of initial consultations were equal. Numbers of diagnostic and curative services were generally larger in the county, and numbers of perinatal and curative services were equal, as we

expected. Numbers of referrals were equal. We expected them to be smaller in the county, if numbers of diagnostic and curative services were larger in the county (as they were), and if these services would make a referral redundant.

Table 3.2: Numbers per week per 1000 registered patients of sampled diagnostic and curative services

	City	County
diagnostic services:		
taking a blood sample	.30	.56
cervical smear	.89	1.44
pregnancy test	.23	.48
proctoscopy	.01	.16
electrocardiogram	.01	.31
haemoglobin measurement	.51	1.16
ESR measurement	.00	.35
blood glucose measurement (photometer)	.04	.28
streptoculture / urine culture	.13	1.99
inoculating for culture	.64	.72
urine test with sticks	1.13	2.77
urine microscopy	.05	.89
urine culture with sensitivity	.05	.91
curative services:		
removing warts	.14	.82
removing ear wax	.41	.84
removing corpora aliena from eye/ear/nose/throat	.02	.09
removing corpora aliena from skin/nail	.06	.22
incision/excision of abscess/tumour	.08	.36
first treatment of large injury	.08	.19
dressing an immobilizing bandage	.09	.40
bladder catheterization	.01	.02

Thus generally our unconditional expectations are confirmed, and the conditional ones rejected. Apparently the conditions mentioned were absent.
Possibly initial consultations, on which we do not have separate data, were smaller in number in the county. This could be explained by a lower level of morbidity that coincides with a lower degree of urbanization in Copenhagen County. We restricted our comparison to Copenhagen City and the surrounding (sub)urban area. Apparently, this does not rule out differences in demand enough for supply-side effects on total numbers of consultations to show. Apparently supply-side effects on numbers of diagnostic and curative services are stronger.
While numbers of diagnostic and curative services were larger in the county, numbers of referrals were not smaller there. Thus substitution between services and referrals does not show in our comparison.

Of the individual diagnostic services, three were rendered equally often in city and county: taking a blood sample, taking a cervical smear and inoculating for culture. We expected

them to be rendered more often in the county.
Taking a blood sample and inoculating for culture are the two diagnostic services in our sample that do not involve diagnostic testing, but are preparations for diagnostic testing elsewhere. Here a GP has three alternatives: preparation and testing can be done elsewhere; preparation can be done by the GP and testing elsewhere; and both can be done by the GP. If out of these three alternatives the first one is chosen more often by city GPs, and the third one is chosen more often by county GPs, it may well be that the middle one is chosen equally often by city and county GPs, as we found. In introducing our expectations we did not distinguish between these alternatives.
As for taking blood samples, in an earlier comparison of services performed by Copenhagen City and other Danish GPs it was noted that in Copenhagen City blood samples are often taken by the Copenhagen General Practitioners Laboratory (KPL/KKS, 1985). Hence for taking blood samples there is evidence that the first of the three alternatives is indeed often chosen by city GPs.
Taking a cervical smear is the third diagnostic service for which we did not find the difference we expected. Some cervical smears are taken in preventive screening of otherwise healthy women, usually according to pre-established time intervals. And others are taken because of symptoms present. Only in the latter case are supply-side effects plausible. If most smears are taken in preventive screening, total numbers of smears are mainly determined by numbers of women of fertile age. These certainly do not differ in the sense of being at least twice as large, between city and county.

In this chapter we have added one comparison to the ones Sandier (1989) reports, building on her conclusions. For that we used data that were different in a number of ways. These differences may have produced a number of biases. They also excluded rigorous statistical testing. Instead of that, we used a rule taken from usage in comparative health services research. Our results turned out to confirm expectations on services performed by GPs with plausible exceptions. They did not confirm expectations on substitution between services and referrals.
Since they are results from cross-sectional comparison, no strong inferences can be drawn on the effects of payment systems on the behavior of GPs. Stronger evidence on effects of payment systems would come from natural experiments when payment systems are introduced. Results of such an experiment are in the next chapter.

4. Effects of introducing fees for services in Copenhagen, Denmark

4.1 Introduction

As Culyer (1989: 30) stated in a recent review, "it seems fairly clear that fee-for-service methods result in both more active treatment and higher incomes for doctors". As an example of more active treatment, he quoted an example from Quebec, Canada, where a doubling of fees for house calls was followed by some 30% increase in their number, while generally in Canada house calls by community-based doctors declined. Culyer went on to remark that "no evidence for Europe exists that is comparable to that for North America". Recently, for Europe such evidence came from Copenhagen, Denmark. A first report of it was made in Krasnik et al. (1990). Here we elaborate on it.

In all of Denmark a tax based health insurance system is operative in which people can choose once a year between two regulations. In one regulation, one has to register with a general practitioner (GP) and can visit another physician only upon referral by his GP, but then one gets one's health care in kind, with the exception of a few out-of-pocket copayments. By the other regulation, one can freely visit any physician, but then out of pocket payments with partial reimbursement are the general rule. By far the majority of Danes (about 95%) choose the former regulation. So for the majority of Danes, GPs are the gatekeepers of health care: they decide whether a patient should be referred to a medical specialist.
In such conditions, more active treatment by GPs should result in lower referral rates, if GPs' services serve to postpone or avoid referrals to medical specialists.

When the Danish health insurance system was introduced in 1961, with it came a GPs' payment system in which by prospective estimation the average GP received about half his income from capitation for registered patients, and about half his income from fees for consultations and for a set of mainly diagnostic and curative services.
Yet in Copenhagen City this payment system was not introduced until October 1987. Until then, Copenhagen City GPs received the major part of their income from capitation, and got fees only for a far smaller set of services (mainly in the realms of contraception and perinatal care) on top of that.
Now if fee-for-service results in more active treatment, the change in payment in Copenhagen City in October 1987 should have caused an increase both in consultations and in the larger set of diagnostic and curative services. If GPs' services serve to postpone or avoid referrals, the change in payment should have caused a decrease in referrals as well.

A natural control group for testing these effects is made up by the GPs in the remainder of Copenhagen County. Copenhagen County, with about one million inhabitants in all, is the larger (sub)urban area of which Copenhagen City, with about half a million inhabitants, is the central part.
With Copenhagen County as a control group, the expected effects just mentioned were put to the test in the report mentioned. Comparing proportional changes in rates, we found that the change in payment in Copenhagen City caused initial increases in consultations, which continued to exist for consultations by telephone but not for face-to-face consultations in the doctor's office; and that it caused sustained increases in diagnostic and curative services. We concluded that services over which doctors have more discretion have increased more upon the introduction of fees. Also, we found that the change in payment caused delayed decreases in referrals to both privately practicing specialists and hospitals. We concluded that the increased numbers of services performed seem to have replaced referrals to some degree.

Here we will elaborate on the choice of comparing proportional changes in rates, rather than sizes of changes in rates. Comparing proportional changes, a 100% increase from 1 to 2 services per registered patient is considered equal to a 100% increase from 10 to 20 services per registered patient. Comparing sizes of changes in rates, it is considered equal to an increase from 10 to 11 services per registered patient. We will show that, given this choice, comparing proportional changes produces more solid results.
Alongside that, we will show that the conclusions drawn before are fully underpinned if assumptions made in testing for effects are changed.
Also, we will show that if services and referrals are further specified, seemingly random results are obtained. This suggests that clear-cut conclusions can be drawn only at the level of specificity chosen.
Building on that, we address one question that did not come up in the previous report. Does the introduction of fees for services stimulate quantity instead of quality? Does quality of care provided deteriorate as a result? This question can be answered in a limited sense only.
To promote health care quality, health authorities already paid fees to Copenhagen City GPs for the smaller set of perinatal and contraceptive services mentioned. Now it may be that an increased performance of diagnostic and curative services leaves a doctor less time to perform perinatal and contraceptive services. If these are performed less as a result, health care quality has deteriorated in one respect generally recognized.
This is the one aspect of health care quality for which data on the Copenhagen County control group are available. Therefore, it is the only aspect that is put to the test. A comparison of the performance before and after the change of services not separately paid for would have been a better test. However, we do not have data on such services.

4.2 Method and data

4.2.1 Methods compared

In the previous report, changes in a sample of Copenhagen City GPs were tested against simultaneous changes in the population of Copenhagen County GPs. As for method, we estimated time-bound doctor-independent rates of performance of a service by assuming additivity between these and doctor-bound time-independent rates or tendencies. Proportional changes in estimated doctor-independent rates in the city sample were assessed, with 95% confidence intervals. Doctor-independent rates in the county GPs' population were measured by dividing total numbers of services performed by total number of patients registered. If the city sample 95% confidence interval of proportional change lay outside the county population proportional change, we concluded that the difference was due to the change in payment in Copenhagen City.

In doing so we by-passed the problem that county data were not available per doctor, but only at a more aggregate level. Here we by-pass that problem by assuming that variances of changes among county doctors are equal to those among city doctors. A priori, supply and demand side characteristics lead to opposing conjectures relative to that assumption.

As for supply side, the city sample of doctors volunteered to record their patient contacts. From this element of self-selection, a homogeneous sample can be assumed, with variances smaller than those in the county population of doctors.

As for demand side, the period of data collection was shorter in the city. From that, complaints presented can be assumed to show more random variation between city doctors, resulting in larger variances. Hence, the equal variance assumption can not be rejected a priori.

4.2.2 Estimating numbers per doctor per week

In the city, data were collected in three one-week periods, in March 1987 (six months before the change in payment), March 1988 (six months after) and November 1988 (one year after the change) respectively. Seventy-one doctors actually recorded their patient contacts for at least three days in all three weeks; one did so for only two days in November 1988. Data on the latter doctor were excluded from our previous report; the remaining incompleteness was left unadjusted. Here we include data on all 72 GPs, and estimate numbers per 1000 registered patients per full five-day working week for each GP.

On county GPs, data on services performed and referrals made by regional groups of GPs during the months mentioned are available. Regional groups were 8 in number, including 326 GPs in all, with a minimum of 10 and a maximum of 82 GPs per region. We estimated average numbers per GP in the county by averaging over regions, weighting by numbers of GPs. The numbers averaged were numbers per 1000 registered patients per full five-day working week in the months concerned.

4.2.3 Changes studied

In the previous report we studied changes in consultations in office and by telephone, excluding house calls. We excluded house calls since no distinction could be drawn between visits to the patients' own homes and those to an institution. Both differ in that GPs' time costs per patient are smaller when visiting an institution. Still, they are larger than with consultations in the office. Here we include house calls, and examine whether lack of discretion produces results similar to those from consultations in the office.

In the previous report we also studied changes in prescription renewals. We found a decrease in measured numbers, which we suspected to represent a redefinition by GPs rather than a behavioral change. Therefore, we leave them aside here.

We left aside perinatal and contraceptive services, since they were already paid for in Copenhagen City before the payment change. Here we include them to answer the health care quality question asked before.

As we did in the previous report, we include diagnostic and curative services, and referrals to hospitals and privately practicing specialists. Data on referrals relate to all specialties. Data on services relate to a sample of 13 out of the 43 diagnostic services, and to a sample of 8 out of the 27 curative services that fees were introduced for. Services were sampled that an average Copenhagen City GP was judged to be already equipped to perform. If so, the effect of introducing fees for these services would not be curbed by postponed investments.

4.2.4 Method elaborated

The method to be used should enable us to decide whether changes in GPs' behavior are caused by the change in payment system. That change took place in the city, while the payment system in the county remained unchanged. We will assume that this was the only difference in relevant changes between city and county. Thus we will not assume that other things remained equal, but will assume that they changed equally in city and county.

We assume that the changes in city doctors' behavior are the summed effects of both general changes in e.g. health care demand and medical culture, and of the change in payment system that was specific to the city. We will use the changes in county doctors' behavior as an estime of the effects of these general changes.

In doing so we will explore both options mentioned before on what are to be considered equal changes in doctors' behavior. Option 1 says that changes equal in size (in numbers per registered patient) are to be considered equal. Option 2 says that changes equal in proportion are to be considered equal, so that to double or to halve the frequency of a unit of behavior is always as big a change in habit or working style, no matter what number is doubled or halved.

From that, a priori option 2 seems to be the more plausible one: the counting unit of behavior seems to be more relevant to the one counting than to the one behaving. Still, option 1 is the more simple one numerically. We will choose the option that produces more unequivocal results.

Hence to begin with, the change among city doctors to be expected from the general

sources of change will be estimated both from the size of the average change among county doctors (option 1), and from their proportional average change (option 2). The differences between these expected changes among city doctors on the one hand and the measured change among city doctors on the other are then regarded as due to the change in payment specific to the city.
To decide on the significance of these differences, we assume that variances of changes among county doctors are equal to those among city doctors. Given that assumption, Student's t can be applied to decide on the significance of differences. It takes the form of

$$t=((chge_{meas.city}-chge_{exp.city})/sd_{chge.meas.city})*$$
$$*\sqrt{(n_{city}*n_{county})/(n_{city}+n_{county})},$$

where

$chge_{meas.city}$ is the measured average change in the city,

$chge_{exp.city}$ is the expected average change in the city,

by option 1 assessed as average change in the county, and by option 2 as (proportional average change in the county) * (initial value in the city),

$sd_{chge.meas.city}$ is the standard deviation of the measured changes in the city, and

n_{city} and n_{county} are the numbers of doctors in city and county.

In testing differences we will apply the usual 5% significance level. If we find a significant difference between measured and expected changes in the city, we will conclude that the difference in changes is real, and is caused by the change in payment system.

4.3 Results

In Copenhagen City, consultations and diagnostic and curative services have been newly paid for since October 1987. In Copenhagen County they have been paid for since 1961. Does the change in payment in the city cause an increase in these, if the changes are assessed by their sizes irrespective of proportionality? Not unequivocally, as Table 4.1 shows.

Table 4.1: Changes in numbers per week of consultations and services per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	differ-ence in changes	Nov. 1988 mean	change from Mch'87	differ-ence in changes
consultations	city	79.27	89.24	+9.97	p≥.05	84.63	+5.35	p≥.05
	county	95.28	103.39	+8.11		100.94	+5.67	
diagnostic services	city	4.27	6.03	+1.76	p≥.05	7.11	+2.83	**p<.01**
	county	12.01	12.98	+.97		13.29	+1.28	
curative services	city	0.90	1.71	+.81	**p<.01**	1.62	+.72	p≥.05
	county	2.95	3.21	+.26		3.48	+.54	

Table 4.2: Proportional changes in numbers per week of consultations, services and referrals per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	differ-ence in changes	Nov. 1988 mean	change from Mch'87	differ-ence in changes
consultations	city	79.27	89.24	+13%	p≥.05	84.63	+7%	p≥.05
	county	95.28	103.39	+9%		100.94	+6%	
diagnostic services	city	4.27	6.03	+41%	**p<.05**	7.11	+66%	**p<.001**
	county	12.01	12.98	+8%		13.29	+11%	
curative services	city	0.90	1.71	+90%	**p<.001**	1.62	+80%	**p<.01**
	county	2.95	3.21	+9%		3.48	+18%	
referrals to private specialists	city	11.26	9.60	-15%	**p<.05**	8.27	-27%	**p<.001**
	county	14.63	14.30	-2%		14.78	+1%	
referrals to hospitals	city	2.39	2.26	-5%	p≥.05	1.85	-23%	**p<.01**
	county	1.38	1.39	+1%		1.45	+6%	
perinatal services	city	4.89	4.26	-13%	p≥.05	4.64	-5%	p≥.05
	county	5.48	4.31	-21%		4.75	-13%	
contraceptive services	city	2.32	1.92	-17%	p≥.05	1.80	-22%	p≥.05
	county	2.53	2.41	-5%		2.40	-5%	

As consultations are initiated by patients as well as by doctors, it is quite possible that the more active treatment held to be caused by the introduction of fees for services does not

include more numerous consultations. But why should it include an initial increase in curative services that is subsequently replaced by an increase in diagnostic services?
Indeed, such a shift may express learning among GPs about which services serve their preferences for income and leisure best. Yet it seems more apt to look for unequivocal results on diagnostic and curative services generally, before going into these particularities. Such unequivocal results are obtained when proportional changes are compared, as Table 4.2 shows.
Changes in referrals as caused by the change in payment are explicable only if diagnostic and curative services change in number as well. And only then does the question on changes in perinatal and contraceptive services indicating a changing health care quality come up. Therefore, these other changes are included in Table 4.2.

According to both precedings tables, the change in payment in the city caused a unequivocal increase in diagnostic and curative services only if proportional increases are compared. Before, we argued that comparing proportional changes would be more plausible a priori. Now it proves to produce more solid results as well.
According to Table 4.2, as a secondary effect of the change in payment the increase in services causes a decrease in referrals. Yet for referrals to hospitals this secondary effect only arises with delay. On the other hand, according to Table 4.2 GPs refer their patients far more often to private specialists than to hospitals. From that it seems quite plausible that a change in practice style affects their style of referring to private specialists more quickly.
Concerning health care quality, we asked whether an increase in diagnostic and curative services would be accompanied by a decrease in perinatal and contraceptive services. According to Table 4.2, this side effect does not arise.

Next, we investigate what effects are to be attributed to the change in payment if services and referrals are further specified. In Table 4.3, we do so for single types of consultations.

Table 4.3: Proportional changes in numbers per week of single types of consultations per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	difference in changes	Nov. 1988 mean	change from Mch'87	difference in changes
consultations in office	city	51.69	56.71	+10%	p≥.05	53.17	+3%	p≥.05
	county	56.62	60.92	+8%		60.51	+7%	
consultations by telephone	city	24.40	28.52	+17%	p≥.05	27.90	+14%	p≥.05
	county	35.78	39.53	+10%		37.85	+6%	
house calls	city	3.18	4.01	+26%	p≥.05	3.56	+12%	p≥.05
	county	2.88	2.94	+2%		2.58	-10%	

Before, we found that the change in payment in the city did not cause an increase in consultations in general. In Table 4.3 this finding is confirmed with all three types of consultations: neither consultations in office nor those by telephone are held more often, nor are patients visited at home or in an institution more often because of the change in payment in the city.

Table 4.4: Proportional changes in numbers per week of single perinatal and contraceptive services per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	difference in changes	Nov. 1988 mean	change from Mch'87	difference in changes
perinatal:								
examination of mothers	city	1.24	1.10	-11%	p≥.05	1.10	-11%	p≥.05
	county	.76	.80	+5%		.85	+12%	
examination of infants	city	1.36	1.54	+14%	p≥.05	1.60	+18%	p≥.05
	county	1.48	1.48	+0%		1.65	+12%	
vaccination of infants	city	2.29	1.62	-29%	p≥.05	1.94	-15%	p≥.05
	county	3.24	2.03	-37%		2.25	-31%	
contraceptive:								
instruction on use	city	.59	.51	-13%	p≥.05	.47	-20%	p≥.05
	county	.59	.57	-4%		.63	+7%	
checkup on use	city	1.34	1.21	-10%	p≥.05	1.03	-24%	p≥.05
	county	1.64	1.58	-4%		1.48	-10%	
examination for abortion/sterilization	city	.27	.11	-59%	**p<.05**	.23	-15%	p≥.05
	county	.16	.15	-6%		.16	+2%	
introducing an IUD	city	.11	.09	-25%	p≥.05	.08	-32%	p≥.05
	county	.14	.12	-14%		.12	-15%	

We also found that the change in payment in the city did not cause a decrease in perinatal or contraceptive services in general. Findings on single perinatal and contraceptive services are in Table 4.4 above. According to the table, only examinations for abortion or sterilization are performed less often because of the change in payment in the city, and only initially so. Leaving aside this one instance out of fourteen, findings are confirmed for single services: perinatal and contraceptive services are not performed less often.

Diagnostic services in general are performed more often because of the change in payment. For twelve of the thirteen single diagnostic services in our sample, findings are in Table 4.5. ESR measurements are left aside because no proportional increase can be assessed: to begin with, they were not performed at all by the sample of city doctors.

Table 4.5: Proportional changes in numbers per week of single diagnostic services per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	differ-ence in changes	Nov. 1988 mean	change from Mch'87	differ-ence in changes
blood sample	city	.32	.22	-30%	p≥.05	.38	+22%	p≥.05
	county	.56	.60	+8%		.58	+4%	
cervical smear	city	.96	1.32	+37%	p≥.05	1.15	+20%	p≥.05
	county	1.44	1.62	+13%		1.65	+15%	
pregnancy test	city	.26	.26	-0%	p≥.05	.31	+19%	p≥.05
	county	.48	.53	+9%		.52	+8%	
proctoscopy	city	.02	.10	+470%	**p<.05**	.07	+332%	p≥.05
	county	.16	.15	-5%		.18	+15%	
electro-cardiogram	city	.01	.07	+746%	p≥.05	.02	+108%	p≥.05
	county	.31	.28	-8%		.31	-1%	
haemoglobin measurement	city	.54	.78	+44%	p≥.05	.83	+54%	**p<.05**
	county	1.16	1.14	-2%		1.18	+2%	
blood glucose measurement	city	.04	.39	+895%	**p<.01**	.29	+633%	**p<.001**
	county	.28	.36	+29%		.41	+48%	
strepto-/urine culture	city	.14	.25	+82%	p≥.05	.44	+223%	**p<.05**
	county	1.99	2.25	+13%		2.22	+12%	
inoculation for cultivation	city	.72	.89	+23%	p≥.05	1.38	+91%	**p<.05**
	county	.72	.82	+14%		.89	+23%	
urine test with sticks	city	1.15	1.49	+29%	p≥.05	1.85	+61%	**p<.01**
	county	2.77	2.96	+7%		3.05	+10%	
urine microscopy	city	.06	.02	-60%	p≥.05	.10	+57%	p≥.05
	county	.89	.93	+5%		.93	+4%	
urine culture with sensitivity	city	.05	.15	+169%	p≥.05	.20	+275%	p≥.05
	county	.91	.98	+7%		1.01	+10%	

Only with blood glucose measurements is the finding with diagnostic services in general confirmed: only for them did the introduction of a fee cause a sustained increase. With four services only a delayed increase is caused by the introduction of fees; with six, no increase at all.

Like diagnostic services, curative services in general are performed more often because of

the change in payment. For seven curative services in our sample, findings are in Table 4.6. Now bladder catherizations are left out, since they were not performed at all.

Table 4.6: Proportional changes in numbers per week of single curative services per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	differ-ence in changes	Nov. 1988 mean	change from Mch'87	differ-ence in changes
removing warts	city	.17	.42	+149%	**p<.05**	.44	+162%	**p<.05**
	county	.82	1.01	+23%		1.21	+48%	
removing ear wax	city	.40	.60	+47%	p≥.05	.47	+17%	p≥.05
	county	.84	.89	+6%		.86	+1%	
removing corp.al.	city	.02	.04	+67%	p≥.05	.01	-62%	p≥.05
eye/ear/nose/thr.	county	.09	.07	-29%		.07	-21%	
removing corp.al.	city	.05	.05	-10%	p≥.05	.07	+41%	p≥.05
from skin/nail	county	.22	.20	-10%		.19	-13%	
in/excision of	city	.08	.16	+94%	p≥.05	.26	+221%	**p<.01**
abscess/tumour	county	.36	.41	+14%		.44	+23%	
treating a	city	.07	.10	+30%	p≥.05	.13	+75%	p≥.05
large wound	county	.19	.21	+6%		.21	+9%	
dressing an immo-	city	.08	.36	+346%	**p<.01**	.24	+194%	**p<.05**
bilizing bandage	county	.40	.40	+1%		.48	+21%	

Two single curative services show a sustained increase, as curative services in general did. One only shows a delayed increase, and four show no increase at all because of the change in payment.

With findings on single diagnostic and curative services so heavily mixed, one wonders how things are with referrals to single specialties. Owing to the change in payment, patients were referred less often to specialists in private practice generally. Findings per specialty are in Table 4.7.

Table 4.7: Proportional changes in numbers per week of referrals to private specialists in single specialties per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	differ-ence in changes	Nov. 1988 mean	change from Mch'87	differ-ence in changes
rehabilitation specialists	city	1.23	.92	-25%	p≥.05	.70	-43%	**p<.001**
	county	.64	.60	-7%		.66	+3%	
gynaecologists/ obstetrians	city	.85	.69	-19%	p≥.05	.46	-45%	**p<.001**
	county	.98	.90	-8%		1.06	+8%	
pathologists	city	.81	.44	-45%	**p<.01**	.68	-16%	p≥.05
	county	2.75	3.00	+9%		3.26	+19%	
ENT specialists	city	1.00	1.08	+7%	p≥.05	.93	-7%	p≥.05
	county	2.32	2.42	+4%		2.42	+4%	
ophthalmologists	city	1.06	.91	-14%	p≥.05	.88	-17%	p≥.05
	county	3.36	3.22	-4%		3.30	-2%	
dermatologists	city	1.66	1.32	-20%	p≥.05	.98	-41%	p≥.05
	county	1.78	1.55	-13%		1.41	-21%	
psychiatrists	city	.29	.29	-1%	p≥.05	.17	-41%	p≥.05
	county	.41	.42	+0%		.34	-18%	
pediatricians	city	.02	.02	-6%	p≥.05	.02	-11%	p≥.05
	county	.10	.08	-16%		.07	-25%	
internists	city	.14	.21	+52%	p≥.05	.23	+67%	p≥.05
	county	.30	.26	-16%		.34	+10%	
surgeons	city	1.22	1.00	-18%	p≥.05	.81	-33%	**p<.05**
	county	1.37	1.32	-4%		1.32	-3%	
anaesthesiologists	city	.30	.25	-17%	p≥.05	.19	-35%	p≥.05
	county	.34	.33	-1%		.29	-12%	
other/unknown	city	2.69	2.48	-8%	p≥.05	2.22	-17%	**p<.01**
	county	.28	.21	-25%		.31	+10%	

With none of the eleven specialties listed is the finding with referrals to private specialists in general confirmed: for no specialty is there a sustained decrease in referrals because of the change in payment.

As a delayed effect of the change in payment, patients were referred less often to hospital as well. Findings per hospital specialty ward are in Table 4.8. Here four specialties are left aside because no patients were referred there at all.

Table 4.8: Proportional changes in numbers per week of referrals to single hospital specialty wards per 1000 patients

		March 1987 mean	March 1988 mean	change from Mch'87	difference in changes	Nov. 1988 mean	change from Mch'87	difference in changes
gynaecology/ obstetrics	city	.49	.24	-51%	**p<.01**	.22	-54%	**p<.01**
	county	.48	.47	-2%		.49	+2%	
ENT department	city	.04	.07	+62%	p≥.05	.07	+55%	p≥.05
	county	.21	.15	-29%		.15	-30%	
dermatology	city	.06	.03	-54%	p≥.05	.01	-76%	**p<.05**
	county	.04	.03	-12%		.04	+12%	
psychiatry	city	.08	.06	-26%	p≥.05	.05	-42%	p≥.05
	county	.01	.00	-74%		.00	-77%	
pediatrics	city	.04	.09	+133%	**p<.05**	.05	+30%	**p<.05**
	county	.05	.01	-78%		.00	-96%	
internal medicine	city	.60	.64	+6%	p≥.05	.43	-29%	**p<.01**
	county	.10	.12	+17%		.13	+24%	
surgery	city	.40	.33	-19%	p≥.05	.41	+2%	p≥.05
	county	.41	.26	-37%		.26	-37%	
other/unknown	city	.64	.76	+19%	**p<.001**	.55	-14%	**p<.001**
	county	.04	.08	+109%		.10	+157%	

Now the finding of a delayed decrease in referrals to hospital generally is confirmed with two specialties in hospital; one specialty shows a sustained decrease, and one even shows a sustained increase in referrals that should be attributed to the change in payment system. Indeed, like the findings on single diagnostic and curative services, findings on referrals to single specialties in private practice and in hospital are heavily mixed.

4.4 Discussion

In the previous report mentioned above (Krasnik et al., 1990), a test of the effects of introducing fees for services in Copenhagen City, Denmark was performed. We assumed additivity between doctor-bound and time-bound tendencies to perform a service, and compared changes in time-bound tendencies or rates. In the present report, instead of that we assumed equal variances among doctors, and compared average changes in rates.
In the previous report we tested the effects of introducing fees for services by comparing

proportional changes in rates. Here we showed that such a comparison produces more unequivocal and thus more solid results than comparing sizes of changes in rates.
In the previous report we concluded that services over which doctors have more discretion have increased more upon the introduction of fees for services. Here that conclusion is fully underpinned. Generally, it is up to the doctor to decide that a diagnostic or curative service is his to perform, while consultations may equally well be initiated or left undone by the patient. Thus doctors have more discretion over diagnostic and curative services than they have over consultations. We found that diagnostic and curative services showed a sustained increase, and consultations showed no increase upon the introduction of fees.
In the previous report we also concluded that the increased numbers of services performed seem to have replaced referrals at least to some degree. Again, that conclusion is fully underpinned. Referrals to specialists in private practice, which are the more common ones, show an immediate decrease in number. With the rarer referrals to hospitals, the decrease is delayed.
Addressing one question not previously answered, we showed that perinatal and contraceptive services have not decreased in number. Yet one could imagine that they would have come to be neglected to some degree, owing to the increased attention paid to other services. In that respect, health care quality has not deteriorated.
Finally, we specified single diagnostic and curative services performed, and referrals to single specialties. Comparing changes in these, we obtained seemingly random results. Here we seem to have reached a limit to the tenability of the assumption introduced.
We assumed equal variances among city and county doctors, of services performed and referrals made. We argued that a priori, in the shorter period of data collection in the city, more random variation in complaints presented was to be expected. This would result in larger variances in the city, violating our assumption. Such a violation will be strongest with actions most dependent on complaints presented. And indeed among the actions investigated, single diagnostic and curative services and referrals to single specialties are most dependent on complaints. If with these the assumption is violated too much, random results like the ones that we produced can be expected.
We conclude that, because of the short period of data collection, no formal test of the effect of the payment change on single services and referrals can be applied. As for formal testing, the right level of specificity was chosen in the previous report.
Still, inspecting the average changes in services one year after the change in Tables 4.5 and 4.6, a clear pattern can be discerned that may be taken as an indication of an effect of the change in payment.
In seventeen instances out of nineteen, an increase in the city coincides with either a decrease in the county or an increase of which the one in the city is a multiple. In Table 4.5 the 20% increase of cervical smears in the city in November 1988 is no multiple of the 15% increase in the county. In Table 4.6 a 62% decrease in removals of corpora aliena occurs. With these exceptions, the pattern holds.
The pattern holds less well among the changes six months after the payment change, in Tables 4.5 and 4.6. In March 1988 five exceptions occur. The pattern holds even less well

for the changes in referrals in Tables 4.7 and 4.8. So the conclusion seems to be tenable that the effect of the change in payment on services performed does not show in half a year, but does show in one year. Its effect on referrals seems to depend, not surprisingly, on the nature of the specialties concerned. One would expect that the decrease in referrals to those specialist is largest, whose services are most akin to those rendered by GPs. If so, the care provided by gynaecologists/obstetricians is especially akin to the care that Copenhagen City GPs provide, according to Tables 4.7 and 4.8.

If our conclusion on GPs' services under study holds, the change in payment has quite varying effects on single services. Diagnostic services generally had increased by 66% in the city one year after the change. This seems to be no small increase at all. Next, for single diagnostic services the increase varied from 19% up to no less than 633%. For curative services the general increase was 80%: again, no small increase at all. Here single increases varied from 17% up to 221%. These variations are quite large as well. They surely deserve further examination.

5. Introducing fees for services with professional uncertainty*

5.1 Introduction

In Denmark, since 1961 a system of general practitioners' remuneration has been operative in which by prospective estimation the average general practitioner (GP) gets about half his income from capitation for patients on his list, and about half from fees for consultations and a set of mainly diagnostic and curative services. In Copenhagen City this system was not introduced until October 1987. Until then, in Copenhagen City a payment system was in use in which GPs received the major part of their income from capitation. Copenhagen City, with about half a million of inhabitants, is the central part of the larger (sub)urban area of Copenhagen County, with about one million inhabitants in all.
We examined the effect of the introduction of fees for services in October 1987, by comparing the change in number of services performed among Copenhagen City GPs with the change among GPs in the remainder of Copenhagen County. The Copenhagen City GPs served as the experimental group, and GPs in the remainder of Copenhagen County, where the mixed system of GP remuneration already existed, served as the control group. We compared numbers performed of a sample of the diagnostic and curative services newly paid for in Copenhagen City, comparing numbers per week per 1000 registered patients six months before October 1987 with numbers six months and one year after that.
Total numbers of both sampled diagnostic and sampled curative services showed a significantly larger increase among Copenhagen City GPs than among GPs in the remainder of Copenhagen County. They did so when numbers both six months and one year after October 1987 were compared to numbers six months before that. These results were stable in that they were obtained equally with differing statistics, in Krasnik et al. (1990) and Chapter 4 of this book.
In Chapter 4 we also investigated the changes in single diagnostic and curative services. We concluded that the introduction of fees produced far larger increases in some services than in others. Here we ask how these differences can be explained.

In health economics, increases like the ones mentioned are explained by stating that income is an argument in GPs' utility function. Yet income maximization is considered to be constrained by professional medical standards on when to perform a service, as another argument in that function. The strength of that constraint however, is considered to vary with consensus on standards (Evans, 1984: 150-151).

---

* The text of this chapter, combined with parts of Chapter 4 and rewritten to analyze relative proportional changes instead of just proportional changes as we do here, has been accepted for publication in Health Care Financing Review. The differing analysis does not lead to differing conclusions.

In health services research, variation over services in consensus on when to perform them is nowadays attributed to varying strength of the scientific foundation of standards, as perceived by the medical community (e.g.: Wennberg, 1987; Wolff, 1989: 99). The weaker the perceived scientific foundation of a standard on when to perform a service, the greater is the lack of consensus, or professional uncertainty relative to it. And the greater the uncertainty, the larger the variation in performance of a service. Comparing surgical interventions performed in hospitals, differences between hospital markets in numbers per capita that are attributed to varying degrees of professional uncertainty range from no differences up to tenfold differences.
Attribution of variation to professional uncertainty results from standardizing rates for age and gender distributions of populations, and excluding random variation under a Poisson assumption (McPherson et al., 1982: 1313-4). The magnitude of the remaining systematic variation is then taken as a measure of professional uncertainty (McPherson, 1989: 11).
Here, in Section 5.3 another measure of professional uncertainty will be introduced that differs in content as well as in form from the one described. As for content, it takes account of morbidity itself, instead of using age and gender as proxies. As for form, it measures uncertainty in terms of indetermination rather than variation. The concept to be measured is (un)certainty among providers about when to perform a service. Therefore, the degree to which performances of a service are determined by (morbidity) conditions assessed reflects the structure of the concept more explicitly.
Using the new measure, in Section 5.4 we will investigate the health economics prediction that the introduction of a fee for a service stimulates its performance more when professional uncertainty concerning it is greater. We will examine whether or not the varying rates of increase of services mentioned above can be explained in this way.

5.2 Data

Our data were recorded by 87 Copenhagen City GPs during one week in March 1987, by 86 GPs for a week in March 1988, and by 76 for a week in November 1988. Among them, 72 GPs participated all three times. During the weeks concerned they recorded their consultations with regard to diagnoses made and services rendered. The recording form that they used included a list of (complaints and) diagnoses derived from ICHPPC-2 (Classification Committee,1983), a sample of 13 out of the 43 diagnostic services newly paid for in Copenhagen City, and a sample of 8 out of the 27 curative services newly paid for.
Comparing March 1987 to March 1988 data, both for diagnostic and for curative services overall increases were shown to be larger in Copenhagen City than in Copenhagen County on a 5% significance level. Comparing March 1987 to November 1988 data, they were so on a 1% significance level (Table 4.2). Hence in the latter comparison the evidence is strongest that services in general are performed more often because fees for them have been introduced. Therefore, the March 1987 - November 1988 comparison will be used to test the prediction on professional uncertainty in Section 5.4.

For the weeks concerned, the numbers of patients registered with the 72 doctors, of consultations and of the sample of services recorded by them are reported in Table 5.1.

Table 5.1: Numbers of patients and services of 72 GPs in Copenhagen City before and after the change in remuneration

	March 1987	November 1988
registered patients	123,964	127,009
consultations	9,516	10,419
diagnostic services	525	892
curative services	106	199

Table 5.1 shows that, compared to March 1987, diagnostic and curative services are performed nearly twice as often in November 1988. Where services are performed more often, the degree to which the performance of a service is determined by morbidity conditions can be better assessed. Therefore, the November 1988 data will be used to assess these degrees of (in)determination.
Formally, in doing so we weaken our test of the predicted effect of professional uncertainty. A greater professional uncertainty among doctors as measured after the change in payment system may be not a modifier of the effect of the introduction of fees, but another effect of it. It may be that fees tend to increase the numbers of services by increasing the range of states of health in which they are performed. And it may be that services vary in their openness to such a tendency. Then in varying degrees "more servicing, which may initially be a response to economic factors, becomes over time the new standard" (Evans, 1984: 152). Our analysis rests on the assumption that one year is too short a time span for that. We introduce that assumption in order to avoid small numbers.
We further avoid small numbers by including data on the four GPs who did not record their consultations in March 1987, but did do so in November 1988. Uncertainty among GPs about when to perform a service is therefore assessed for 76 GPs, in Section 5.3.
Morbidity conditions as assessed by these GPs were recorded using a list of (complaints and) diagnoses. More than one diagnosis could be recorded. The numbers of consultations with differing numbers of diagnoses are in Table 5.2.

Table 5.2: Numbers of consultations with 0 through 5 diagnoses, as coded from November 1988 records of 76 Copenhagen City GPs

no diagnosis	1 diagnosis	2 diagnoses	3 diagnoses	4 diagnoses	5 diagnoses	sum total
4,001	6,239	614	73	17	3	10,947
37%	57%	6%	1%	0%	0%	100%

The large number of consultations with no diagnosis partly reflects the fact that in general

practice a diagnosis is not always made. It may be still unclear what diagnosis should be made, or the reason for the consultation may be of an administrative nature (e.g., getting the result of a diagnostic test).
Partly, the large number of consultations without any diagnosis is also an artefact of data collection. Our recording form contained a list of 34 (complaints and) diagnoses, with the request either to pick an item on that list or else to write down the diagnosis. The written diagnoses most closely related to one another and largest in number were grouped into eleven extra items afterwards. So many diagnoses were lost both in collecting and in processing the data.
With no diagnosis coded, the degree to which the performance of a service is determined by the diagnosis cannot be assessed; with more than one diagnosis coded complexities of comorbidity are introduced with hardly any profit in numbers of consultations analyzed, as Table 5.2 shows. Therefore, uncertainty among GPs on when to perform a service is assessed with data on 6,239 one-diagnosis consultations, in Section 5.3.
In Table 5.3 the diagnoses are listed that were made in these consultations and, moreover, were made five times or more while a service in our sample was performed.

The diagnoses listed are derived from ICHPPC-2 by selection and combination. Next, we grouped them into umbrella categories, like 'acute infections'. Diagnoses are not listed if they were never made when a service was performed, or if suspicion of measurement error is due.
The combinations of diagnoses and services found have been checked on their medical plausibility. Combinations that were clearly produced by measurement error never arose in more than four consultations. Therefore all combinations that arose in less than five consultations were excluded. Hence only diagnoses coinciding with services in at least five consultations are listed here.
Diagnoses not listed are put together in the residual categories, such as 'other acute infections', and in the umbrella category 'psychiatric and neurological disorders'. No psychiatric or neurological disorder was ever diagnosed in combination with a service in our sample (numbers under suspicion of measurement error excluded).

The 18 diagnoses listed (residual categories excluded) can be used in measuring professional uncertainty on when to perform the 21 services in our sample. However, for diagnoses and services similar restrictions hold. Not all 21 services were performed in combination with one of the diagnoses listed in sufficient numbers to be above suspicion of measurement error.
For this reason, nine services are omitted in the next section. These are: taking a blood sample, proctoscopy, electrocardiogram, ESR measurement, urine microscopy, removing ear wax, removing corpora aliena (with two specifications), and bladder catheterization. A tenth one, performing a pregnancy test, was omitted because its most obvious reason is not included in either ICHPPC-2 or our recording form: a woman's question whether or not she is pregnant.

For the remaining 11 services, professional uncertainty on when to perform them in terms of the 18 diagnoses listed could be measured. The results are in the next section.

Table 5.3: Numbers of November 1988 one-diagnosis consultations in which any service under study was performed at least five times, by diagnosis made

acute infections			1,536	25%
URI (head cold)	521	34%		
acute tonsillitis	139	9%		
urinary infection	209	14%		
conjunctivitis	117	8%		
other acute infections	550	36%		
acute injuries			37	1%
open wound	37			
chronic disorders			1,189	19%
hypertension	334	28%		
diabetes mellitus	98	8%		
anaemia	55	5%		
other chronic disorders	702	59%		
psychiatric and neurological disorders			1,219	20%
gynaecological and obstetric disorders			625	10%
menstrual disorders	164	26%		
fluor vaginalis	193	31%		
menopause symptoms	69	11%		
other gynaecological disorders	69	11%		
obstetric disorders	130	21%		
joint and muscle disorders			975	16%
non-degenerative joint pains	258	26%		
other joint and muscle disorders	717	74%		
dermatologic disorders			658	11%
warts/condylomata acuminata	122	19%		
dermatitis/eczema/rash	366	56%		
abscesses, other localized inflammations	109	17%		
benign tumour	61	9%		
sum total			6,239	100%

5.3 Measuring professional uncertainty

As was argued in Section 5.1, professional uncertainty is measured here in terms of indetermination of performances of a service by morbidity conditions assessed.

With full professional certainty among doctors, there would be full consensus on standards on when to perform a service. Such a standard would include a full description of the conditions in which a service should be performed. Full consensus on such a standard would show when every doctor always performed the service when he assessed a condition described by the standard, and no doctor ever performed the service in other conditions.
In such a state of affairs, the number of times any doctor performed a service would equal the sum of numbers of times he assessed the relevant conditions.
Now assume that the ideal state of affairs is still there, but conditions are not described fully in the data. For example, every doctor always removes ear wax with hearing problems and when the external ear is infected. No doctor ever removes ear wax in other conditions. From the data we know of hearing problems as such, but we know only of infections of the ear generally. Next, assume that the relevant but unknown particulars within the conditions that we do know of are equally distributed in every doctor's practice. To continue the example, assume that in every doctor's practice half of the ear infections are in the external ear.
In that case every doctor would remove ear wax a number of times equal to his assessments of hearing problems plus half his assessments of ear infections.
In a regression over all doctors, the non-standardized coefficient (b) of ear wax removals on hearing problems would be 1, and on ear infections it would be ½. Still, the coefficient of determination (R^2) would be 1, expressing full consensus among doctors on when to remove ear wax.
Thus if the assumptions are valid, an empirical value of R^2 would be an appropriate measure of the degree to which a real state of affairs approximates the ideal one: then R^2 is an appropriate measure of professional certainty among doctors.

Now looking back at Table 5.3, it seems safe to assume that for none of the services are the relevant conditions described fully by the diagnoses listed.
By the same token, there is more hazard in the assumption of equal distributions of relevant particulars. A priori, the equal distributions assumption is more valid when the conditions on which we do have data are more specific themselves. On the face of it, the ones in Table 5.3 are quite unspecific indeed.
Hence when the diagnoses that are listed in the previous section enter the measurement of professional uncertainty by R^2, their global character makes content validity questionable a priori. However, the test of the health economics prediction in the next section will decide whether content validity is good enough for construct validity to show.

In Table 5.4, R^2 is determined for eleven services. These services are listed in boldface vertically. For each service, the diagnoses that were made when it was performed are listed horizontally. The numeric content of Table 5.4 is best described per service. For this we pick the service listed first, taking a cervical smear as an example.
Cervical smears are taken with three diagnoses: menstrual disorders, fluor vaginalis and menopause symptoms. These three diagnoses are made in 164, 193 and 69 consultations respectively; these numbers of consultations from Table 5.3 are repeated in Table 5.4.

Since our total number of one-diagnosis consultations is 6,239, some other diagnosis is made in 5,813 consultations. That number is at the end of the first line on cervical smears.
The next two lines in Table 5.4 describe in different ways to what degree services are rendered when diagnoses are made.
The middle line on cervical smears shows the number of times that the service is rendered when each of the three diagnoses and when some other diagnosis is made. Thus in 164 consultations menstrual disorders are diagnosed, and in these consultations a servical smear is taken 15 times. In the 5,813 consultations in which anything different from menstrual disorders, fluor vaginalis or menopause symptoms is diagnosed, a cervical smear is taken 16 times in all. Yet with no other diagnosis taken separately, a cervical smear is made more than four times, and thus above suspicion of measurement error.
The numbers discussed so far are stepping stones in assessing professional uncertainty among the 76 Copenhagen City doctors about when to take a cervical smear. In the third line that uncertainty is assessed. The first two lines gave information on the level of consultations; the third one gives information on the level of GPs.
In the third line the results of a regression analysis are reported in which GPs are the units of analysis. The dependent variable is the number of times each GP makes a cervical smear. The independent variables are the numbers of times each GP diagnoses menstrual disorders, fluor vaginalis and menopause symptoms. The number of times other diagnoses are made does not enter the regression equation as an independent variable. Yet the 16 times that a cervical smear was made with other diagnoses are included in the values of the dependent variable.
With this reservation, in the third line the empirical counterpart of our ideal state example is shown for cervical smears. One reads that on average GPs take a cervical smear one out of four times that they assess menstrual disorders: the non-standardized regression coefficient is +.24. Yet counted over all GPs, they take a cervical smear fewer than one out of ten times that they assess menstrual disorders: 15 out of 164 times, in the first two lines. The difference disappears when a service is hardly performed with 'other diagnoses', as is shown further down the table with 'urine culture with sensitivity' and 'removing warts'. Hence the difference stems from excluding 'other diagnoses' while including services performed with these in the total numbers of performances entering the regression equation. This enlarges the values of the non-standardized regression coefficients.
As another effect, it also reduces the value of R^2 when a service is often performed with 'other diagnoses'. Performances of a service with other diagnoses than those entering the regression equation cannot be exlained in the equation. But then again, when performances of a service are heavily scattered over 'other diagnoses' this can be taken as a secondary indication of indetermination. That is: if measurement errors are in the recording of diagnoses, and if these errors do not vary in number with the service performed. Under these assumptions R^2 should be lower for heavily scattered services.
We could have avoided both effects by excluding the performances of the service with other diagnoses from the values of the dependent variable. But then the secondary indication of indetermination just discussed would not have been catched by R^2, and the measure

would have lost empirical coverage.
Given the enlarged regression coefficients, one reads in the third line that on average GPs take a cervical smear one out of four times that they assess menstrual disorders. Likewise, they do so one out of six times with fluor vaginalis (+.17), and one out of three times when menopause symptoms are assessed (+.32). Deviations from these averages indicate professional uncertainty as we introduced it in our ideal state example, under the equal distributions assumption discussed before. These deviations and the scattered performances of the service taken together let R^2 sink from 1 to .48.

As measured in Table 5.4, the degree of professional certainty about when to perform the services listed varies from .01 for dressing an immobilizing bandage to .49 for an inoculation for culture.
One would like to judge the content validity of this measure. In principle one could do so by examining whether the reported dependences of services performed on diagnoses made reflect medical knowledge. One could proceed by neglecting non-significant coefficients, taking a significant positive coefficient as evidence of agreement that a diagnosis is a good reason to perform a service, and taking a significant negative coefficient to show agreement that a diagnosis is a bad reason to perform the service.
In the latter instance the finding would be that although a service is sometimes performed with a disorder, doctors who see the disorder more often perform the service less often, other things being equal. This would reflect medical knowledge if experience shows (to the experienced) that the service generally does not help or give a definitive answer with the disorder at hand.
In Chapter 8 we will apply the measure presented here to more precise diagnostic information, and will indeed show content validity with few exceptions, following the rules just described.
However, owing to the global character of the diagnostic categories used in the table above, here one would inevitably run into speculation on the precise nature of complaints and diagnoses. Therefore here it seems more fruitful to examine whether content validity is good enough to produce construct validity, as we will do in the next section.

Table 5.4: Numbers of consultations and performances of a service by diagnosis, with dependences* of performances on diagnoses made, for 11 out of 21 sampled services performed by 76 Copenhagen City GPs in November 1988

cervical smear	menstrual disorders	fluor vaginalis	menopause symptoms	other diagnoses	R^2
consultations	164	193	69	5,813	
performances	15	26	6	16	
dependences	+.24	+.17	+.32		.48

haemoglobin measurement	diabetes mellitus	anaemia	obstetric disorders	other diagnoses	R^2
consultations	98	55	130	5,956	
performances	5	16	5	33	
dependences	+.12	+.67	+.06		.25

blood glucose (photometer)	diabetes mellitus	other diagnoses	R^2
consultations	98	6,141	
performances	14	7	
dependence	+.14		.10

strepto/urine culture	URI (head cold)	acute tonsillitis	urinary infection	other diagnoses	R^2
consultations	521	139	209	5,370	
performances	6	5	18	12	
dependences	+.05	+.09	+.11		.06

inoculation for culture	URI (head cold)	acute tonsillitis	urinary infection	conjunctivitis
consultations	521	139	209	117
performances	11	14	8	7
dependences	-.02	+.39	+.03	-.07

(continued)	menstrual disorders	fluor vaginalis	other gyn. disorders	abscesses/ local inflamm.
consultations	164	193	69	109
performances	5	46	5	5
dependences	+.02	+.47	+.23	+.53

(continued)	other diagnoses	R^2
consultations	4,718	
performances	19	
dependences		.49

(continued on next page)

(Table 5.4 continued)

urine test with sticks	urinary infection	hyper-tension	diabetes mellitus	fluor vaginalis
consultations	209	334	98	193
performances	63	7	19	10
dependences	+.53	+.00	**+.35**	-.08

(continued)	obstetric disorders	other diagnoses	R^2
consultations	130	5,275	
performances	6	29	
dependences	+.07		**.31**

urine culture with sensitivity	urinary infection	other diagnoses	R^2
consultations	209	6,030	
performances	19	2	
dependence	+.11		**.14**

removing warts	warts/ cond.acum.	other diagnoses	R^2
consultations	122	6,117	
performances	43	3	
dependence	+.35		**.33**

in/excision of abscess/tumour	abscesses/ local inflamm.	benign tumour	other diagnoses	R^2
consultations	109	61	6,069	
performances	12	7	7	
dependences	+.05	+.08		**.02**

treating a large wound	open wound	other diagnoses	R^2
consultations	37	6,202	
performances	11	4	
dependence	+.35		**.44**

dressing an immobilizing bandage	non degen-erative joint pains	other diagnoses	R^2
consultations	258	5,981	
performances	8	9	
dependence	+.03		**.01**

* non-standardized regression coefficients b (with coefficients of determination R^2-adjusted printed in boldface), after aggregation into numbers per doctor

5.4 Testing for construct validity

To measure professional uncertainty among doctors concerning services, we analyzed data on one-diagnosis consultations of 76 doctors in November 1988. Now we consider the R^2 values obtained as characteristics of services, and assume that they are not affected yet by the change in payment in October 1987. We compare them to changes in performance of services in all consultations by 72 doctors between March 1987 and November 1988.

R^2 values were obtained for 11 out of 21 services. In Table 5.5, changes in performance are reported for 19 of them. ESR measurement and bladder catheterization are omitted. The 72 GPs did not measure ESR at all in their week of March 1987, and did not catheterize any bladder in their week of November 1988. Hence for ESR measurement, no proportional increase can be assessed, and attributing a 100% decrease (from .02 to 0.00 per 1000) to bladder catheterization seems to be overstretching the data. Proportional changes are argued to be a better measure than sizes of change in Chapter 4.

Table 5.5: Changes in numbers of services per week per 1000 registered patients performed by 72 Copenhagen City GPs for 19 out of 21 sampled services

	March 1987	November 1988	prop. change
diagnostic services			
taking a blood sample	.32	.38	+ 22%
cervical smear	.96	1.15	+ 20%
pregnancy test	.26	.31	+ 19%
proctoscopy	.02	.07	+332%
electrocardiogram	.01	.02	+108%
haemoglobin measurement	.54	.83	+ 54%
blood glucose (photometer)	.04	.29	+633%
streptoculture/urine culture	.14	.44	+223%
inoculation for culture	.72	1.38	+ 91%
urine test with sticks	1.15	1.85	+ 61%
urine microscopy	.06	.10	+ 57%
urine culture with sensitivity	.05	.20	+275%
curative services			
removing warts	.17	.44	+162%
removing ear wax	.40	.47	+ 17%
removing corpora aliena from eye, ear, nose or throat	.02	.01	- 62%
removing corpora aliena from skin or from under nail	.05	.07	+ 41%
incision/excision of abscess/tumour	.08	.26	+221%
treating a large wound	.07	.13	+ 75%
dressing of immobilizing bandage	.08	.24	+194%

Among these nineteen services, eighteen increase in number. Yet the increases range from a mere 17% for removing ear wax up to 633% for blood glucose measurement. The question to be answered here is how these varying rates of increase can be explained.
More specifically, the question is whether they can be explained from varying degrees of professional uncertainty among GPs about when to perform them. The prediction to be tested is: services about which GPs are less certain professionally show a higher rate of increase when fees for them are introduced.
For eleven services, this prediction is tested by rank order correlation in Table 5.6.

Table 5.6: Professional certainty relative to 11 sampled services among 76 Copenhagen City GPs and proportional change in the performance of these services by 72 of these GPs

	professional certainty R^2	proportional change
inoculation for culture	.49	+ 91%
cervical smear	.48	+ 20%
first treatment of large wound	.44	+ 75%
removing warts	.33	+162%
urine test with sticks	.31	+ 61%
haemoglobin measurement	.25	+ 54%
urine culture with sensitivity	.14	+275%
blood glucose (photometer)	.10	+633%
streptoculture/urine culture	.06	+223%
incision/excision of abscess/tumour	.02	+221%
dressing of immobilizing bandage	.01	+194%
	rank order correlation = -.64; p = .035	

On a 5% significance level, services about which doctors are less certain professionally, as measured by a lower R^2, show a larger proportional increase.

5.5 Discussion

In Section 5.3, we noted that content validity of the measure of professional uncertainty introduced was questionable a priori because of the global character of the diagnostic categories used. In the previous section, we examined whether content validity was good enough to produce construct validity. Indeed, it did: the health economics prediction introduced in Section 5.1 was confirmed in the previous section.
The globality of diagnostic categories and the frequency of measurement errors are the main limitations of the analysis reported in this chapter. Globality made it impossible to judge content validity. Due to measurement errors we had to introduce extra assumptions in order to avoid loss of empirical coverage of our measure.

Using a more precise diagnostic classification, the content validity of our measure will be shown in Chapter 8. In future research using such a classification, the health economics prediction investigated here should be tested in natural experiment conditions like the one in Copenhagen, Denmark in 1987.

Confirmation of the prediction would then support the assertion that GPs react to the introduction of fee-for-service as income maximizers under the constraint of professional standards on when to perform a service, while the strength of that constraint varies with professional certainty about standards. For health politics such results would imply that, if the introduction of fees for services is considered, services should be chosen for which constraints on income maximization are strong, and constraints should be strengthened further.

Thus services should be chosen with a strong scientific foundation of standards on when to perform them, as perceived by the medical community. That constraint could be strengthened further by developing and publishing standards, and by stimulating adherence to them through postgraduate education and peer review.

6. General practice under capitation: a comparison of Dutch and Copenhagen urban practice

6.1 Introduction

6.1.1 Hypothesis and context

Differences in performance of health services across physicians and geographic areas have been amply documented. Best documented are differences in rates of surgical procedures performed in hospitals. For example, Bunker (1970) reported that over all surgical procedures, per capita twice as many were performed in the United States as in Great Britain. He commented that differences in disease prevalence in populations "cannot reasonably be invoked to explain" such a difference (op.cit.: 141).
Recognizing that disease prevalences generally vary with age and gender, Vayda (1973) compared gender-specific age-standardized surgical rates in Canada and in England and Wales, and still found rates to be many times as large in Canada for single surgical procedures. Summarizing these and more recent but similar findings, McPherson (1989) states that among western industrialized countries, "demography should not be a major determinant of large variations in rates" of health services (op.cit.: 12).
Since demography is introduced as an indicator of morbidity, McPherson's conclusion implies that differences in performance of health services between western industrialized countries are only to a small extent explained by differences in morbidity. This in its turn implies that to a greater extent they are explained by differing tendencies among physicians to perform health services.

We now turn this implication into a stronger hypothesis: comparing western industrialized countries, **differences in performance of health services are fully explained by differing tendencies to perform them**. If so, differing performances can be taken to stand for differing tendencies to perform. A priori, only with the same disorder presented can differing performances of a service be taken to stand for differing tendencies to perform it. Thus the hypothesis is corroborated when **overall numbers of performances differ only if numbers differ when presentations of the same disorder are compared**. This prediction will be tested here.
If the prediction is corroborated, a minimal condition is fulfilled to generalize effects of a change in GPs' payment system on overall performances of services from Copenhagen to Holland while disregardig morbidity, as we will do in Chapter 7. The condition is minimal in that one still has to make two additional assumptions for such a generalization. For one, one has to assume that conclusions holding for equalities and differences in performance (in terms of statistical significance) hold for degrees of performance as well. And for another, one has to assume that changes in overall performance do depend on morbidity

just as little as differences do.
Still, not only morbidity and payment are relevant to the performance of services. If the prediction is corroborated, so is the hypothesis that overall performances are explained by tendencies to perform. This raises the question as to what explains these tendencies. McPherson (op.cit.: 16) offers a heuristic for answering that question: "if the variation observed is out of proportion to feasible morbidity differences, then the influence of culture, education, or availability on clinical decisions would be the next most obvious explanation". If the prediction is corroborated, this heuristic may be useful for explaining any overall differences that we find in testing it. If the prediction is falsified, the heuristic may at least be useful for explaining differences in services performed with the same disorders presented.

The conclusions that we build on here are mainly drawn from comparisons of surgical procedures performed in hospitals. Generally, age and gender are used as proxies for morbidity in these comparisons. On the one hand, we extend their range of application by comparing services performed in general practice. Numbers of diagnostic and curative services performed by general practitioners (GPs) in Copenhagen, Denmark and in urban Holland will be compared. Thus small area variation research is extended to primary care. On the other hand, in testing the prediction for these services in general practice we directly control for morbidity presented.
To be able to apply the heuristic to our results, we have to assess whether differences are to be expected from the influences listed, when GPs in Copenhagen and urban Holland are compared. That assessment will take an argument of considerable proportions, which fills subsections 6.1.2. and 6.1.3. There we leave aside the influence of overall characteristics of culture. Such characteristics (like Cartesianism in France and Romanticism in Germany) can be argued to influence medical practice (Payer, 1988). However, there is no body of knowledge from which one can deduce whether differences are to be expected from that influence in other instances.
Thus we will explore the heuristic that equalities and differences in the performance of services should be explained by equalities and differences in the education of providers and in availability of their services. A test of predictions resulting from this heuristic is only by implication: if morbidity cannot explain the overall performance of services, the explanation should be found in the other factors mentioned.
Morbidity cannot explain overall performance if the prediction at issue to begin with is corroborated. For testing it, morbidity and services have to be made comparable for different health care systems. This will take an argument of considerable size as well, in Section 6.2. Differences in sample design, recording instructions to GPs, coding systems of complaints and diagnoses, and in services recorded all have to be dealt with.
These difficulties are common to all attempts to construct meaningful Ambulatory Visit Groups (Schneider et al., 1988). AVGs are measures of case mix in ambulatory and primary care, comparable to Diagnosis Related Groups in in-patient care. AVGs are an important tool in comparative research in primary care. The development of this methodology, however, is still in an infant stage (McACE, 1990). This chapter illustrates the difficulties inherent

in making complaints, diagnoses and services comparable over health systems.
In summary, the relevance of our comparison is threefold. Firstly, it provides an extension of the application of small area variations research to primary care services. Secondly, it directly controls for morbidity differences and thus provides a better test of the conclusions that we build on. Finally, it illustrates the difficulties inherent in making morbidity and services comparable over health systems, and thus contributes to the development of the increasingly important concept of Ambulatory Visit Groups.

6.1.2 Education and availability

To begin with, we examine whether differences between Copenhagen and urban Holland are to be expected from the influences of education and availability. We will deal with these two influences one by one. As for availability, three determinants of availability of GPs' services will be pointed to. It will be argued that their influence is constrained by the availability of alternatives to GPs' services.

Professional education of GPs differs markedly between the two countries. In Denmark since 1976 after graduation one has had to attend a year and a half of follow-up education to qualify as a physician. Internships in hospital wards of prescribed specialties had to be attended for one year, and internships in other hospital wards or general practice for six months. To qualify as a GP, one had to attend two years of internships to conclude, including at least one year in general practice. In 1983, the concluding two-year period of internships was extended to a five-year period.
In Holland, on the other hand, since 1973 compulsory post-graduate education has consisted of only one year of training in general practice. From 1988 training has been extended to two years, partly to be attended in hospital as well. Thus at least comparing younger Danish and Dutch GPs, the Danish have attended longer hospital internships.
From this difference in education, one would expect Danish GPs to have more technical medical skills, and therefore to perform the diagnostic and curative services to be investigated here more readily.

As for availability, in principle three determinants can be pointed to. For a GP's service to be available to anyone, a GP should be present, he should in some way get paid for his services (if he is not prepared to render them for free), and one should be able to pay for them. Generally, ability to pay varies with health insurance. Hence three determinants of availability are GP-to-population ratio, GPs' payment system, and health insurance. Two health insurance schemes coexist both in Denmark and Holland, with GPs' payment systems of their own. Therefore, GP-to-population ratios within insurance schemes will also be discussed here, as will population ratios between insurance schemes.
As for health insurance, in Denmark it is a public, tax-based affair generally. In Holland public health insurance (called 'sick fund' insurance) is obligatory to those below, and closed to those above, an income level set by law, as was 'group I' insurance in Denmark before 1973. In Denmark one can now choose between two public insurance schemes, of

which group I insurance is very similar to Dutch sick fund insurance. In both schemes one has to register with a GP, who manages access to specialist and hospital care by referral. Here GPs' services to Danish group I and to Dutch sick fund insured patients will be compared.

As noted, ability to pay varies with health insurance. In any insurance scheme, health care is either supplied in kind, with payment by the insurer to the provider, or health care costs are reimbursed by the insurer to the insured. Also, any insurance scheme may imply some degree of copayment by the insured, be it through out-of-pocket payment complementing supply in kind, or through reimbursement that is only partial. With partial reimbursement, ability to pay will vary most with net income. With supply in kind without copayments, it does not vary with net income at all.

Both in Danish group I and in Dutch sick fund insurance, health care is supplied in kind. Small copayments for pharmaceuticals prescribed were a common characteristic at the time of investigation. For the services to be investigated here, no copayments were required in either insurance scheme. Thus ability to pay for these services, as one determinant of availability, did not differ between populations covered by both insurance schemes.

The most obvious determinant of availability is the presence of GPs. Here it is specified as GP-to-population ratios within Danish group I and Dutch sick fund insurance schemes, with population ratios between insurance schemes in both Denmark and Holland as a side condition.

In Denmark at the time of investigation, 2,956 GPs served 3,909,100 group I insured persons (aged 16 or over: younger children, coinsured with their parents, are not counted in Danish statistics), so that there were 76 GPs per 100,000 population insured. Likewise, in Holland 5,826 GPs served 7,347,384 sick fund insured persons (excluding those under 16), making up 79 GPs per 100,000. Thus GP-to-population ratios within group I/sick fund insurance hardly differed between Denmark and Holland. To that extent presence of GPs as a determinant of availability of GPs' services hardly differed between the two.

However, population ratios between insurance schemes are a relevant side condition here. In both Denmark and Holland the group I/sick fund insured are the largest part of the population. Yet in Denmark they were 95%, and in Holland only 61% of the population. Hence while there were virtually as many GPs per population within group I/sick fund insurance in Denmark and Holland, Dutch GPs had more patients otherwise insured, who took up their time as well to some degree at least. Consequently, presence of GPs as a determinant of availability of their services was less for Dutch sick fund insured than for Danish group I insured persons.

The third and last determinant of availability is the way GPs get paid for their services, specified here as GPs' payment systems within insurance schemes. For group I/sick fund insured patients, both Copenhagen and Dutch GPs were mainly paid by capitation at the time of investigation. In Denmark the other patients have chosen the 'group II' public insurance scheme; in Holland they are those above the 'sick fund income level', and are privately insured. For both groups GPs are paid by fee-for-service.

Now with fee-for-service, GPs who devote extra time to a patient by performing a service receive extra income in return. With capitation, however, GPs only lose time. So, leaving medical ethics aside, GPs can be expected to be more inclined to make time for a service to a group II/privately insured patient, where they get a fee for the service, than to do so for a group I/sick fund insured patient, where they do not. Since privately insured patients in Holland are larger in number than group II insured patients in Denmark, this will tend to distract more time from Dutch sick fund insured patients than from Danish group I insured. Therefore, availability of GPs' services to Dutch sick fund insured patients will be less.

These influences on the numbers of GPs' services performed will be constrained by the availability to GPs of alternatives to their services. If alternatives are hardly available, medical ethics will restrain GPs from denying a patient a service. What the alternatives to GPs' services are depends on the structure of health care.
As noted above, both Danish group I and Dutch sick fund insured persons have to register with a GP, who manages their access to specialist and hospital care by referral. Thus from the structure of health care, referring a patient is an alternative to providing a service equally open to GPs in Denmark and Holland.
Availability of that alternative depends on specialist and hospital medical personnel-to-population ratios, and on hospital bed-to-population ratios. In the numbers of medical personnel, one might include trainees as well. And indeed one should do so, since the longer compulsory postgraduate education in Denmark that was described above produces larger numbers of medical trainees. According to the description above, for Denmark in 1987 hospital-based physicians including GP trainees and specialists in private practice should be included. For Holland in 1987, only specialists and specialist trainees should be included, since Dutch GP trainees were trained only in general practice then. Thus counted, numbers of hospital and specialist medical personnel were 171 in Denmark and 94 in Holland per 100,000 population. The numbers of hospital beds per 1000 population were 6.9 in Denmark and 4.7 in Holland, with hardly different bed occupancy rates of 81% and 77%.
Thus in terms of both personnel and bed capacity, referring a patient as an alternative to performing a service was an option less available to Dutch than to Danish GPs. Owing to medical ethics, this puts a stronger constraint on distracting their time from sick fund/group I insured patients to Dutch than to Danish GPs.

6.1.3 Explaining equalities and differences

Here we will examine whether equal or different overall numbers of services are explained by equal or different tendencies to perform them, when comparing Copenhagen and urban Dutch GPs. If that hypothesis is corroborated, an explanatory problem shows up: what does explain equal or different tendencies to perform a service? Here we solve that problem by assuming that the inventory of relevant equalities and differences produced above is exhaustive. If it is, our results answer the question which of the influences listed are decisive.

If we find no differences in services performed (with the same disorders presented, or generally as well), equal health insurance of group I/sick fund insured persons and equal GP-to-population ratios within these insurance schemes are decisive among the influences listed.
If services are performed more often in Copenhagen, decisive influences are exerted by the more elaborate Danish professional education, the smaller part of the population otherwise insured, and the correspondingly smaller distraction of GPs' time to these otherwise insured owing to fees paid for services rendered to them.
If services are performed more often in urban Holland, the fewer opportunities to refer a patient as an alternative to performing a service are decisive.

6.2 Data and method

6.2.1 Samples compared

In Copenhagen City 283 GPs were practicing in the spring of 1987. They were all asked to record their patient contacts for one week each in March 1987. Finally, 98 GPs did so. They recorded 13,193 consultations with group I insured patients. Willingness to volunteer was the main selection criterion among GPs. They all practiced in a municipality of 470,000 inhabitants, which together with bordering municipalities constitutes the (sub)urban area of Copenhagen County, with 1,030,000 inhabitants in all.
For urban Holland, a sample of contact record data was drawn from the Dutch National Study on morbidity and interventions in general practice (Bensing, Foets, Van der Velden & Van der Zee, 1991). For this larger study, out of the population of 5,826 GPs practicing in Holland a stratified random sample of 161 GPs was drawn, who recorded their patient contacts for three months each in four subsequent groups. Overall, patient contacts were recorded from April 1987 through March 1988. The sample was stratified by urbanization, region and distance from hospital. GPs were paid to participate in the study. Numbers of GPs by urbanization are reported in Table 6.1.

Table 6.1: GPs participating in Dutch National Study by urbanization

practicing in municipalities with:	number of GPs
up to 30,000 inhabitants	60
from 30,000 to 50,000 inhabitants	62
from 50,000 to 600,000 inhabitants	29
over 600,000 inhabitants: Amsterdam, Rotterdam	10

In Holland only Amsterdam and Rotterdam with their respective bordering municipalities constitute urban areas that approximate the Copenhagen County size. For comparison with Copenhagen City GPs, we chose the 39 GPs in municipalities with over 50,000 inhabitants. For comparison with the one-week contact records by Copenhagen City GPs, we took one week's contact records by urban Dutch GPs. The weeks were drawn randomly

from each doctor's three-month recording period. Two GPs did not record any patient contacts in their sampled weeks, because of holidays or post-graduate courses. The other 37 urban Dutch GPs recorded 3,670 consultations with sick fund insured patients in the weeks concerned.
The samples compared can be expected to differ in three respects. The Copenhagen GPs practiced in a more urban area, all recorded their contacts in the month of March, and all were willing to co-operate for free.
In Holland, more urban areas have been shown to have higher chronic and psychosocial morbidity; that is presented to a smaller extent in general practice, however (Van der Velden, 1990: 606). If such differences apply to our samples as well, they are eliminated in controlling for morbidity presented, as we will do in testing the prediction at issue.
The month of March will have its own seasonal morbidity. Again, in controlling for morbidity this difference is eliminated, and in testing the prediction it is checked whether the elimination makes any difference.
If the Copenhagen GPs are more willing to co-operate for free, they place less value on monetary rewards. If so, the more the argument holds that their time is less distracted from patients under capitation to those under fee-for-service. In the previous section this was argued to be one difference that would explain larger numbers of services rendered to patients under capitation in Copenhagen City.

6.2.2 Services and morbidity

Copenhagen City GPs recorded a total of 45 diagnoses that they made or complaints presented to them: if no diagnosis was made the complaint presented was recorded. 34 items were in a pre-established list on the Copenhagen recording form. GPs could either pick an item from that list or write down complaints and diagnoses. The ones they wrote down most often were grouped into 11 items afterwards. All 45 complaints and diagnoses were derived from ICHPPC-2 (Classification Committee, 1983) by selection and combination. GPs recorded a selection of 13 diagnostic and 8 curative services. Both the selections of services and of listed complaints and diagnoses were based on high expected frequency.
The GPs in the Dutch National Study wrote down their diagnosis or the complaint presented. These were coded into ICPC (Lamberts & Wood, 1987) with data entry by medically trained personnel. GPs recorded a list of 26 curative and 24 diagnostic services.

Of the 21 services recorded by Copenhagen GPs, one (proctoscopy) cannot be defined in Dutch National Study data. Three other services (strepto culture or urine culture, urine test with sticks, and urine culture with sensitivity), can be defined in Dutch data only when taken together, leaving streptoculture out (the result is called 'urine tests' here). Thus as regards definitions, performances of 18 services by Copenhagen and urban Dutch GPs can be compared. Of these, five are left aside because of small numbers (pregnancy test, electrocardiogram, removal of corpora aliena from skin or from under nail, and from eye, ear, nose or throat, and bladder catheterization). In both the Copenhagen and the urban Dutch sample, they were either not performed at all or in no more than 0.1% of consultations in the

weeks concerned. Therefore, performances of 13 services will be compared in the next section.

In comparing the performance of these services, we will control for morbidity presented as measured by diagnoses or complaints recorded by GPs. Full recording by GPs of diagnoses and complaints and full coding of these are time-consuming indeed. Therefore, a pre-established list was used in Copenhagen data collection. As a consequence, however, for many Copenhagen consultations we do not know what diagnosis was made or what complaint presented. In these instances morbidity cannot be controlled for: when no pre-established item was marked and nothing was written down on the recording form, one of the eleven items construed afterwards may still habe been applicable. On the other hand, when more than one diagnosis or complaint has been recorded, one can expect comparisons to become invalid owing to differing comorbidities. Therefore only consultations in which one and only one diagnosis (or complaint) came up will enter the analysis. For Copenhagen GPs, these are 7,088 (or 54%) of the 13,193 consultations with group I insured patients they recorded.
For urban Dutch GPs, the diagnoses they made (or the complaints presented to them) were coded into ICPC at the moment of data entry. For our analysis, these codes were converted into ICHPPC-2 by the conversion table (op.cit.: ch.12). Now for Dutch GPs, only consultations with one diagnosis that could be converted into a diagnosis recorded by Copenhagen GPs will enter the analysis. These are 1,348 (or 37%) of the 3,670 consultations with sick fund insured patients that the urban Dutch GPs recorded. Here the reduction in numbers (to 37%) is stronger than that (to 54%) among Copenhagen GPs. To control for morbidity to the same extent as we are able to do with the Copenhagen data, we limit the analysis to consultations in which diagnoses were made that we also know of if they were made by Copenhagen GPs.

6.2.3 Method

We will examine whether overall numbers of services performed can be taken to stand for tendencies to perform them. We will do so by examining whether overall performances differ if and only if performances differ when comparing presentations of the same disorder, as measured by complaints and diagnoses recorded.
For numbers of services performed, average numbers per consultation will be used, limiting the analysis to the consultations specified above. For readability, average numbers per 100 consultations will be reported. Numbers will be considered different when comparison of consultations yields t-test p-values under .05.
For each service, both overall performances and performances with the same disorder presented should be compared. Thus for each service a disorder should be identified for comparison of performances. Now that comparison can yield results different from comparing overall performances only as far as the service is specific to the disorder. For minimal specificity we apply the criterion that the service is performed with the disorder presented at least half of the time it is performed, in either Copenhagen, urban Holland or both.

For each service we therefore looked for a set of medically related complaints and diagnoses that accounts for at least half of the performances of the service in one or both of the instances.
By that criterion, seven of the thirteen services whose overall performances will be compared proved to be sufficiently specific to any disorder. Therefore, the prediction will be tested for seven services.

6.3 Results

In Table 6.2 the prediction is tested that overall numbers of performances of a service differ if and only if numbers differ when presentations of the same disorder are compared.

Table 6.2: Performances of 7 services per 100 consultations, idem when relevant disorders are presented, and frequencies of these disorders per 100 consultations, in Copenhagen and urban Holland

	Copenhagen	Holland	p
cervical smear	.9	.2	**.000**
- with gynaecol./obstetric disorders	7.9	2.9	**.033**
gynaecological or obstetric disorders	9.0	5.1	
inoculating for culture	1.0	.5	.061
- with gynaecol./obstetric disorders	5.7	4.4	.653
gynaecological or obstetric disorders	9.0	5.1	
urine microscopy	.1	2.9	**.000**
- with urinary infection	2.4	56.3	**.000**
urinary infection	3.6	3.6	
urine tests	1.5	3.4	**.000**
- with urinary infection	20.1	58.3	**.000**
urinary infection	3.6	3.6	
removing warts	.2	.4	.361
- with warts/condylomata acuminata	10.4	27.8	.137
warts/condylomata acuminata	1.8	1.3	
incision/excision of abscess or tumour	.2	.5	.142
- with abscess/inflammation/benign tumour	6.9	17.7	.129
abscess/inflammation/benign tumour	2.4	2.5	
treating a large wound	.1	.8	**.005**
- with a bite or cut	25.9	57.9	**.029**
bites and cuts	0.4	1.4	

In Table 6.2 the prediction at issue is corroborated: among these seven services, overall performances differ in number if and only if performances with the same disorder presented differ, between Copenhagen and urban Holland.
A priori, performances of a service with the same disorder presented can be taken to stand for GPs' tendencies to perform it.
In Table 6.2, performances with the same disorder presented differ with four services out of seven. For three services out of these four, numbers are higher in urban Holland.
Since the prediction is corroborated, overall performances of a service may also be taken to stand for GPs' tendencies to perform it. Overall performances of thirteen services are in Table 6.3.

Table 6.3: Performances of 13 services per 100 consultations, in Copenhagen and urban Holland

	Copenhagen	Holland	p
taking a blood sample	.2	3.3	**.000**
cervical smear	.9	.2	**.000**
haemoglobin measurement	.6	.4	.339
ESR measurement	.0	.4	**.040**
blood glucose measurement	.1	.4	.075
inoculating for culture	1.0	.5	.061
urine microscopy	.1	2.9	**.000**
urine tests	1.5	3.4	**.000**
removing warts	.2	.4	.361
removing ear wax	.3	0.0	---
incision/excision of abscess or tumour	.2	.5	.142
treating a large wound	.1	.8	**.005**
dressing an immobilizing bandage	.0	.4	.050
all 13 services: sum totals	5.1	13.4	**.000**

In Table 6.3, overall performances differ with six services out of thirteen. For five services out of these six, numbers are higher in urban Holland.
Thus both with performances with the same disorders presented in Table 6.2 and with overall performances in Table 6.3, not always do numbers differ, but if they do, more often than not numbers are higher in urban Holland. When performances are summed over services in Table 6.3, numbers are higher in urban Holland again.
To relate these findings to the referring alternative discussed in Section 6.1, numbers of referrals in the consultations analyzed are compared in Table 6.4.

Table 6.4: Numbers of referrals per 100 consultations, in Copenhagen and urban Holland

	Copenhagen	Holland	p
all specialties: sum totals	13.7	7.1	.000

While services are performed more often in urban Holland, patients are referred less often there.

6.4 Discussion

In a recent summary of findings in comparative health services research (McPherson, 1989), it is implied that differences in performance of health services between western industrialized countries are for the greater part explained not by differences in morbidity, but by physicians' differing tendencies to perform these services. Direct tests of this hypothesis are not found, since morbidity differences have usually been estimated by differences in age and gender composition. Here a stronger version of the hypothesis is directly tested by controlling for morbidity presented: that differences in performance are fully explained by differing tendencies to perform. If so, differing overall performances of a service can be taken to stand for differing tendencies to perform it.
A priori, only with the same disorder presented can differing performances of a service be taken to stand for differing tendencies to perform. Thus here the prediction is tested that overall performances of a service differ if and only if performances with the same disorder presented differ. To test that prediction, differences in data collection and classifications used are dealt with that typically complicate attempts to construct Ambulatory Visit Groups.

The prediction is tested and corroborated for seven services performed by general practitioners in Copenhagen, Denmark and in urban Holland. This strengthens the validity of simply taking different overall performances of a service to stand for differing tendencies to perform it.
This result poses an explanatory problem: what does explain these differing tendencies? If GPs in western industrialized countries are utility maximizers pursuing equal goals, these differences should be explained by differing resources available and differing restrictions imposing themselves. Here we elaborated a heuristic proposed by McPherson (1989) that identified education and availability as resources and restrictions. We now apply that heuristic to our results.
According to our results, not for all of thirteen services do overall numbers differ between Copenhagen and urban Holland. But if they do, more often than not numbers are higher in urban Holland, as are numbers summed over services. Generalizing from that, we claim that GPs' services tend to be performed more often in urban Holland. From a difference in samples compared, we expected to find higher numbers in Copenhagen. This only strengthens our claim.

Elaborating the heuristic, we assumed that the resulting inventory of relevant equalities and differences between Denmark and Holland is exhaustive. If it is, and if our generalizing claim is valid, the fewer referring opportunities in Holland are the decisive one of the relevant differences from Denmark. Despite educational and economic influences to the contrary, they bring about that urban Dutch GPs tend to refer patients less often, as we also found, and therefore tend to render a service themselves more often. The alternative would be to deny medical care to a patient, which would run contrary to medical ethics.
Thus lacking referring opportunities as a restriction to the referring alternative to performing a service are more decisive than an elaborate education as a resource in performing it. That one of both alternatives should be realized follows if medical ethics constrain the pursuit of money.

7. From Denmark to Holland: predicting the effects of introducing fees for services

7.1 Introduction

In Denmark since 1961 the majority of the population (about 95%) has been covered by a public health insurance scheme (called 'group I insurance') in which a patient can be admitted to a hospital, or visit a specialist in hospital or private practice, only if he is referred by the general practitioner with whom he is registered. The 5% ('group II insured') minority are not registered with a GP.

Danish general practitioners (GPs) are paid both capitation fees for patients registered with them, and fees for services: both for consultations, and for a set of diagnostic and curative services. Fees are set in such a way that by prospective estimation an average GP gets about half his income from capitation fees, and about half from fees for consultations and services.

In Copenhagen City, this payment system was not introduced until October 1987. Before then, Copenhagen City GPs were mainly paid by capitation fees for patients on their list.

In Krasnik et al. (1990) and in Chapter 4 of this book the changes in rates of consultations, services and referrals among Copenhagen City GPs around October 1987 were reported, and tested against those among GPs in the surrounding (sub)urban area of Copenhagen County. Using differing statistics, both reports concluded that the change in GPs' payment caused no clean-cut increase in consultations, did cause an increase in services, and caused a decrease in referrals that was smaller than the increase in services performed.

In Chapter 4 we found a 7% increase in consultations among Copenhagen City GPs (including consultations by telephone and home visits), which did not differ significantly (on a .05 level) from the 6% increase among Copenhagen County GPs. We concluded that consultations were too much patient-initiated for the introduction of fees to have any influence on their numbers. On the other hand, we found a 66% increase in diagnostic services, an 80% increase in curative services, and a 26% decrease in referrals in Copenhagen City; these changes were significantly larger (on a .01 level) than those in Copenhagen County.

In Chapter 5 we found that the effect of the introduction of fees on services performed varied with a characteristic of services. That characteristic is the professional (un)certainty of GPs relative to the performance of a service. Professional certainty was measured by the degree to which the performance of a service is determined by morbidity conditions. We found that the proportional increase in performance of a service coinciding with the introduction of fees is smaller if its performance is more diagnosis-determined.

This finding can be related to the fact that the decrease in referrals is smaller than the increase in services performed. If the performance of a service is fully diagnosis-determined among GPs, they fully agree on the appropriateness of the service, given a com-

plaint presented by a patient. If they do, it seems safe to assume that if the patient had been referred to a specialist, he would agree on the appropriateness of the service too (and probably would insist on more services in his specialty). Then the performance of the service by a GP always substitutes for its performance by a specialist after referral. Therefore, its performance by a GP always serves to either avoid or postpone a referral, depending on the development of the complaint.
Now if the services for which fees are introduced are not all fully diagnosis-determined, this argument does not hold completely. It follows that the decrease in referrals will be smaller than the increase in services performed, as we found.
If the explanation proposed here is correct, the referral-reducing effect of the introduction of fees for services depends on the diagnosis-determinedness of these services among the GPs (and specialists) concerned. In this sense the effect depends on local medical culture.

In the Netherlands, a smaller majority of the population than in Denmark (of about 65%) are covered by a public health insurance scheme (called 'sick fund insurance') similar to the Danish group I insurance scheme. Until now, Dutch GPs have been paid by capitation for sick fund patients on their list, as were Copenhagen City GPs for group I insured patients. Recently, however, the Dutch general practitioners' association and the association of Dutch sick funds agreed in principle on the introduction of a GPs' payment system like the Danish one, with fees paid for a smaller set of diagnostic and curative services rendered to these patients as a first step.
The question to be answered here is what the relative size of the referral reducing effect will be when such a change in payment system is introduced in the Netherlands. In predicting changes we will take into account both diagnosis-determinedness of services as part of local medical culture, and characteristics of GPs and their practices.

We will answer the question at hand by combining Copenhagen and Dutch data. In Section 7.2 on data and method, the differences in data collection between Copenhagen and the Netherlands are discussed, as are the assumptions introduced to overcome these differences. Also, two ways of aggregating consultation-level data are introduced. To take into account diagnosis-determinedness of services, data on each single service are averaged over GPs. To take into account characteristics of GPs, data are summed over services for each GP.
In Section 7.3 on results, first we examine what was to be expected in the Netherlands if fees were introduced for the services reimbursed in Denmark. We do so taking into account characteristics of GPs in the first subsection, and characteristics of services in the second. In the third subsection, we examine what is to be expected when fees are introduced for the services agreed upon in the Netherlands. Both approaches developed in the first two subsections are applied again.
Finally, in Section 7.4 the referral-reducing effects to be expected from introducing fees for these two sets of services in the Netherlands are discussed.

7.2 Data and method

7.2.1 Data

To start with, we will assess whether or not the changes in consultations, services and referrals among Copenhagen City GPs depend on characteristics of GPs and their practices, and on characteristics of services as part of local medical culture. We will do so by re-analyzing the data described in Chapters 4 and 5.

These data were collected by 72 Copenhagen City GPs who voluntarily recorded diagnoses made or else complaints presented, a sample of diagnostic and curative services performed, and referrals made in each of their consultations with group I insured patients during one week in March 1987 (n=9,516) and during one week in November 1988 (n= 10,419). Hence data stem from six months before and one year after the change in payment. The 72 GPs concerned are a 31% sample of the 233 GPs that practiced in Copenhagen City during that period. Their selection was guided mainly by their willingness to volunteer (Krasnik et al., 1990: 1699).

Complaints presented and diagnoses made were either marked on a list that was derived from ICHPPC-2 (Classification Committee, 1983) by selection and combination, or they were written down in the recording form and coded into ICHPPC-2 afterwards. Using these diagnostic categories, the diagnosis-determinedness of services among Copenhagen City GPs was assessed in Chapter 5.

With the change of payment system in Copenhagen City, fees were introduced for a set of 43 diagnostic and 27 curative services. Of these services, a sample of 13 diagnostic and 8 curative services was included in the recording form. The services sampled were performed frequently in Copenhagen County, and it was judged that the average Copenhagen City GP would be equipped to perform them. Here the sample is assumed to be representative of the larger set of services and it is assumed that the changes in performance of these services can be generalized to the larger set. The services involved are discussed one by one in Chapter 8, which serves as an appendix to the present chapter.

We will use the results of our analysis of the change of payment system in Copenhagen City to predict the effects of a similar change in the Netherlands.

Data on Dutch GPs are from the Dutch National Study on morbidity and interventions in general practice (Bensing, Foets, Van der Velden & Van der Zee, 1991). In that study, from the 5,826 GPs working in the Netherlands in January 1987 a random stratified sample of 161 GPs was drawn. The sample was stratified by region, urbanization and distance from a hospital. Participating GPs recorded their consultations with sick fund insured patients during three months each in four subsequent groups, from April 1987 through March 1988 (n=169,501 for consultations, which were unfolded into 197,752 reasons for encounter).

Complaints and diagnoses were always written down by the GP on the Dutch recording form, and coded into ICPC (Lamberts & Wood, 1987) afterwards. The diagnosis-determinedness of services among Dutch GPs is assessed in terms of this diagnostic classification in Chapter 8.

In the Netherlands, the parties involved agreed in principle on the introduction of fees for 4 diagnostic services, 12 curative services, and one container service. These are discussed in Chapter 8 as well. The Dutch recording form generally allows for definition of both these services and the ones measured in the Copenhagen City recording form. Again, exceptions are in Chapter 8.

7.2.2 Method

Basically, our method is to assess dependences among Copenhagen City GPs by estimating regression equations from Copenhagen data, and to apply these to Dutch GPs by inserting Dutch values of independent variables into them. Predicted Dutch values of dependent variables then follow from the equations.

When dependences of changes on characteristics of GPs and their practices are assessed, GPs are the units of analysis. In that analysis we will sum over types of consultations, over single diagnostic, and single curative services, and over referrals to single specialties both in hospital and in private practice. These summed changes will be assessed for each GP, and regarded as characteristics of GPs.

When the dependence of changes on a characteristic of services is assessed, single diagnostic and curative services are the units of analysis. In that analysis we will aggregate over GPs, relating changes averaged over GPs in single services performed to the degrees to which among GPs the performance of those services is determined by diagnoses made. Thus changes are assessed for each service, and regarded as characteristics of services.

To start with, we examine whether changes in numbers of consultations and diagnostic and curative services per 1000 registered group I patients depend on characteristics of GPs and their practices. We examine whether these changes depend on the initial numbers of services per 1000 registered group I patients, on the numbers of group I patients registered with GPs, on the numbers of contacts with group II patients (not registered with a GP by definition), and on year of graduation, and gender of GPs.

We do so because a GP who already has many consultations or performs many services per group I patient, has many such patients on his list or has many contacts with group II patients, may have little opportunity left for any increase. Also, a more recently graduated GP may be more apt to change his style of practice than a GP established for a longer time. And a male GP, who typically is the main breadwinner of a household, may be more sensitive than a female GP to the money-earning opportunities offered by the introduction of fees for services.

Now if consultations, diagnostic or curative services performed e.g. by a not too busy young male GP increase in number, and if these services serve to postpone or avoid referrals, his referrals should decrease in number. Thus next we examine whether changes in numbers of referrals per 1000 registered group I patients depend on changes in numbers of consultations, diagnostic or curative services.

We first examine all these dependences pairwise, removing data on GPs that by inspection behave as outliers first pairwise and then listwise (which results in the removal of data on 6

out of 72 GPs). Next, we examine whether dependences showing up pairwise do so with listwise regression as well. With each step, a .05 significance level is used. The resulting regression equations describe the dependences between characteristics of Copenhagen City GPs that are to be applied to Dutch GPs.

Next it is assessed to what degree the changes in numbers of single services depend on their diagnosis-determinedness. In Chapter 5 (Table 5.6), this dependence between characteristics of services was established with eleven services. Of these services, measuring blood glucose shows an increase of 633%, and by inspection behaves as an outlier. That service is removed from the aggregate data. The resulting regression equation estimates the extent to which the increase in performance of a service depends on its diagnosis-determinedness. The degrees of diagnosis-determinedness of services reflect the medical culture in Copenhagen City. This regression equation will be applied to the Netherlands to predict increases in services performed by Dutch GPs. These increases are predicted from the services' diagnosis-determinedness among Dutch GPs, which reflects Dutch medical culture.

When estimation results from Copenhagen City GPs are applied to Dutch GPs, the differences in sampling method can be expected to influence results. The Copenhagen City data are from metropolitan GPs collecting data voluntarily. For both reasons they can be expected to be a more homogeneous sample than the Dutch one. Dutch GPs were paid a fee for data collection, and are a sample stratified by urbanization. They can be expected to be artificially heterogeneous concerning urbanization, with GPs from less urban areas overrepresented. And indeed, from rural to metropolitan areas the size of the Dutch sample decreases monotonically from 3.4% to 1.3%.
Differences in homogeneity are taken into account here by standardizing independent variables. Since these are numbers per patient, and therefore non-negative, after standardization a more heterogeneous sample can still be expected to yield larger positive skewnesses of distributions. This is taken into account by inspecting skewnesses and qualifying predictions accordingly. To have more ground to stand on when doing so, this is done for the Copenhagen City sample, for the (sub)urban GPs in the Dutch sample (who were established in municipalities of over 50,000 inhabitants, and are 39 in number), and for all of the Dutch sample. The qualifications will prove to imply that only one-sided limiting values of changes can be predicted.
Both in standardizing and in qualifying for skewness, it is assumed that a linear relationship fitted on one range of values of an independent variable holds equally well for another range of values.
When changes in numbers of summed diagnostic and curative services performed are predicted by applying dependences between characteristics of GPs, these will be applied both to the sample of the services that fees were introduced for in Copenhagen City, and to the set of services that fees are agreed to be introduced for in the Netherlands. In doing so it is also assumed that these dependences hold regardless of the difference in content

between the two sets of services.

Finally, a dependence between characteristics of services in Copenhagen medical culture is applied to services in Dutch medical culture. To be able to do so, the diagnosis-determinedness of services in Dutch medical culture is assessed in Chapter 8. This dependence is assessed in the Danish data using the degree to which the performances of a service are determined by complaints and diagnoses presented, as classified in the ICHPPC-2. It is applied using complaints and diagnoses as classified by either ICPC or a clustering thereof. The use of the clustering is explained in Chapter 8. Since the ICHPPC-2 is used in Copenhagen data collection and the ICPC in Dutch data collection, we have to assume that the dependence at hand holds regardless of differences between the two classifications.
And again, the dependence is applied both to the sample of the services that fees were introduced for in Copenhagen City, and to the services that fees are agreed to be introduced for in the Netherlands. Thus finally it is assumed that this dependence holds regardless of the differences in content between the services in these sets.
Changes in numbers performed of single services as predicted from this dependence are summed over both diagnostic and curative services for comparison with the changes predicted from dependences between characteristics of GPs. Since the latter are one-sided limiting values, they are narrowed down by combining them with this final prediction.
To set off any violations of the assumptions discussed, predictions are then turned into simple fractions: a half, a third, a quarter.

7.3 Results

7.3.1 Assessing and applying dependences between characteristics of GPs

In examining on what characteristics of GPs the changes in numbers of consultations, diagnostic and curative services depend among Copenhagen City GPs, we obtained a fairly simple result: they all depend on their own initial numbers only. The more services of any type a doctor performs as measured before the change in payment, the less he increases his performances of that type of service when fees are introduced. To some degree this may reflect the general statistical tendency of regression to the mean in repeated measurements. However, it also indicates that the busier a doctor is in performing a type of service, the fewer opportunities he has for any increase.
This dependence does not hold when different types of services are related to each other: a doctor who has more consultations per 1000 registered group I patients before the payment change does not hold back on the increase in diagnostic or curative services performed, and vice versa. Nor do the other aspects of busyness investigated (number of group I patients registered with a GP, number of contacts with group II patients) have any influence here. Any possible influences of year of graduation and gender of GPs are already caught by the simple relations described.
Changes in number of referrals depend only on changes in consultations, and not on

changes in diagnostic or curative services. More consultations entail more referrals, and fewer entail fewer. More will be said about this when the equations are discussed.

Before quantifying these dependences and applying them to Dutch GPs, we will look into the distributions of our independent variables: initial numbers of consultations and services. Means and standard deviations are in Table 7.1.

Table 7.1: Means and standard deviations of consultations and of 11 diagnostic and 8 curative services that fees were to be introduced for in Copenhagen City (numbers per week per 1000 registered patients)

	Copenhagen GPs		urban Dutch GPs		all Dutch GPs	
	mean	(sd)	mean	(sd)	mean	(sd)
consultations	79.05	(16.88)	78.26	(45.95)	73.50	(70.61)
diagnostic services	3.67	(2.68)	9.92	(7.22)	9.56	(13.46)
curative services	.93	(.88)	2.10	(1.70)	2.49	(2.32)

Compared to Dutch GPs, on the average Copenhagen City GPs initially had somewhat more consultations and performed far fewer services per patient. As we expected, Copenhagen City GPs are more homogeneous than urban Dutch GPs, who are again more so than all Dutch GPs. This is taken into account by standardizing initial numbers of consultations and services. After standardization, the dependences found between characteristics of GPs are described by equations (1) through (4).

(1) $\text{prop.chge}_{\text{cons.pat}} = .0823 - .1474 * z_{\text{cons.pat}}$ (p=.0000)

(2) $\text{prop.chge}_{\text{rfrl.pat}} = .5990 * \text{prop.chge}_{\text{cons.pat}} - .3015$ (p=.0001)

(3) $\text{prop.chge}_{\text{diag.pat}} = .8595 - .4156 * z_{\text{diag.pat}}$ (p=.0160)

(4) $\text{prop.chge}_{\text{cura.pat}} = .8012 - .5879 * z_{\text{cura.pat}}$ (p=.0063)

where
prop.chge = 'proportional change of',
cons = consultations,
rfrl = referrals,
diag = summed diagnostic services in the sample,
cura = summed curative services in the sample,
.pat = 'number per week per 1000 registered patients',
z = 'standardized values of'.

In equations (1), (3) and (4) coefficients are negative. This means that a doctor more active in performing each type of service as measured before the change in payment system increases that activity to a smaller extent. In equation (2) the coefficient is positive, indicating that more consultations entail more referrals, and fewer entail fewer.
As stated in the opening section, the introduction of fees for consultations and services in Copenhagen City caused no change in consultations, an increase in services performed, and a decrease in referrals. The decrease in referrals can be caused only by the increase in services that was caused by the introduction of fees for them. Yet at the level of individual Copenhagen City GPs this dependence is not repeated. According to equation (2), a decrease in referrals is either explained by a decrease in consultations, or 'explained' by a negative constant. The direct and substantial explanation of the decrease in the number of referrals is therefore not corroborated.
Copenhagen City GPs have recorded their consultations for periods of one week only. Sampled services rendered by a GP during a week will be heavily influenced by random fluctuations in complaints presented to him. Over GPs these random influences may well be canceled out, while comparing GPs they are not. If so, the negative constant in equation (2) represents the total impact on referrals of both services increasing in number and complaints presented randomly.

Before applying the equations to Dutch GPs, we will look into the standardized distributions of initial numbers of consultations and services. Skewnesses are in Table 7.2.

Table 7.2: Skewness of (standardized) distributions of consultations and of 11 diagnostic and 8 curative services (numbers per week per 1000 registered patients).

	Copenhagen GPs	urban Dutch GPs	all Dutch GPs
consultations	+.23	+2.28	+6.28
diagnostic services	+.92	+1.38	+6.95
curative services	+.76	+1.81	+5.39

All distributions have a right-hand tail, as is to be expected with numbers per patient that are non-negative by definition. From left to right in the table skewness increases heavily, while heterogeneity of samples increases, as shown in Table 7.1. In our samples of Dutch GPs there are far more cases in the right-hand tails of distributions than in the sample of Copenhagen GPs from which equations (1) through (4) have been estimated. Thus when the equations are applied to Dutch GPs, the negative terms in equations (1), (3) and (4) will tend to be too large. Therefore, applying the equations to Dutch GPs the changes in numbers of consultations, services and referrals predicted will tend to be biased to the negative side: decreases predicted will tend to be too large, and increases predicted will tend to be too small. Predictions will have to be qualified for that.

Now to predict changes among Dutch GPs, their measured numbers of services have been

standardized. Using these standardized values, for each Dutch GP proportional changes are predicted from equations (1) through (4). Next, these predicted proportional changes are applied to the non-standardized measured numbers to predict future numbers for each GP. Averages over doctors of measured and predicted numbers are reported in Table 7.3.

Table 7.3: Changes in consultations, referrals, and 11 diagnostic and 8 curative services among Dutch GPs, as predicted by eqs. 1-4 (numbers per week per 1000 registered patients).

		measured	predicted	predicted change
consultations	urban GPs	78.26	78.10	-0%
	all GPs	73.50	69.20	-6%
referrals	urban GPs	5.11	3.69	-28%
	all GPs	4.71	3.19	-32%
diagnostic services	urban GPs	9.92	15.52	+56%
	all GPs	9.56	12.22	+28%
curative services	urban GPs	2.19	2.97	+36%
	all GPs	2.49	3.14	+26%

Indeed, in the sample of all Dutch GPs, where distributions are even more skew to the right than in the sample of urban Dutch GPs, decreases are larger and increases are smaller. The minimum size of the negative bias produced by skewness can be estimated from the predictions on consultations. Applying that estimate in qualifying the predictions on services and referrals, one-sided limiting values of changes can be predicted.

For consultations, decreases are predicted in Table 7.3. We claim that their changes in number should be non-negative. When fees are introduced for consultations, they either do not change in number because of patient-initiatedness, or they increase in number. Then it follows from Table 7.3 that the minimum negative bias is less than .5% for urban GPs, and is 6% for all GPs. From that and Table 7.3 it follows that:

- the maximum real decrease in referrals is 28% for urban GPs and 26% for all GPs,
- the minimum increase in diagnostic services is 56% for urban GPs and 34% for all GPs, and
- the minimum increase in curative services is 36% for urban GPs and 32% for all GPs.

For referrals and curative services, percentages for urban and all GPs do not differ much. For diagnostic services they do. Table 7.2 shows that for diagnostic services skewness among all GPs is about five times as large as among urban GPs, while with consultations and curative services it is about three times as large. Skewness of the distribution of the number of consultations affects the number of referrals predicted by equation (2). And consultations are our yardstick in correcting for skewness. Hence while our correction for negative bias in predicted change for all GPs can be trusted with referrals and curative serv-

ices, with diagnostic services it can be suspected to be insufficient. The 34% minimum increase in diagnostic services for all GPs may well be too low. The 56% minimum increase for urban GPs seems to be the better prediction here.
We now allow for that, and turn predictions into simple fractions to set off any violations of the assumptions discussed in the previous section. Then these predictions follow from applying the regression equations estimating dependences between characteristics of GPs. If in the Netherlands fees were introduced for consultations and for the set of services that fees are paid for in Denmark, then for Dutch GPs:

- referrals would decrease by a quarter at the most,
- the set of diagnostic services would increase by at least half, and
- the set of curative services would increase by at least a third.

To narrow down predictions on the sets of diagnostic and curative services involved, we now switch to assessing and applying dependences between characteristics of single services, aggregating over GPs.

7.3.2 Assessing and applying dependences between characteristics of services

Our earlier analyses showed that the more the performance of a service is diagnosis-determined among GPs, the smaller is its increase when fees for services are introduced. As estimated for ten sampled services performed by Copenhagen City GPs, this dependence is described by equation (5).

$$(5)\ \text{prop.chge}_{\text{service.pat}} = 2.32 - 3.75 * \text{diagdet}_{\text{service}} \qquad (p=.0063)$$

where
prop.chge = 'proportional change of',
service.pat = number of performances of a service per week per 1000 registered patients, averaged over GPs,
diagdet = diagnosis-determinedness among GPs, of a service in the sample.

To be able to apply equation (5) to Dutch GPs, in Chapter 8 we have assessed the diagnosis-determinedness among Dutch GPs of the sample of services that fees are paid for in Denmark. Thus the data used in doing so are Dutch.
Each service's diagnosis-determinedness among Dutch GPs as assessed in Chapter 8 is entered into the equation. The resulting predicted proportional change is applied to GPs' average number of performances of the service. The resulting predicted average numbers of services are summed over diagnostic and curative services for comparison with the predictions in Table 7.3, which are based on dependences between characteristics of GPs.
Equation (5) cannot be applied to one diagnostic service included in the set to which equation (3) was applied*. For the sake of comparability equation (3) is applied again to the

* The service of making a cervical smear can be defined in Dutch data, but its diagnosis-determinedness cannot be assessed, as is explained in Chapter 8.

smaller set of diagnostic services. Measured numbers and numbers predicted by applying equations (3), (4) and (5) are in Table 7.4.

Table 7.4: Changes in 10 diagnostic and 8 curative services among Dutch GPs, as predicted by eqs. 3-4 vs. eq. 5 (numbers per week per 1000 registered patients).

		measured	by equations 3-4: predicted	by equations 3-4: predicted change	by equation 5: predicted	by equation 5: predicted change
diagnostic services	urban GPs	9.44	14.80	+57%	17.49	+85%
	all GPs	9.12	12.00	+32%	17.10	+88%
curative services	urban GPs	2.19	2.97	+36%	2.76	+26%
	all GPs	2.49	3.14	+26%	3.21	+29%

Table 7.4 shows that changes predicted by equation (5) are hardly different for the samples of urban GPs and all GPs. Comparison with Table 7.3 shows that the exclusion of one diagnostic service from the set that equation (3) was originally applied to, hardly makes any difference for the changes predicted. Therefore, the predictions in Table 7.4 can be compared with the predictions made before.

Before, we predicted that diagnostic services will increase by at least half. Here we find that they will increase by about seven eighths. We now narrow down the first prediction by combining it with the second one, and turn predictions into simple fractions again. Then this prediction follows from applying dependences between characteristics of both GPs and single services. If in the Netherlands fees were introduced for the set of diagnostic services that fees are paid for in Denmark, then for Dutch GPs:

- these diagnostic services would increase by at least half, and by doubling at the most.

Also, we predicted that curative services will increase by at least a third. Here we find that they will increase by between a quarter and a third. Thus in terms of simple fractions, the first prediction on curative services is completely narrowed down by combining it with the second one. If fees were introduced for the Danish set of curative services, then among Dutch GPs:

- these curative services would increase by about a third.

Now to make more relevant but also more uncertain predictions, we switch to applying dependences of changes on characteristics of GPs and single services to the diagnostic and curative services that fees are agreed to be introduced for in the Netherlands. In doing so, we assume that equations (3) through (5) hold regardless of differences in content between services. Hence we assume that the equations hold regardless of whether they are applied to the sample of Danish services or to the set of Dutch services.

7.3.3 Applying dependences to other services

Before applying the dependences between characteristics of GPs, described by equations (3) and (4), to this other set of services performed by Dutch GPs, we look into the standard-

ized distributions of initial numbers. Skewnesses are in Table 7.5.

Table 7.5: Skewness of distributions of 5 diagnostic and 12 curative services that fees are to be introduced for in the Netherlands.

	urban GPs	all GPs
diagnostic services	+3.30	+3.09
curative services	+1.60	+4.25

Here diagnostic services show an irregularity: in the generally more heterogeneous sample of all GPs, skewness to the right is smaller. Apparently, GPs in the right hand tail of the distribution of these diagnostic services are mainly urban.
Equation (3) on diagnostic services was estimated from a sample with skewness +.92, and equation (4) on curative services came from a sample with skewness +.76, as Table 7.2 showed. In Table 7.5 skewnesses are larger. Hence in applying equations (3) and (4) here, predictions will be negatively biased again. They will be less negatively biased where skewnesses are indeed larger, but are so to a smaller extent. Thus with diagnostic services, predictions on all GPs will be less negatively biased. And with curative services, predictions on urban GPs will be less negatively biased.
To be able to apply the dependence between characteristics of single services described by equation (5) to the services in this other set as performed by Dutch GPs, in Chapter 8 we have assessed the diagnosis-determinedness of these services among Dutch GPs. Again, each service's diagnosis-determinedness as assessed in Chapter 8 is entered into the equation.
For these services, measured numbers and numbers predicted by applying equations (3), (4) and (5) are in Table 7.6.

Table 7.6: Changes in 5 diagnostic and 12 curative services among Dutch GPs, as predicted by eqs. 3-4 vs. eq. 5 (numbers per week per 1000 registered patients).

		measured	by equations 3-4:		by equation 5:	
			predicted	predicted change	predicted	predicted change
diagnostic services	urban GPs	.74	.88	+19%	1.41	+91%
	all GPs	.74	1.04	+41%	1.43	+93%
curative services	urban GPs	.86	1.19	+38%	1.26	+47%
	all GPs	1.15	1.35	+17%	1.68	+46%

The differences between the changes predicted with equations 3-4 and equation 5 tend to be larger here than they were in Table 7.4. We attribute this phenomenon to the more uncertain character of these predictions: they rest on the assumption that the equations hold regardless of differences in content. The greater differences between changes indi-

cate that this assumption does not fully hold.
This difference is reduced, however, when we take into account the negative biases in the changes predicted with equations (3) and (4). Predicting changes with equations (3) and (4), the 41% increase in diagnostic services among all GPs will still be negatively biased, as will the 38% increase in curative services among urban GPs.
We now allow for that, combine predictions, and turn them into simple fractions again. By turning predictions into simple fractions we further reduce the difference discussed. Now this prediction follows. When in the Netherlands fees are introduced for the diagnostic and curative services agreed upon:

- these diagnostic services will increase by at least half and by doubling at the most, and
- these curative services will increase by about half.

This prediction stands if, by taking the subsequent steps just described, we do full justice to the uncertainties discussed in the previous section.

7.4 Discussion

As was quoted in the opening section, the introduction of fees for services in Copenhagen City caused an increase by two thirds in sampled diagnostic services, and an increase by four fifths in sampled curative services performed, together with a decrease by one fourth in referrals. We argued that the decrease in referrals falls short of the increase in services because services performed are not fully determined by diagnoses made among GPs. We asked what the relative referral-reducing effect will be when fees for services are introduced in the Netherlands.
We first answer that question assuming that in the Netherlands fees would be introduced for the sets of diagnostic and curative services that fees are paid for in Denmark. If so, we predicted that in the Netherlands these diagnostic services would come to be performed one and a half time to twice as often, and the performance of these curative services would increase by about a third, while referrals would decrease by up to a quarter.
Thus for diagnostic services the referral-reducing effect would seem to be in the same order as it was in Copenhagen City. For curative services it would seem to be larger. Inspection of the findings in Chapter 8 shows what might be a substantial explanation for that: among Dutch GPs the performance of these curative services is generally more diagnosis-determined (.20-.90) than the performance of these diagnostic services is (.02-.47).
Next, we answer the question assuming that fees will be introduced for the diagnostic and curative services agreed upon in the Netherlands. If so, we predicted that these diagnostic services would again come to be performed one and a half times to twice as often, but that these curative services would increase by about half. We claim that this difference is stripped of artefacts by the steps described in the previous section. Now will this larger increase in these curative services entail a larger decrease in referrals as well? Will the referral-reducing effect be maintained?

If that effect depends on diagnosis-determinedness of services, it will not. Inspection of the findings in Chapter 8 shows that among Dutch GPs the performance of the curative services agreed upon in the Netherlands is generally less diagnosis-determined (.17-.72) than their performance of the Copenhagen sample of curative services is (.20-.90).

Looking back at our results, the proportional increases in services performed that we predict are quite large. A priori it may seem implausible that diagnostic services would come to be performed no less than twice as often when fees are introduced. However, in Copenhagen we found an increase in the same order of magnitude. And moreover, what we predict are large proportional increases in small numbers. Fees are agreed to be introduced for five diagnostic services in the Netherlands. Together these are now performed .74 times per week per 1000 registered patients by Dutch GPs, as Table 7.6 showed. Thus a GP with 2700 registered patients now performs any one of these services twice a week. We predict that he will come to do so three or four times a week when fees are introduced.

In making our predictions we left two issues aside. For one, we left aside the value placed on monetary rewards by the GPs in the Copenhagen City sample on which our results are based. And for another, we left aside the part of income made up by fees for services.
As we noted in Section 7.2, the Copenhagen City data are from metropolitan GPs collecting data voluntarily. Now it seems implausible that their metropolitan character has influenced results. In fact, in cross-sectional analysis substitution between services and referrals has been assessed among urban and rural Dutch GPs alike (Groenewegen, 1990: 610). From their voluntary co-operation, however, they can be expected to place a low value on monetary rewards (Krasnik et.al., 1990: 1700). It does seem plausible that the dependences we assessed are weakened by this. This might imply that the changes we have predicted are underestimates.
On the other hand, the payment system introduced in Copenhagen City implies that the average GP gets about half his income from capitation fees and about half from fees for consultations and services. Now if about half a GP's income depends on a routine of billing for consultations and services, he has good reason to incorporate such a routine in daily practice. However, a far smaller part of income would depend on billing for the set of services agreed upon in the Netherlands. In such a setting, a GP would have less reason to incorporate the performance and billing of these services in daily practice. If so, the changes we predicted are overestimates.

These uncertainties add to the ones discussed in Section 7.2. Yet, amidst all uncertainty, from our results is seems advisable to anticipate a strong dependence of referral reduction on diagnosis-determinedness of services that fees are introduced for. To maximize referral reduction, fees should be introduced for strongly diagnosis-determined services, and their diagnosis-determinedness should be guarded and enhanced.
The diagnosis-determinedness of a service stands for the professional certainty of GPs

about when to perform it, as was noted in the introduction. That certainty can be argued to vary with the strength of the scientific foundation of standards, as perceived by the medical community (e.g.: Wennberg, 1987; cf. Chapter 5). Thus fees should be introduced for services with a strong scientific foundation of standards on when to perform them. Moreover, such standards should be published, and adherence to them should be guarded and enhanced by postgraduate education and peer review.

8. Assessing the diagnosis-determinedness of services among Dutch general practitioners*

8.1 Introduction

In Chapter 5 we assessed the diagnosis-determinedness among Copenhagen City GPs of a sample of the services that fees were introduced for in Copenhagen City in October 1987. Copenhagen City GPs were the units of analysis in assessing a service's diagnosis-determinedness, which was then regarded as characteristic of the service in Copenhagen medical culture.
Here we will assess the diagnosis-determinedness of these services among Dutch GPs, using data from the Dutch National Study on morbidity and interventions in general practice (Bensing, Foets, Van der Velden & Van der Zee, 1991). Thus here Dutch GPs are the units of analysis, and the resulting diagnosis-determinedness characterizes the same services in a different context. It is regarded as characteristic of the services in Dutch medical culture.
In the Netherlands, parties have agreed that fees will be introduced for a set of services partly different from the sample mentioned. The diagnosis-determinedness of these services will be assessed as well.
In Section 8.2 the services under study are discussed and the procedure followed in assessing their diagnosis-determinedness is introduced. In Section 8.3 the procedure is applied, and in Section 8.4 its content validity is evaluated.

8.2 Data and method

8.2.1 Data

The sample of the services that are now reimbursed in Copenhagen City consists of 13 diagnostic and 8 curative services. For 7 out of 13 diagnostic services, diagnosis-determinedness among Dutch GPs will be assessed service by service in Section 8.3. One diagnostic service identified as such by Danish health authorities, 'bacterial and fungal culture', was represented by two major parts in Copenhagen City data collection: 'streptoculture' and 'urine culture'. Of these two, 'streptoculture' cannot be defined in Dutch National Study data, and 'urine culture' can only be defined together with two other services in the Copenhagen City sample: 'urine test with sticks' and 'urine culture with sensitivity'. The result is called 'urine tests' in Section 8.3. The three sample services not covered at all are making a cervical smear, proctoscopy, and inoculation for culture. The latter two services cannot be

---

* This chapter serves as an appendix to Chapter 7.

defined in Dutch data. The first one can, but its diagnosis-determinedness cannot be assessed, as will be explained in the next subsection.
All 8 sample curative services are covered one by one in Section 8.3, except for 'removing foreign bodies from eye, ear, nose or throat' and 'from skin or from under nail'. These are combined into 'removing foreign bodies' to avoid small numbers.
In the Netherlands, parties have agreed on the introduction of fees for four diagnostic services, twelve curative services, and one container service.
The container service is check-up and counseling with type II diabetes mellitus. It could not be defined as such in Dutch National Study data. However, a central element in it is blood glucose measurement. This diagnostic service is taken to stand for the container service here. It is included in the sample of services reimbursed in Copenhagen City.
Of the four diagnostic services that fees are agreed to be introduced for, one is making an electrocardiogram. This service is included in the Copenhagen City sample as well. The other three diagnostic services will be dealt with separately in Section 8.3.
Of the twelve curative services, four are again included in the Copenhagen City sample. These are stitching an open wound (or treating a large wound), bladder catheterization, and the removal of foreign bodies from both groups of locations mentioned above. One other curative service is part of a more broadly defined service included in the Copenhagen sample. 'Incision of an abscess' is part of 'incision or excision of an abscess or tumour'. Here the more broadly defined service in the Copenhagen sample is taken to stand for its part. The remaining seven curative services will be dealt with separately.
The services whose diagnosis-determinedness will consequently be assessed are listed in Table 8.1.

8.2.2 Method

To determine the degree of diagnosis-determinedness of these services, the classification of complaints and diagnoses in the International Classification of Primary Care (ICPC) is used, in the version that was developed in the Dutch National Study (Van der Velden et al., 1989)*.
For each service, a choice was made between that classification itself and a clustering thereof that was developed in the study mentioned (Claessens & Van der Velden, 1990). In that clustering, ICPC codes of complaints and diagnoses are ordered into a four-level hierarchy of organic system, general, and specific ethiological category, and full description. Of the 17 organic systems distinguished, 11 will be referred to in abbreviated notation in the tables in the next section. These are: general and unspecific (gen); digestive system (dig); ear (ear); circulatory system (circ); musculoskeletal system (musc); respiratory system (resp); skin and subcutaneous tissue (skin); urinary tract (urin); pregnancy, childbearing and family planning (or 'reproduction') (repr); female genital system (fem); and male genital system (male).

---

* ICPC codes represent complaints presented, diagnoses made or services rendered. Here the ICPC code representing a service rendered is used to define making a cervical smear. As a consequence, the diagnosis-determinedness of the service cannot be assessed.

Table 8.1: The services to be analyzed

reimbursed in Copenhagen	to be reimbursed in the Netherlands
diagnostic	**diagnostic**
taking a blood sample	electrocardiogram(2)
pregnancy test	blood glucose measurement(1)(2)
electrocardiogram	examining vision
haemoglobin measurement	examining vaginal discharge
ESR measurement	audiometry
blood glucose measurement	
urine microscopy	**curative**
urine tests(1)	removing foreign bodies(1)(2)
	in/excision of an abscess/tumour(1)(2)
	stitching an open wound(2)
curative	bladder catheterization(2)
removing warts	bandaging a sprained ankle
removing ear wax	excising a sebaceous cyst
removing foreign bodies(1)	excising a nevus or mole
in/excision of an abscess/tumour	excising wedge of ingrown toenail
treating a large wound	puncture/injection with bursitis
bandaging	adjusting/removing an IUD
bladder catheterization	adjusting/removing a pessary

1 modified for measurement

2 also in the Copenhagen sample

For each service, the choice between the ICPC and its clustering was made as follows. With each of the two classifications, complaints and diagnoses are selected that were made in at least 2.5% of the instances in which the service was performed. Then the classification was chosen that best met the criterion that no more than ten complaints and diagnoses cover at least half of the instances in which the service was rendered. The complaints and diagnoses selected within the classification chosen were further analyzed.

Next, we assessed the degree to which the performance of a service is fixed among GPs with their assessment of these complaints and diagnoses. That degree is measured as the conventional coefficient of determination R^2 over doctors, of performances of a service, by assessments of complaints and diagnoses. The logic of that procedure is reasoned out in Chapter 5. Here we will only introduce its application in the tables to follow. We will do so by taking as an example the service in the first table in the next section: taking a blood sample. With that service, the ICPC clustering was chosen as classification to be used in the analysis, as is implied by the abbreviations 'resp' for respiratory system, 'gen' for general and unspecific, and 'musc' for musculoskeletal system in the upper lines of the table. Since the hierarchical ordering is not complete in the clustering, not all complaints and diagnoses in the table have a prefix for the organic system involved.

These are three of the eight categories of disorders that cover more than 2.5% of the instances in which a blood sample was taken: 'resp:infections', or infections of the respiratory system; 'gen:infirmity/fatigue', or infirmity or fatigue as a general and unspecific dis-

order; and 'musc:pain', or pain in the musculoskeletal system. The results concerning these three will be introduced here. Each represents one of the three types of outcome to be met in the next section.

Infections in the respiratory system were established in 21,252 reasons for encounter (or rfe's). In 370 of these, a blood sample was taken. Now for each doctor total numbers are assessed of the times a blood sample was taken, and of the times infections in the respiratory system and the other disorders listed were established. Then a simple regression analysis over doctors is performed of the number of times that the service was rendered on the numbers of times that the disorders were established. For infections in the respiratory system the non-standardized regression coefficient is +.13. On a 5% significance level, this is a significant coefficient, as is implied by the boldface printing. Thus doctors who establish infections in the respiratory system more often take a blood sample more often. We take that as evidence that doctors agree that infections in the respiratory system are a good reason for taking a blood sample.

Secondly, infirmity or fatigue as a general and unspecific disorder was established in 916 reasons for encounter. In 265 of these, a blood sample was taken. The regression over doctors of blood samples taken on disorders established produces a non-significant coefficient of +.02 here. We take that as evidence that doctors do not agree on infirmity or fatigue as a reason for taking a blood sample.

Thirdly, pain in the musculoskeletal system was established in 13,555 reasons for encounter. In 189 of these, a blood sample was taken. Regression produces a significant negative coefficient of -.15 here. The sign of the coefficient shows that doctors who establish pain in the musculoskeletal system more often take a blood sample less often. So while sometimes blood samples are taken with pain in the musculoskeletal system, doctors who are more familiar with the complaint avoid doing so. The significance of the negative coefficient is taken as evidence that, generally, doctors agree that pain in the musculoskeletal system is a bad reason for taking a blood sample. And indeed, experience shows that most of the time it provides no information in that case.

Regression produces a significant overall degree of determination R^2 of .47. Hence between the disorders listed doctors agreed to a degree of .47 that they are a good or bad reason for taking a blood sample. Or among doctors, the taking of a blood sample is fixed with their assessment of these disorders to a degree of .47. To put it briefly: among doctors, taking a blood sample is diagnosis- (or complaint-) determined to that degree.

8.3 Results

8.3.1 Services reimbursed in Copenhagen City: diagnostic services

Table 8.2: Taking a blood sample

	neurotic disorders	diabetes mellitus	resp:infections	hypertension
rfe's	8,978	2,207	21,252	10,996
performances	467	373	370	285
dependences	+.04	**+.72**	**+.13**	**+.14**

	gen:infections	gen:infirmity/fatigue	anaemia	musc: pain
rfe's	2,567	916	1,628	13,555
performances	280	265	215	189
dependences	**+.18**	+.02	-.13	**-.15**

	other diagnoses	R^2
rfe's	135,653	
performances	3,276	
dependences		**.47**

Besides the disorders discussed above, diabetes mellitus, hypertension and infections in general are agreed to be good reasons for taking a blood sample among Dutch GPs. Agreement between doctors as shown by the figures reflects common medical knowledge.

Table 8.3: Pregnancy test

	pregnancy	menstrual cycle disorders	pregnancy complications	other diagnoses	R^2
rfe's	1,775	1,411	360	194,206	
performances	354	81	27	146	
dependences	+.03	**+.24**	**+.39**		**.39**

Menstrual cycle disorders, and pregnancy complications (such as blood loss in early pregnancy) are agreed to be good reasons for performing a pregnancy test.
With 'pregnancy' the apparent disagreement may well be an artefact of the fact that reasons for encounter are the units of analysis here. Generally, a pregnancy test is indicated only once during an episode of pregnancy. However, that episode may encompass various numbers of consultations, all containing pregnancy as one recurrent reason for encounter. Given the artefact, performing a pregnancy test is somewhat less determined by conditions established (.39) than taking a blood sample is (.47).

Table 8.4: Electrocardiogram

	circ:de-generation	cardiac arrhythmia	neurotic disorders	hyper-tension
rfe's	4,421	670	8,978	10,996
performances	18	15	10	7
dependences	+.01	-.03	-.00	-.00

	circ: anxiety	musc: pain	other diagnoses	R^2
rfe's	205	13,555	158,927	
performances	6	4	16	
dependences	**+.11**	+.00		.02

Only anxiety concerning the circulatory system is agreed to be a good reason for a GP to make an electrocardiogram. Reassurance may well be the main reason for doing so in general practice. As the insignificant (non boldface) R^2 of .02 shows, over the disorders listed making an electrocardiogram is simply not determined by disorders established at all.

Table 8.5: Haemoglobin measurement

	anaemia	neurotic disorders	pregnancy	gen:infirm-ity/fatigue
rfe's	1,628	8,978	1,775	916
performances	508	198	133	125
dependences	**+.49**	-.00	+.15	-.04

	resp:in-fections	menstrual cycle disorders	repr:pre-vention	other diagnoses	R^2
rfe's	21,252	1,411	1,105	160,687	
performances	81	56	55	768	
dependences	+.03	**+.41**	-.05		**.23**

Haemoglobin measurement is specific to anaemia, and menstrual cycle disorders may well lead to anaemia. Doctors agree on both as good reasons for measuring haemoglobin. Disagreement on the other disorders as reasons for doing so brings the procedure's determinedness by disorders established to a relative low of .23.

Table 8.6: ESR measurement

	resp:in-fections	musc: pain	neurotic disorders	gen:infirm-ity/fatigue	
rfe's	21,252	13,555	8,978	916	
performances	154	77	68	60	
dependences	-.01	+.02	+.05	-.19	
	musc:immunity/ system disorders	**gen:in-fections**	**dig:in-fections**	**dig: pain**	
rfe's	1,226	2,567	2,767	1,437	
performances	49	44	36	35	
dependences	-.00	+.06	**+.44**	+.14	
	dig:non-infectious inflammations	**irritable bowel**	**urin:in-fections**	**other diagnoses**	**R^2**
rfe's	2,749	1,486	2,636	138,183	
performances	33	32	30	371	
dependences	-.13	-.18	+.06		**.15**

ESR measurement is a non-specific procedure. Only when infections of the digestive system are suspected is ESR measurement a simple means for either rejecting or confirming that diagnosis. Infections of the digestive system are the only (suspected) disorder that doctors agree on as a good reason for measuring ESR. Disagreement on the other disorders brings the procedure's determinedness by disorders to a further low of .15.

Table 8.7: Blood glucose measurement

	diabetes mellitus	other diagnoses	R^2
rfe's	2,207	195,545	
performances	1,002	407	
dependences	**+.55**		**.39**

Blood glucose measurement is specific to diabetes mellitus, which is all that the figures express.

Table 8.8: Urine microscopy

	urin:in-fections	dig: pain	musc: pain	kidney/ureter/ bladder stone
rfe's	2,636	1,437	13,555	319
performances	1,584	160	154	143
dependences	**+.93**	**+.34**	+.00	+.91

	irritable bowel	male:in-fections	other diagnoses	R^2
rfe's	1,486	482	177,837	
performances	124	104	1,438	
dependences	-.06	+.78		**.43**

Urine microscopy is indicated with infections of the (genito)urinary tract. And when a female patient has pain in the abdomen, urine microscopy is indicated to rule out the possibility of urinary infections. The figures reflect agreement among doctors on that.

Table 8.9: Urine tests

	urin:in-fections	diabetes mellitus	pregnancy	contra-ception
rfe's	2,636	2,207	1,775	5,687
performances	1,153	686	573	443
dependences	**+.55**	**+.50**	**+.43**	+.13

	repr:pre-vention	other diagnoses	R^2
rfe's	1,105	184,342	
performances	185	1,899	
dependences	+.36		**.35**

With urinary infections a urine test is indicated to identify the class of infecting organism. With diabetes mellitus, the urine sugar level is assessed with sticks. And with pregnancy both testing for urinary infections and assessing the urine sugar level are proper preventive actions.

8.3.2 Services reimbursed in Copenhagen City: curative services

Table 8.10: Removing warts

	warts	other diagnoses	R^2
rfe's	1,463	196,289	
performances	862	0	
dependences	**+.83**		**.82**

Concerning warts the figures express agreement on the obvious among Dutch GPs. This brings the procedure's determinedness by conditions assessed to a high of .82.

Table 8.11: Removing ear wax

	ear wax	ear:in-fections	other diagnoses	R^2
rfe's	1,881	4,305	191,566	
performances	1,567	206	118	
dependences	**+.93**	**+.03**		**.88**

Concerning ear wax, agreement on the obvious shows up once more. It produces a further high in the procedure's determinedness of .88.
With infections of the ear, doctors agree that in some instances removing ear wax is indicated, as shown by the small but significant coefficient. No doubt these are instances in which either the external ear is infected, or in which wax hampers the inspection of the ear drum to decide on infection of the middle ear.

Table 8.12: Removing foreign bodies

	foreign body in eye	foreign body in skin	foreign body in ear	other diagnoses	R^2
rfe's	489	123	28	197.112	
performances	84	59	7	0	
dependences	**+.16**	**+.46**	**+.70**		**.32**

With foreign bodies, doctors agree that removal by the GP is indicated in part of the instances. That part is smallest with foreign bodies in the eye, whose removal requires the skill least common among GPs. This produces a degree of determinedness far smaller than with removing warts and ear wax.

Table 8.13: In/excision of abscess/tumor

	benign neoplasm skin	boil/carbuncle/ cellulitis	lipoma skin	other diagnoses	R^2
rfe's	393	688	195	196,476	
performances	162	107	22	3	
dependences	**+.51**	**+.13**	+.15		**.51**

With a benign neoplasm of the skin, excision is indicated. With boils, carbuncles and cellulitis, incision is indicated only when an abscess develops from them. Doctors agree that with boils, carbuncles and cellulitis minor surgery is indicated in a smaller part of the instances than with benign neoplasms.

Table 8.14: Stitching an open wound

	laceration/ cut	bite	other diagnoses	R^2
rfe's	999	116	196,637	
performances	775	55	0	
dependences	**+.78**	**+.60**		**.90**

With the stitching of an open wound agreement on the obvious shows up again. It produces a final high in the procedure's determinedness of .90. With bites, complications are more probable than with cuts and lacerations.

Table 8.15: Bandaging

	musc: traumata	circ:non-infectious inflammations	skin: traumata	musc:non-infectious inflammations
rfe's	7,450	742	3,703	4,275
performances	1,040	175	162	132
dependences	**+.32**	**+.52**	+.12	+.07

	musc: pain	skin:infections	circ:vascular disorders	other diagnoses	R^2
rfe's	13,555	7,202	1,276	159,549	
performances	84	68	64	279	
dependences	**-.11**	+.06	+.02		**.42**

Doctors agree that traumata in the musculoskeletal system (such as sprained ankles) and non-infectious inflammations of the circulatory system (such as a crural ulcer) are good reasons for bandaging. They also agree that pain in the musculoskeletal system is a bad reason for bandaging. Joint pains belong to that category, and with many joints bandaging is either inadvisable or impossible. Over disorders, a multi-purpose procedure like bandaging is far less determined by disorders established (.42) than single-purpose procedures like removing warts (.82), removing ear wax (.88) and treating a large wound (.90).

Table 8.16: Bladder cathererization

	urine retention	urine incontinence	benign prostate hypertrophy	paralysis
rfe's	31	256	99	103
performances	16	11	6	4
dependences	**+.54**	**+.10**	+.11	**+.10**

	cystitis	other diagnoses	R^2
rfe's	2,346	194,917	
performances	4	39	
dependences	+.01		**.20**

With urine retention, bladder catheterization is indicated. With urine incontinence and paralysis, it is indicated to counteract leaking and bedsores if that outweighs the risk of causing an infection. Doctors agree on that: for all three disorders the coefficients are significant and positive, but for the latter two disorders they are small. As a result, the overall degree of determinedness is low as well.

8.3.3 Services to be reimbursed in the Netherlands: diagnostic services

Table 8.17: Examining vision

	refractive errors	vision complaints	other diagnoses	R^2
rfe's	322	248	197,182	
performances	143	116	0	
dependences	**+.33**	**+.36**		**.40**

If a diagnosis is made when vision complaints are presented, almost always refractive errors will be diagnosed. Doctors agree that both are equally good reasons for examining vision.

Table 8.18: Examining vaginal discharge

	vaginal discharge	other diagnoses	R^2
rfe's	186	197,566	
performances	29	0	
dependences	**+.12**		**.12**

The vision complaints and refractive errors just mentioned always need specific measurement before a remedy is in order. Vaginal discharge, however, can often be cured by anti-infectives without a specific diagnostic examination. Doctors agree that, compared to vision complaints, vaginal discharge is a good reason for the relevant diagnostics in a smaller part

of the instances. As a consequence, examining vaginal discharge is less determined by disorders established (.12) than examining vision (.40).

Table 8.19: Audiometry

	ear:in-fections	ear:de-generation	ear: traumata	resp:in-fections
rfe's	4,305	100	152	21,252
performances	21	5	3	2
dependences	+.00	-.06	**+.18**	+.00

	other diagnoses	R^2
rfe's	171,943	
performances	23	
dependences		**.04**

Doctors agree that traumata of the ear are a good reason for performing audiometry.
With infections of the ear and of the respiratory system, a more serious infection may produce less hearing ability, but it need not do so. Doctors do not agree on these infections as reasons for performing audiometry.
With degeneration of the ear, the apparent disagreement may well be an artefact of the fact that reasons for encounter are the unit of analysis. As was the case with pregnancy tests during pregnancy, generally audiometry is indicated only once during an episode of degeneration. (With continuing degeneration it may be indicated once a year, for example). Still, such an episode may encompass various numbers of consultations, all containing degeneration as one recurrent reason for encounter.
Thus audiometry renders inconclusive evidence with infections and is of infrequent use with degeneration. These are two reasons why not every general practitioner equips his practice with an audiometer to begin with. All of this taken together brings the procedure's determinedness to a further low of .04.

8.3.4 Services to be reimbursed in the Netherlands: curative services

Table 8.20: Bandaging a sprained ankle

	sprained ankle	other diagnoses	R^2
rfe's	1,191	196,561	
performances	472	0	
dependences	**+.59**		**.72**

'Bandaging a sprained ankle' is part of 'bandaging', which is included in the Copenhagen sample of services, and was dealt with above. Yet here the whole is not taken to stand for the part, since agreement on bandaging a sprained ankle can be expected to be stronger

than agreement on bandaging in general. Indeed, determinedness by disorders established is larger here (.72) than it was with bandaging in general (.42).

Table 8.20: Excising a sebaceous cyst

	sebaceous cyst	other diagnoses	R^2
rfe's	800	196,952	
performances	202	0	
dependences	**+.27**		**.51**

A sebaceous cyst is not as acute a disorder as a sprained ankle, while excising it also requires less common skills, more equipment and more organizing than bandaging does. Accordingly, performing the service is less determined by the assessment of the disorder.

Table 8.22: Excising a nevus or more and ordering a pathologic-anatomic test

	nevus/ mole	other diagnoses	R^2
rfe's	255	197,497	
performances	20	0	
dependences	**+.13**		**.21**

Generally, the presence of a nevus or mole is a condition even less acute than a sebaceous cyst. Performing the service is even less determined by the assessment of the condition.

Table 8.23: Excising a wedge of ingrown toenail

	ingrown toenail	other diagnoses	R^2
rfe's	403	197,349	
performances	122	0	
dependences	**+.32**		**.40**

Like a sebaceous cyst, an ingrown toenail is a disorder not as acute as a sprained ankle, and excising a wedge of it requires a skill less common than bandaging. The degree of determinedness is in the middle range again.

Table 8.24: Puncture or injection with bursitis

	bursitis	other diagnoses	R^2
rfe's	515	197,237	
performances	98	0	
dependences	**+.32**		**.45**

With puncture or injection after assessing bursitis, things are in the middle range in all respects once more.

Table 8.25: Adjusting or removing an IUD

	contra-ception	menstrual cycle disorders	sub/in-fertility	fem:in-fections
rfe's	5,687	1,411	279	1,564
performances	170	11	7	6
dependences	**+.01**	+.02	+.06	**+.07**

	other diagnoses	R^2
rfe's	188,811	
performances	37	
dependences		**.17**

When contraception is the reason for encounter, not all the time is an IUD chosen as the contraceptive device. If it is, most of the time its location has to be checked only. Doctors agree that in only a small part of the instances is contraception a good reason for adjusting an IUD: the coefficient is very small, but significant. Hence its smallness may be partly due to the fact that an IUD is not the only contraceptive device to be chosen. And partly it may be an artefact of the fact that reasons for encounter are the units of analysis here, instead of episodes of contraception.
An infection of the female genital system in a woman wearing an IUD may be caused by the IUD, which is then a good reason for removing it. Doctors agree that female infections are a good reason for removing an IUD in a small part of the instances. Evidently these are the instances in which such infections occur in women wearing it.
Both conditions help to bring the determinedness of the service by conditions assessed to a low of .17.

Table 8.26: Adjusting or removing a pessary

	uterovaginal prolapse	stress in-continence	other diagnoses	R^2
rfe's	378	40	197,334	
performances	183	4	43	
dependences	**+.67**	-.26		**.68**

With uterovaginal prolapse, figures show agreement on the obvious again. It produces a relative high in the procedure's determinedness by conditions assessed of .68.
With stress incontinence, adjusting a pessary is one way to counteract it, which may produce complications, however. As figures show, doctors do not agree on stress incontinence as a reason for adjusting a pessary.

8.4 Discussion

Reviewing our results in assessing the diagnosis-determinedness of services, these assessments seem quite valid. In three instances the apparent disagreement on the performance of a service in a condition seems to be an artefact of the fact that reasons for encounter are the units of analysis. These instances are performing a pregnancy test with pregnancy, performing audiometry with degeneration of the ear, and adjusting an IUD with contraception. Otherwise (dis)agreement among doctors as expressed by the figures corresponds to medical knowledge.

In Chapter 5 we noted that we could not decide about the content validity of our measurement of the diagnosis-determinedness of services in Copenhagen medical culture. We attributed that uncertainty to the global character of the diagnostic categories listed on the Copenhagen recording form. Those categories were derived from ICHPPC-2 (Classification Committee, 1983) by selection and combination.

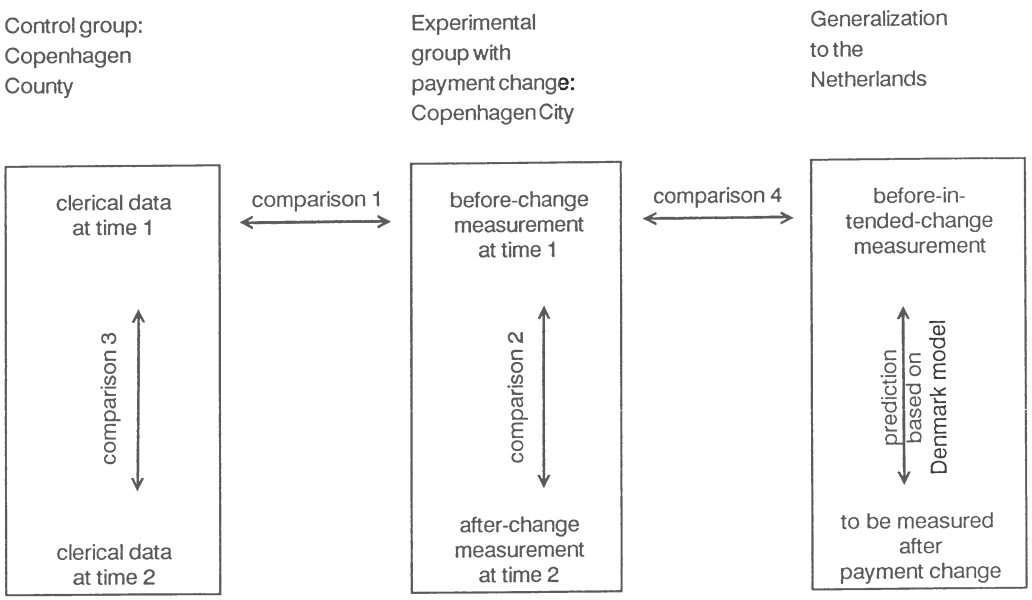
Here that uncertainty on content validity has been overcome. This we credit to the quality of the Dutch National Study data on complaints and diagnoses. These were written down by GPs as they assessed them, and were centrally coded into ICPC by medically trained personnel.

9. Conclusions

9.1 Introduction

In Section 1.4 we outlined our study in figure 1.1, containing a set of comparisons. For ease of reference it is reproduced here as figure 9.1.

Figure 9.1: Design of the study



In Section 9.2, the results from these comparisons are themselves compared. A difference in results then catches the attention. That difference is explained by introducing two hypotheses in Section 9.3. By introducing these hypotheses, the scientific context of the study is elaborated. In Section 9.4 two sets of falsifiers of both hypotheses are identified. It is argued that one or the other set of falsifiers is claimed to hold true in the policy context of the study. And, as an epilogue, the scientific context is evaluated from the perspective of the policy context in Section 9.5.

9.2 Results compared

Comparison 1 in Figure 9.1 was reported in Chapter 3. This comparison is a cross-sectional analysis of differences between the mixed system of remuneration in Copenhagen County and the capitation system in Copenhagen City, as they co-existed before the change in payment system. Generally, a cross-sectional comparison provides only a weak test of hypotheses on differences in numbers of services performed. This one provides a stronger test than most cross-sectional studies reported in the literature, however. There is no apparent patient selection for either of the two conditions, and the broader institutional context of health care is the same in both. From this comparison we concluded that consultations were not held more often in Copenhagen County, where fees were already paid for them. Diagnostic and curative services were performed more often there. Patients were not referred less often however. Thus from the cross-sectional comparison alone, one would conclude that fees for services do not have a referral-reducing effect.

The core of the study, reported in Chapter 4, is the comparison of changes over time in Copenhagen City (as assessed in comparison 2) to those in Copenhagen County (assessed in comparison 3), where no change in payment system took place. In its design the core study is a natural experiment with a control group. This design allows for clear conclusions on the cause of the changes measured. We concluded that the introduction of fees for services in Copenhagen City caused no change in consultations, did cause an increase in diagnostic and curative services, and caused a decrease in referrals that was smaller than the increase in services performed. So, from the natural experiment alone, one would conclude that fees for services do have a referral-reducing effect, be it a limited one.

The cross-sectional comparison showed that diagnostic and curative services were performed three times as often in Copenhagen County as in Copenhagen City, in March 1987. Still, referrals were also somewhat larger in number in Copenhagen County. Fees for these services had been introduced in Copenhagen County in 1961. In Copenhagen City they had not yet been introduced. In the assessment of changes in Copenhagen City we found that services increased by more than half between March 1987 and November 1988, while referrals decreased by a quarter. Fees for services were introduced there in October 1987. If one were to take these comparisons by themselves, thus leaving aside differences in time spans, from the cross-sectional comparison one would conclude that fees for services make a larger difference for numbers of services performed than one would conclude from the assessment of changes. Also, from the cross-sectional analysis alone one would conclude that, if anything, fees for services have a referral-inducing effect, while from the assessment of changes one would conclude that they have a limited referral-reducing effect.

In Chapter 5 the change in Copenhagen City has been analyzed in more detail. From that we concluded that services whose performance was less determined by diagnoses made

among GPs increased more strongly when fees for services were introduced.
We estimated the services' diagnosis-determinedness from their performance one year after the introduction of fees. We recognized that therefore our results could be otherwise interpreted as well. It may be that fees tend to increase the numbers of services performed by increasing the range of health conditions in which they are performed. And it may be that services vary in their openness to such a tendency.
However, we assumed that such an effect of payment system on medical culture takes more than one year. If so, time spans are relevant once more.

As a first step towards estimating the consequences of a change in payment in the Netherlands, in Chapter 6 we compared Copenhagen City before change to urban general practice in the Netherlands. We concluded that differences in morbidity between Copenhagen City and urban Holland did not explain differences in numbers of services performed. Hence differences in numbers of services had to be explained by differences in institutional settings and their distributional characteristics. Next we concluded that urban Dutch GPs tend to perform services more often owing to fewer referring opportunities as a decisive factor.
In Section 1.3 we argued that, where referring opportunities are fewer, the decrease in referrals will fall short of the increase in services to a greater degree. On the other hand, we argued that it will do so to a smaller degree with more certainty about standards and less on task definitions. Certainty about task definitions (defining GPs' versus specialists' tasks) has been left aside here. Certainty about a standard includes certainty about how and when to perform a service. Certainty about how to perform a service has been left aside again. Certainty about when to perform it, or diagnosis-determinedness of a service among GPs, was assessed for Copenhagen City GPs in Chapter 5, and for Dutch GPs in Chapter 8. For the services involved in both assessments, results are compared below. This rounds off our comparison between Copenhagen City before change and the Netherlands.

Table 9.1: Diagnosis-determinedness of services among GPs in Copenhagen City and the Netherlands

	Copenhagen City	Netherlands
haemoglobin measurement	.25	.23
blood glucose measurement	.10	.39
removing warts	.33	.82
in/excision of abscess/tumour	.02	.51
treating a large wound	.44	.90
bandaging	.01	.42

These differences may to some degree be an artefact of differences in data collection. In particular, the smaller number of more broadly defined diagnostic categories in terms of which the degrees of determinedness were assessed for Copenhagen City GPs may have partly caused the generally smaller values that we found for them. Still, with haemoglobin

measurement we found a somewhat larger value for Copenhagen City GPs based on a smaller number (3 vs. 7) of diagnostic categories. Thus at least in part, these are real differences as well.
To that extent, certainty about when to perform a service is generally stronger among Dutch than among Copenhagen City GPs, while referring opportunities are fewer for Dutch GPs. With more certainty the referral-reducing effect of introducing fees for services will be larger, but with fewer referring opportunities it will be smaller than we predicted for the Netherlands.

Our prediction on the effect of the change in payment system in the Netherlands was developed in Chapter 7. It read as follows. When fees are introduced for the set of services that parties have agreed on in the Netherlands, these will come to be performed one and a half time to twice as often, while referrals will decrease by up to a quarter as a result.
Thus for the Netherlands we predicted a referral-reducing effect that is at the most equal in size to the one we assessed in Copenhagen City. From that, the chance that we overestimated it seems to be limited a priori.

9.3 A difference explained: elaborating the scientific context

In Section 1.3 we postulated a utility function for physicians in which the trade-off between two arguments is constrained by a third one: the trade-off between time and money is constrained by compliance with medical standards. Time and money operate more strongly as arguments in that function if standards are less definite. In Chapter 5 it was shown that Copenhagen City GPs did in fact react to the introduction of fees for services by utility maximization according to that function. In Chapter 7 we simulated Dutch GPs' reaction to the introduction of fees for services using the parameters found among Copenhagen City GPs relative to the utility function.
Now we claim that this approach is valid only in the short run. Only in the short run does medical ethics limit the influence of a payment system without any influence of the payment system on medical ethics. In the long run such an opposite influence takes effect as well. This claim clears the way for an explanation of the difference in results between cross-sectional comparison and natural experiment. The cross-sectional comparison by itself led to the conclusion that, if anything, fees for services have a referral-inducing effect. On the other hand, from the natural experiment alone one would conclude that fees for services have a referral-reducing effect, be it a limited one.

To explain that difference, to begin with we introduce an **assumption**: that the introduction of fees for services in Copenhagen County in 1961 had the same initial effect that it had in Copenhagen City in 1987. We therefore assume that the introduction of fees in Copenhagen County some twenty-five years before also resulted in a limited decrease in the number of referrals. If so, it follows from our results that, while the decrease in referrals already falls

short of the increase in services to begin with, after some 25 years it has disappeared completely.

Next, the question should be answered why the decrease in referrals has disappeared. To do so, we introduce two hypotheses. The first one was already touched upon in Section 5.2 and again in the previous section. The second one was touched upon in Section 7.1.

> **Hypothesis 1**: the longer the time that a fee has been paid for a service, the more GPs come to see the service as a means of earning income. And therefore, the more they are inclined to perform the service in a wider range of health conditions, in which its appropriateness becomes questionable. Thus the longer the time that a fee has been paid for a service, the less diagnosis-determined it becomes among doctors.
>
> **Hypothesis 2**: the less diagnosis-determined a service is among GPs, the less they believe it to make a referral redundant, so the less are referrals avoided by performing the service.

Hypothesis 1 describes an effect of payment system on medical culture. Before, we assumed that such an effect will not show in one year. Now we introduce a counterpart **assumption**: that the effects described by both hypotheses will fully show in 25 years, so that no referrals are avoided by the increased performances of the service any more.
This would explain why the referral-reducing effect of the introduction of fees for services has disappeared after some 25 years, as it seems to have done in Copenhagen County.

9.4 The explanation tested: returning to the policy context

When services come to be performed in a wider range of health conditions, in which their appropriateness becomes questionable, quality of care deteriorates. Will this happen inevitably when fees for services are introduced? No, the Dutch GPs' association claims: not if a GP is only partially dependent on fees for services for earning his income, and if his performance of services is monitored by standards (Section 1.2).
Under the Danish mixed payment system, a GP is indeed only partially dependent on fees for services. On the other hand, his performance of services is not monitored by standards. The only norm applied is that the number of services he performs should not exceed averages too much (Section 2.4). So, if our explanation in the previous section holds, it is for lack of monitoring by standards, according to the association's claim.
We claimed a long-term influence of payment system on medical ethics. Here the GPs' association claims that this influence can be counteracted by active monitoring of compliance with ethics. The association claims that our first hypothesis in the previous section can thus be falsified.

By the year 2006, the Dutch GPs' association plans to have made registration and five-year re-registration as a GP conditional on tested ability to perform services according to standards (Section 1.2). This is planned to be the final point of a development in which GPs' performance of services is monitored with gradually increasing strictness.
Participation in peer review and in refresher courses that are focussed on the contents of published standards is planned to be a prerequisite for five-year re-registration in 1996. Participation in formative testing is planned to be added as a prerequisite in 2001. Summative testing is planned to be added in 2006.
Participation in peer review and in refresher courses is already quite common among Dutch GPs. To an increasing degree these courses are focussed on the contents of standards. And increasingly they are equipped with formative tests. Still, the choice of refresher courses taken is a free one as yet. The degrees to which GPs' performance of any service is monitored by standards therefore show quite some variation at present. In this context, a falsifier of our first hypothesis would be:

> among GPs who subject themselves to some minimally strict monitoring of their performance of a service, the service will not become less diagnosis-determined in the course of time, after introduction of a fee for the service.

A corresponding falsifier of our second hypothesis would be:

> among GPs who subject themselves to that minimally strict monitoring of their performance of a service, an initial decrease in referrals to related specialties accompanying an increase in performances of the service will not fade out in the course of time, after introduction of a fee for the service.

As was noted in Section 1.3, the association's internal policy relating to medical standards can be taken to aim at increasing the profession's demonstrable societal value. In its income policy pursued alongside that, in 1990 the association made an agreement with the association of Dutch sick funds. In principle both associations agreed on the introduction of fees for a set of GPs' services. Now this set is partly different from the set of services on which standards have been published. In this context, an even stronger falsifier of our first hypothesis would be:

> among GPs who subject themselves to some minimally strict monitoring of their performance of a service, the service will become more diagnosis-determined in the course of time, whether or not a fee for the service was introduced.

A correspondingly stronger falsifier of our second hypothesis would be:

> among GPs who subject themselves to that minimally strict monitoring of their performance of a service, an initial decrease in referrals to related specialties accompanying an increase in performances of the service will grow relative to that increase in the course of time, whether or not a fee for the service was introduced.

The first set of falsifiers would imply that the long-term influence of payment system on medical ethics can be balanced by active monitoring of compliance with ethics. The second

set would imply that it can thus be overruled.

9.5 Epilogue: evaluating the scientific from the policy context

As the previous section showed, hypotheses developed from the scientific context of our study can be directly linked to policy claims. This follows from the nature of our scientific context.

For one, a utility function is claimed to be equally valid for all members of a profession. Thus, no assumptions are introduced on differences in attitudes or preferences and in perceptions between individual general practitioners. This would have led us to go back into the socialization history of individuals for explanations. In its turn, this would have reduced the testability of explanations owing to reduced availability and quality of (retrospective) data. Likewise, it would have reduced the relevance of our explanations to policy. Policy implications would relate only to future recruitment and socialization of GPs, with delayed effects promised.

Instead, behavioral differences are here claimed to result from differing present resources for and restrictions on utility maximization according to the utility function. These resources and restrictions are claimed to be given with institutional settings. On top of that, long-term overall behavioral changes are claimed to represent long-term effects of institutional settings on the arguments of the utility function. The contrast worked out in the previous section relates to the question what institutional setting is then of decisive influence: the association's monitoring by standards or the payment system.

Whatever the answer, institutional settings are open to policy interventions if anything is.

Verandering van honorering van huisartsen: samenvatting

1. Beleids- en wetenschappelijke context en ontwerp van de studie

De voorgenomen verandering van het Nederlandse stelsel van gezondheidszorg behelst veranderingen in de condities van vraag en aanbod van zorg. Aan de vraagzijde zijn de te verwachten effecten redelijk goed uit onderzoek bekend. Aan de aanbodzijde is dat veel minder het geval. Het betreft dan effecten van een overgang naar contracteervrijheid voor zorgaanbieders en verzekeraars, waarbij de LHV als vertegenwoordiger van huisartsen als zorgaanbieders, pleit voor een uniform stelsel van vergoeding deels per abonnement en deels per verrichting, voor de zorg aan alle patiënten. De te verwachten effecten van de invoering van zo'n vergoedingsstelsel voor huisartsen (via een standaardcontract?) is het onderwerp van deze studie. Daarbij beperken we ons tot de effecten voor de zorg aan ziekenfondspatiënten, voor welke de invoering van zo'n vergoedingsstelsel een overgang zou inhouden van volledige abonnementshonorering naar een gemengde abonnements- en verrichtingenhonorering.
Als eerste aanzet tot zo'n overgang zijn LHV en VNZ in 1990 in principe overeengekomen dat zeventien verrichtingen op het grensvlak van huisarts- en specialistische zorg afzonderlijk zullen worden gehonoreerd. Daarbij is vastgelegd dat de verrichtingen in kwestie dienen te worden uitgevoerd volgens NHG-standaarden, voorzover die inmiddels zijn gepubliceerd. En daarbij wordt geclaimd dat honorering van deze verrichtingen, indien zo uitgevoerd, verwijzingen zal voorkomen.
Dit weerspiegelt de positie van huisartsen als 'onvolkomen professionele agenten' van hun patiënten, die primair het belang van hun patiënten dienen, maar die de speelruimte die dat hen laat zullen benutten om zich een combinatie van geld en vrije tijd te verschaffen die hun belang het beste dient. Daarmee is een nutsfunctie beschreven. Het standaardenbeleid van het NHG kan worden opgevat als beleid om de afwegingen die men volgens die nutsfunctie maakt te veranderen door de speelruimte in kwestie te verkleinen, en zo de aantoonbare maatschappelijke waarde van de beroepsgroep te vergroten.
Hier wordt aangenomen dat huisartsen hun nut maximaliseren volgens de gepostuleerde nutsfunctie, binnen de mogelijkheden daartoe die zijn gegeven met de institutionele en verdelingskenmerken van hun situatie. A priori kan uit die aanname worden afgeleid dat bij overgang van abonnements- naar gemengde honorering de gehonoreerde verrichtingen vaker zullen worden uitgevoerd. Dat zal des te sterker gelden voor verrichtingen waarvoor de speelruimte in kwestie, de professionele onzekerheid, groter is. Ook kan worden afgeleid dat de afname van de verwijzingen zal achterblijven bij de toename van verrichtingen. En dat zal des te sterker gelden naarmate de professionele onzekerheid omtrent verrichtingen groter is, en het aantal specialisten per hoofd van de bevolking kleiner. Of en in welke mate deze effecten zullen optreden, wordt hier onderzocht door effecten te schatten in een natuurlijk experiment elders, en die te generaliseren naar Nederland.

Een overgang van een honorering overwegend per abonnement naar een meer gemengde honorering, heeft in oktober 1987 plaatsgevonden in de stad Kopenhagen. De institutionele context van de huisartsenhulp aan het overgrote deel van de Deense bevolking is verregaand gelijk aan de context van de huisartsenhulp aan ziekenfondspatiënten in Nederland. Als we (een steekproef van) de huisartsen in de stad Kopenhagen als experimentele groep beschouwen, vormen de huisartsen in de rest van de provincie Kopenhagen de aangewezen controlegroep: die provincie wordt gevormd door een aaneengesloten stedelijk agglomeraat. Het natuurlijk experiment met deze experimentele en controlegroep vormt de kern van onze studie. Daarnaast worden de stad en de rest van de provincie Kopenhagen voor de verandering vergeleken: daaruit kan blijken in hoeverre de uitkomsten van een natuurlijk experiment gelijk zijn aan die van een cross-sectionele analyse, waartoe men zich in ander onderzoek vaak noodgedwongen beperkt. De invloed van institutionele en verdelingskenmerken van de context van de huisartsenhulp in Kopenhagen en Nederland wordt onderzocht door Kopenhagen voor de verandering met stedelijk Nederland te vergelijken. De gegevens omtrent verrichtingen en verwijzingen door huisartsen in het stedelijk deel van Nederland worden ontleend aan de Nationale Studie van ziekten en verrichtingen in de huisartspraktijk (Bensing, Foets, Van der Velden en Van der Zee, 1990). De uitkomsten worden gebruikt ter kwalificatie van de voorspelde effecten in Nederland. Die effecten worden voorspeld door samenhangen van veranderingen met karakteristieken van huisartsen en verrichtingen in Kopenhagen te schatten, en die toe te passen op Nederlandse beginwaarden, die weer afkomstig zijn uit de Nationale Studie.

2. Het Deense voorbeeld: een gemengd systeem van honorering per verrichting en abonnement

Deense huisartsen hebben kleinere praktijken dan Nederlandse. Ze hebben aanzienlijk langere co-assistentschappen in ziekenhuizen achter de rug. Het aantal ziekenhuisbedden en opnamen in het ziekenhuis per capita is in Denemarken groter. De ziekenhuisplanning, het vestigingsbeleid van huisartsen en de afhandeling van hun declaraties is de verantwoordelijkheid van de provinciale overheden. De abonnementsvergoedingen worden driemaandelijks betaald, en de verrichtingenvergoedingen maandelijks. De verrichtingen die afzonderlijk worden vergoed zijn consulten en zogeheten 'toegevoegde' verrichtingen. De vergoedingen worden er bij vooruitberekening op afgestemd dat in een gemiddelde praktijk de abonnements- en consultvergoedingen een gelijk deel van het inkomen vormen, en de 'toegevoegde' verrichtingenvergoedingen 7,3% van het abonnementsdeel. Tweejaarlijks worden de vergoedingen opnieuw daarop afgestemd in landelijke onderhandelingen tussen overheid en huisartsenvereniging. Provinciale samenwerkingscommissies van overheid en huisartsenvereniging onderzoeken de verrichtingendeclaraties van huisartsen die te ver boven het provinciale gemiddelde liggen.
De consultvergoedingen zijn verschillend voor huisbezoeken, praktijkconsulten, telefonische consulten en herhaalreceptuur. De vergoedingen voor huisbezoeken verschillen

naar af te leggen afstand. Van de 'toegevoegde' verrichtingen worden diagnostische verrichtingen het meest uitgevoerd, gevolgd door verrichtingen van perinatale zorg, curatieve verrichtingen, verrichtingen ter begeleiding van anticonceptie, verrichtingen die overleg behelzen met gemeentelijke diensten voor openbare gezondheidszorg en maatschappelijk werk, en het uitschrijven van certificaten van overlijden en voor opname in een instelling voor geestelijke gezondheidszorg, in die volgorde. Perinatale zorg, overleg met gemeentelijke diensten en certificaten zijn extra reguliere taken van Deense huisartsen, vergeleken met hun Nederlandse collega's. De drie meest-uitgevoerde groepen van verrichtingen (diagnostisch, perinataal, curatief) lijken echter meer technisch georiënteerd te zijn, en de drie minst-uitgevoerde (begeleiding, overleg, certificaat) meer op de persoon van de patiënt gericht. Technisch handelen is dan ook gemakkelijker in afzonderlijk te honoreren deelhandelingen op te splitsen. Het lijkt erop dat vergoedingen per verrichting een medisch-technische gerichtheid in stand houden, die aanvankelijk is aangekweekt in langdurige co-assistentschappen in ziekenhuizen.
Het kan worden aangevochten of dit de kwaliteit van de zorg bevordert, terwijl de beheerskosten van een stelsel als het hier beschrevene aanzienlijk zullen zijn.

3. Abonnements- en verrichtingenhonorering in de Deense huisartspraktijk

Tot oktober 1987 ontvingen de huisartsen in de stad Kopenhagen geen afzonderlijke vergoeding per consult. Van de hiervoor genoemde 'toegevoegde' verrichtingen werden de diagnostische en curatieve verrichtingen (en het overleg met de gemeentelijke dienst) in de stad Kopenhagen niet afzonderlijk vergoed. Het abonnementshonorarium was in de stad Kopenhagen hoger.
In een cross-sectionele vergelijking tussen de stad en de rest van de provincie Kopenhagen voor oktober 1987 (in casu in maart 1987), zou men dan verwachten te vinden dat er in de provincie per ingeschreven patiënt meer consulten plaatsvinden (door meer terugbestellingen, als er evenveel eerste consulten zijn), en meer diagnostische en curatieve verrichtingen worden uitgevoerd. Als dat klopt en inzoverre dat verwijzingen overbodig maakt, zou men verwachten dat patiënten in de provincie minder vaak worden verwezen naar medische specialisten.
Statistisch wordt deze vergelijking bemoeilijkt doordat de data voor stad en provincie Kopenhagen van verschillend aggregatieniveau zijn. We omzeilen deze moeilijkheid door een vuistregel te hanteren. We beschouwen de gemiddelde aantallen pas als verschillend wanneer het ene een veelvoud van het andere is. Volgens die vuistregel geldt dan het volgende.
In de provincie worden per ingeschreven patiënt evenveel consulten gehouden als in de stad. Er worden meer diagnostische en curatieve verrichtingen uitgevoerd. En er vinden evenveel verwijzingen plaats.
Het lijkt er dus op dat de speelruimte van een huisarts om meer consulten te houden, kleiner is dan zijn speelruimte om in deze consulten meer diagnostische en curatieve verrich-

tingen uit te voeren. En het lijkt erop dat die verrichtingen geen verwijzingen overbodig maken.

4. Gevolgen van de invoering van verrichtingenvergoedingen in Kopenhagen

We onderzoeken het effect van een (meer) gemengd honoreringsstelsel niet alleen cross-sectioneel, maar ook in een natuurlijk experiment. We onderzoeken de veranderingen die een half jaar en een jaar na invoering van dat stelsel in de stad Kopenhagen zijn opgetreden ten opzichte van de stand van zaken een half jaar voor invoering. We vergelijken die veranderingen met die welke in dezelfde tijd zijn opgetreden in de rest van de provincie Kopenhagen. Bij het toetsen van het verschil tussen die veranderingen ontmoeten we weer de moeilijkheid van het verschil in aggregatieniveau: hoewel de huisarts de eenheid van analyse zou moeten zijn, beschikken we voor de provincie alleen over data per regionale groep van huisartsen. We lossen dat nu op door de aanname in te voeren dat de variantie tussen huisartsen in de provincie gelijk is aan die in de stad Kopenhagen. We onderzoeken proportionele veranderingen in aantallen per 1000 ingeschreven patiënten per week. Voor de experimentele groep in de stad ontlenen we die aantallen aan een vrijwillige registratie door huisartsen gedurende telkens één week. Voor de controlegroep in de provincie ontlenen we de aantallen aan ambtelijke data over perioden van een maand.
We vinden dat er per ingeschreven patiënt niet meer consulten worden gehouden door de invoering van verrichtingenvergoedingen. We onderzoeken een steekproef van de nieuw gehonoreerde diagnostische en curatieve verrichtingen. Daarover samen genomen, worden er zowel meer diagnostische als meer curatieve verrichtingen uitgevoerd door de verandering van honorering. Medische specialismen samen genomen, vinden er ook minder verwijzingen plaats door de verandering. Onderzoeken we de veranderingen per verrichting en per specialisme, dan ontstaat in termen van significanties een willekeurig beeld.
De toetsing berust op de aanname van gelijke variantie. In de stad betreffen onze data een kortere periode dan in de provincie. In een kortere periode is de aard van de gepresenteerde klachten meer aan het toeval onderhevig. In de mate waarin het al dan niet uitvoeren van een handeling afhangt van de gepresenteerde klacht, zal de variantie in het uitvoeren daarvan dus groter zijn naarmate de data een kortere periode betreffen. In die mate wordt dus de aanname van gelijke variantie geweld aangedaan, en kunnen we in termen van significanties een willekeurig beeld verwachten. En inderdaad zijn de afzonderlijke verrichtingen en de verwijzingen naar afzonderlijke specialismen het meest van de aard van de klacht afhankelijk.
Per verrichting en specialisme keren we dus terug naar onze vuistregel. We beschouwen een toename in de stad van een verrichting als volledig veroorzaakt door de invoering van verrichtingenvergoedingen als hij samengaat met een afname in de provincie. En we beschouwen zo'n toename als merendeels daardoor veroorzaakt als hij ten minste een veelvoud is van een toename in de provincie. Voor de verwijzingen laten we het omge-

keerde gelden.
Dan ontstaat het duidelijkste beeld bij de verrichtingen zoals uitgevoerd een jaar na de verandering. Bij het overgrote deel daarvan (17 van de 19) is sprake van een toename die merendeels of volledig is veroorzaakt door de verandering. Bij de verrichtingen een half jaar na de verandering is het beeld minder eenduidig; bij de verwijzingen is het beeld nog minder eenduidig.
We concluderen dat de invloed van de verandering op het uitvoeren van de nieuw gehonoreerde verrichtingen een half jaar na dato nog niet blijkt, en een jaar na dato wel. Wat betreft de verwijzingen besluiten we met een tautologie: de specialismen waarvoor de verwijzingen het sterkst dalen (zoals gynaecologie), hebben dus blijkbaar de meeste overlap met de huisartsgeneeskunde zoals bedreven in de stad Kopenhagen.

5. Het invoeren van verrichtingenvergoedingen bij professionele onzekerheid

Er zijn grote verschillen in de mate van invloed van de invoering van verrichtingenvergoedingen op de uitvoering van die verrichtingen een jaar na dato. De ene verrichting wordt 17% vaker uitgevoerd, de andere maar liefst 633% vaker dan anderhalf jaar eerder. We veronderstellen (en hebben theoretisch afgeleid) dat dit komt doordat er onder huisartsen voor de ene verrichting meer dan voor de andere, onzekerheid bestaat omtrent de omstandigheden waaronder hij zeker wel, en die waaronder hij zeker niet moet worden uitgevoerd. We verwachten een sterkere toename bij verrichtingen waarvoor deze professionele onzekerheid groter is. Om dat te onderzoeken bepalen we per verrichting de mate van deze onzekerheid.
De omstandigheden in kwestie zijn gegeven met de gepresenteerde klacht (voor diagnostische verrichtingen) of met de vervolgens gestelde diagnose (voor curatieve verrichtingen), als we even een vereenvoudigende tweedeling hanteren. Een grotere zékerheid onder huisartsen blijkt, wanneer het aantal malen dat een huisarts een verrichting uitvoert, meer vastligt met de aantallen malen dat hem bepaalde klachten worden gepresenteerd, dan wel met de aantallen malen dat hij bepaalde diagnosen stelt. De mate waarin dat vastligt, wordt uitgedrukt door de multipele correlatiecoëfficiënt R^2, bij een regressie van het aantal verrichtingen op de aantallen relevante klachten of diagnosen, waarbij huisartsen de eenheden van analyse zijn. De multipele correlatiecoëfficiënt is dus naar zijn vorm een geschikte maat voor de professionele zekerheid onder artsen met betrekking tot een verrichting. De afzonderlijke regressiecoëfficiënten drukken de mate uit waarin de klachten of diagnosen de huisartsen aanleiding geven om de verrichting uit te voeren. De inhoudsvaliditeit van R^2 kan men dus in principe beoordelen door de regressiecoëfficiënten te betrekken op het huisartsgeneeskundig denken: geldt een klacht of diagnose binnen dat denken inderdaad als een goede reden om de verrichting uit te voeren? Echter: hoe globaler de informatie omtrent klachten en diagnosen waarover we beschikken, des te groter is de kans dat die vraag zonder nadere informatie niet kan worden beantwoord.
De huisartsen in de stad Kopenhagen die hun consulten voor en na de verandering

hebben geregistreerd, hebben ook genoteerd welke klacht of diagnose in een consult aan de orde was. Dat is genoteerd (of later gecodeerd) in termen van 45 (groepen van) klachten en diagnosen. Voor ieder van 11 verrichtingen selecteren we op numerieke gronden één tot acht relevante klachten of diagnosen, en bepalen we R^2 als de mate waarin de uitvoering van de verrichting onder Kopenhaagse huisartsen gebonden is aan de klachten of diagnosen in kwestie. Zoals het aantal van 45 klachten en diagnosen al doet vermoeden, is de informatie waarover we aldus beschikken steeds te globaal om de inhoudsvaliditeit van deze maat te beoordelen.
We onderzoeken vervolgens de constructvaliditeit van deze maat voor professionele zekerheid omtrent, of diagnosegebondenheid van verrichtingen. We doen dat door te onderzoeken of de minder diagnosegebonden verrichtingen inderdaad een sterkere toename vertonen, bij de invoering van vergoedingen per verrichting. Dat blijkt het geval te zijn: over de elf verrichtingen vinden we een significante rangcorrelatiecoëfficiënt van -0,64.

6. De huisarts onder abonnementshonorering: een vergelijking van huisartsen in Kopenhagen en stedelijk Nederland

Het kan zijn dat de uitvoering van een minder diagnosegebonden verrichting over het geheel genomen sterker toeneemt bij invoering van een vergoeding per verrichting. Maar is onze werkwijze niet te globaal als we alleen daarmee rekening houden? Heeft een huisarts niet bij de ene klacht of diagnose meer speelruimte om een verrichting al dan niet uit te voeren dan bij de andere? En komt de ene klacht niet meer voor in Kopenhagen, en de andere meer in Nederland? Is met andere woorden de mate van de toename niet afhankelijk van de verdeling van gepresenteerde klachten en gestelde diagnosen?
Mogelijk, maar dat brengt ons op een andere, voorafgaande vraag. Is het verschil in voorkomen van verrichtingen tussen Kopenhagen en Nederland onder abonnementshonorering niet ook al het gevolg van verschillen in voorkomen van aandoeningen?
Dan zouden verschillen weg moeten vallen als we het voorkomen van een verrichting bij dezelfde aandoening vergelijken. We onderzoeken dat alleen voor Kopenhagen en stedelijk Nederland. We onderzoeken het ook alleen in die zin dat we nagaan of significante verschillen insignificant worden. En we onderzoeken het alleen door het globale verschil in voorkomen van een verrichting te vergelijken met het verschil in voorkomen bij één groep van onderling verwante aandoeningen. We vinden dan dat significante globale verschillen altijd significant blijven, en insignificante altijd insignificant.
Verder vinden we dat als een verrichting in stedelijk Nederland en Kopenhagen in verschillende mate wordt uitgevoerd, hij meestal het meest wordt uitgevoerd in stedelijk Nederland. Ook vinden we dat de verrichtingen bij elkaar genomen in stedelijk Nederland vaker worden uitgevoerd. We stellen de vraag hoe dat komt.
We inventariseren relevante verschillen tussen Denemarken en Nederland. De geringere verwijsmogelijkheden in Nederland doen meer verrichtingen in Nederland verwachten; een aantal andere verschillen doet juist minder verrichtingen in Nederland verwachten. We

concluderen dat de geringere verwijsmogelijkheden dan toch de doorslag geven. Dat is des te meer reden om rekening te houden met dit verdelingskenmerk van de situatie van huisartsen.
En zo zijn we dan van een antwoord dat geen uitsluitsel geeft omtrent een gestelde vraag, gekomen op een antwoord dat wel uitsluitsel geeft omtrent een niet gestelde, maar wel ter zake doende vraag.

7. Van Denemarken naar Nederland: voorspelling van de gevolgen van de invoering van verrichtingenvergoedingen

In Kopenhagen steeg het aantal uitgevoerde diagnostische verrichtingen met twee derde en het aantal curatieve verrichtingen met vier vijfde, terwijl het aantal verwijzingen met een kwart daalde. In die mate bleef dus de afname van verwijzingen achter bij de toename van verrichtingen. We vragen ons af of dat in Nederland in dezelfde mate het geval zou zijn. Als er nu in Nederland al minder verwezen wordt eenvoudigweg doordat daartoe minder mogelijkheden aanwezig zijn, zoals we vonden, dan blijft er des te minder ruimte voor afname van verwijzingen over. Maar die afname zal weer minder achterblijven bij de toename van verrichtingen als die verrichtingen onder Nederlandse huisartsen meer diagnosegebonden zijn.
Anderzijds heeft ook de toename van verrichtingen zijn grenzen: als er nu in Nederland al meer verrichtingen worden uitgevoerd, zoals we ook vonden, dan blijft er wellicht ook des te minder ruimte voor toename van de verrichtingen over. Vergelijken we Kopenhaagse huisartsen onderling, dan treedt dit verschijnsel inderdaad op: hoe meer consulten per patiënt een arts om te beginnen al hield, des te minder stijgt dat aantal als een vergoeding per consult wordt ingevoerd; voor diagnostische en curatieve verrichtingen geldt hetzelfde. Bij een vergelijking van huisartsen vinden we níet: hoe groter de toename van diagnostische of curatieve verrichtingen, des te groter de afname van verwijzingen -- ook niet als we die toename per verrichting wegen met zijn diagnosegebondenheid, of als we ons per verrichting beperken tot de relevante specialismen. We wijten dit eraan dat ons onderzoek zich beperkt tot registraties gedurende één week, van een steekproef van 21 van in totaal 70 nieuw gehonoreerde verrichtingen. Bij een vergelijking van Kopenhaagse huisartsen onderling vinden we alleen dat de verandering van het aantal verwijzingen varieert met de verandering van het aantal consulten. We concluderen dat die samenhang het netto resultaat is van toegenomen aantallen (ook niet geregistreerde) verrichtingen en toevallig in de weken in kwestie gepresenteerde klachten.
We kwantificeren de gevonden samenhangen en passen ze toe op de Nederlandse beginwaarden. De verdelingen van de aantallen per 1000 ingeschreven patiënten zijn schever in de meer heterogene steekproef van Nederlandse huisartsen. We schatten het minimale effect daarvan uit de voorspelde verandering van het aantal consulten. We concluderen dat de verwijzingen in Nederland met ten hoogste een kwart zullen afnemen.
Wat betreft de verrichtingen kwantificeren we ook de in hoofdstuk 5 gevonden samenhang

van hun toename met hun diagnosegebondenheid. We laten de vraag naar de inhoudsvaliditeit van onze maat voor diagnosegebondenheid dus even rusten. We passen de samenhang in kwestie toe op de diagnosegebondenheid van de verrichtingen in kwestie onder Nederlandse huisartsen, die we bepalen in hoofdstuk 8. We combineren de resulterende voorspellingen, en concluderen dat de diagnostische verrichtingen in kwestie in Nederland anderhalf à twee maal zo vaak zouden worden uitgevoerd na invoering van verrichtingenvergoedingen, terwijl de curatieve verrichtingen met een derde in aantal zouden toenemen.
Tenslotte combineren we beide voorspellingen voor de verrichtingen waarvoor in Nederland in principe een verrichtingenvergoeding is overeengekomen. Voor de diagnostische verrichtingen onder deze volgt dan weer dat ze anderhalf à twee maal zo vaak zouden worden uitgevoerd, terwijl deze curatieve verrichtingen niet met een derde maar met de helft in aantal zouden toenemen.
Zal de grotere toename van deze curatieve verrichtingen ook een grotere afname van verwijzingen met zich meebrengen? Niet als dat aan hun diagnosegebondenheid ligt, betogen we met een verwijzing naar hoofdstuk 8: onder Nederlandse huisartsen is de diagnosegebondenheid van de curatieve verrichtingen waarvoor een vergoeding is overeengekomen, in het algemeen geringer dan die van de Deense steekproef van curatieve verrichtingen.

8. Bepaling van de diagnosegebondenheid van verrichtingen onder Nederlandse huisartsen

De inhoudsvaliditeit van onze bepaling van de diagnosegebondenheid van een verrichting met de multipele correlatiecoëfficiënt R^2, kan men in principe beoordelen door de regressiecoëfficiënten te betrekken op het huisartsgeneeskundig denken. De regressiecoëfficiënten drukken de mate uit waarin de klachten of diagnosen de huisartsen aanleiding geven om de verrichting uit te voeren. De vraag naar inhoudsvaliditeit luidt dan: geldt een klacht of diagnose binnen het huisartsgeneeskundig denken inderdaad als een goede reden om de verrichting uit te voeren?
Voor de Nederlandse huisartsen beschikken we over voldoende nauwkeurige diagnostische informatie om die vraag te beantwoorden, en dat antwoord luidt op luttele uitzonderingen na bevestigend. We beperken ons hier tot één voorbeeld.
Drie aandoeningen zijn voor de onderzochte huisartsen aanleiding tot blaascatheterisatie: urineretentie, urine-incontinentie en verlamming. Maar voor urineretentie geldt dat in een duidelijk sterkere mate dan voor de andere twee aandoeningen: voor urineretentie heeft de regressiecoëfficiënt de waarde +0,54, en voor ieder van beide andere aandoeningen heeft hij de waarde +0,10. En in het huisartsgeneeskundig denken is blaascatheterisatie bij urineretentie zonder meer geïndiceerd, terwijl het bij de andere aandoeningen weliswaar kan dienen om 'lekkage' en doorliggen te voorkomen, maar dan alleen in die gevallen waarin dat belang zwaarder weegt dan het risico dat men een infectie veroorzaakt.

9. Conclusies

De gemeten diagnosegebondenheid van verrichtingen is onder Nederlandse huisartsen in het algemeen groter dan onder Kopenhaagse. Voor een deel zal dat komen doordat we voor Nederlandse huisartsen over nauwkeuriger diagnostische informatie beschikken. Maar dat lijkt niet alle gevonden verschillen te kunnen verklaren. Inzoverre verrichtingen onder Nederlandse huisartsen werkelijk meer diagnosegebonden zijn, is te verwachten dat de afname van verwijzingen in Nederland minder zal achterblijven bij de toename van verrichtingen dan we hebben voorspeld. We voorspelden de verandering in verwijzingen immers alleen uit hun samenhang met consulten.
Maar op de lange termijn kan die diagnosegebondenheid weer worden beïnvloed door het honoreringsstelsel. We stellen twee hypothesen op. Hoe meer tijd verstrijkt na de invoering van vergoedingen per verrichting, des te meer gaan huisartsen de vergoede verrichtingen zien als middel tot inkomensverwerving, des te breder wordt daardoor het bereik aan aandoeningen waarbij de verrichting wordt uitgevoerd, en des te minder diagnosegebonden wordt de verrichting daardoor. En hoe minder diagnosegebonden een verrichting onder huisartsen is, des te minder zullen zij geloven dat de verrichting een verwijzing overbodig maakt, en des te minder nemen de verwijzingen dus af door een toename van verrichtingen.
Op den duur kan het dan zijn dat er helemaal geen verwijzingen meer worden vermeden door de toegenomen verrichtingen. Dat zou verklaren dat er volgens onze cross-sectionele analyse in de rest van de provincie Kopenhagen meer verrichtingen worden uitgevoerd, en er toch niet minder wordt verwezen dan in de stad Kopenhagen voor de verandering. Zo kan een cross-sectionele analyse uitgevoerd naast een korte-termijn natuurlijk experiment, indicaties leveren omtrent mogelijke lange-termijn effecten.
Maar in Nederland zijn deze effecten alleen te verwachten als de lange-termijn invloed van het honoreringsstelsel sterker is dan die van het standaardenbeleid: door de invloed van het standaardenbeleid zou de uitvoering van verrichtingen mettertijd juist meer diagnosegebonden moeten worden.

References

Arrow, K.J.; Uncertainty and the welfare economics of medical care. **The American Economic Review** 1963; 53(5): 941-973.

Bensing, J.M., M. Foets, J. van der Velden, J. van der Zee; De Nationale Studie van ziekten en verrichtingen in de huisartspraktijk. Achtergronden en methoden. **Huisarts en Wetenschap** 1991; 34(2): 51-61.

Bentzen, N.; Gezondheidszorg in Denemarken. **Medisch Contact** 1983; 38(52): 1633-1638.

Bunker, J.P.; Surgical manpower. A comparison of operations and surgeons in the United States and in England and Wales. **New England Journal of Medicine** 1972; 282: 135-144.

Classification Committee of WONCA (World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians); **ICHPPC-2-Defined**. International Classification of Health Problems in Primary Care. 3rd edition. Oxford, Oxford University Press, 1983.

Commissie Structuur en Financiering Gezondheidszorg; **Bereidheid tot verandering**. The Hague, Distributiecentrum Overheidspublicaties, 1987.

Culyer, A.J.; Cost containment in Europe. **Health Care Financing Review**, 1989 Annual Supplement: 21-32.

Derksen, W.; Het abonnementshonorarium dient gehandhaafd te blijven. **Huisarts en Wetenschap** 1984; (suppl.8): 65-67.

Evans, R.G.; Supplier-induced demand: some empirical evidence and implications, in: M. Perlman (ed.); **The economics of health and medical care**. Proceedings of a conference held by the International Economic Association at Tokyo. London/Basingstoke, MacMillan, 1974.

Evans, R.G.; **Strained Mercy**. The Economics of Canadian Health Care. Toronto, Butterworth & Co., 1984.

Glaser, W.A.; **Paying the doctor**. Systems of Remuneration and Their Effects. Baltimore/London, Johns Hopkins, 1970.

Groenewegen, P.P.; Verrichtingen in de huisartspraktijk. Een presentatie van gegevens uit de Nationale Studie. **Medisch Contact** 1990; 45: 608-611.

Groenewegen, P.P., R. Willemse; **Naar een sterkere eerste lijn?** Deel 2: Buitenlandse ervaringen. Utrecht, NIVEL, 1987.

Ham, C., R. Robinson, M. Benzeval; **Health Check**. Health care reforms in an international context. London, King's Fund Institute, 1990.

Hechter, M., K.-D. Opp, R. Wippler (eds.); **Social Institutions**. Their Emergence, Maintenance and Effects. New York, Walter de Gruyter, 1990.

Janssen, R.T.J.D.; **Honorering van huisartsen**. Een verkenning van effecten op kosten en kwaliteit van de gezondheidszorg. Den Haag, Instituut voor Onderzoek van Overheidsuitgaven, 1988.

KPL/KKS; **Redegørelse** vedrørende kontakt- og ydelsestatistik for alment praktiserende laeger under Københavnsoverenskomsten for perioden uge 16 i 1983 til uge 15 i 1984. Copenhagen, Københavns Praktiserende Laeger / Københavns Kommunes Sundhedsdirektorat, 1985.

Krasnik, A., L. Launsø, J. Ølgod; **Den primaere sundhedstjeneste i Danmark**. Copenhagen, Sundhedsstyrelsen, 1982.

Krasnik, A., P.P. Groenewegen, P.A. Pedersen, P. van Scholten, G. Mooney, A. Gottschau, H.A. Flierman, M.T. Dramsgaard; Changing remuneration systems: effects on activity in general practice. **British Medical Journal** 1990; 300: 1698-1701.

Krogsgaard, M.R. et al.; Skadestuebelastning og indlaeggelsesmonster under PLO-konflikten i efteraret 1984. **Ugeskrift for Laeger** 1985; 147: 4025-4030.

Lamberts, H., M. Wood (eds.); **ICPC**. International classification of primary care. Oxford, Oxford University Press, 1987.

LHV; **Naar een nieuwe honorering**. Utrecht, Landelijke Huisartsen Vereniging, 1981.

LHV; **Basistakenpakket van de huisarts**. Utrecht, Landelijke Huisartsen Vereniging, 1983.

LHV; **De positie van de huisarts in de toekomst**. Utrecht, Landelijke Huisartsen Vereniging, 1987.

LHV; **Kwaliteits- en deskundigheidsbevordering**. Utrecht, Landelijke Huisartsen Vereniging, 1990.

McACE consortium, edited by W.G.W. Boerma & D. Parkin; **European Ambulatory Care Information Systems: An Exploratory Study of Harmonisation**. Volume 2: Identification of opportunities and problems in current health information systems. McACE consortium (CEC AIM Project A1009), 1990.

McPherson, K.; International differences in medical care practices. **Health Care Financing Review**, 1989 Annual Supplement: 9-20.

McPherson, K., J.E. Wennberg, O.B. Hovind, P. Clifford; Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. **The New England Journal of Medicine** 1982; 307: 1310-1314.

Ministerie van Welzijn, Volksgezondheid en Cultuur; **Werken aan zorgvernieuwing**. Actieprogramma van het beleid voor de zorgsector in de jaren negentig. Bijlage bij de Handelingen van de Tweede Kamer der Staten Generaal 1989-1990, 21 545, nr.1.

Newhouse, J.P., W.G. Manning, C.N. Morris et al.; Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance. **The New England Journal of Medicine** 1981; 305 (25): 1501-1507.

Payer, L.; **Medicine & Culture**. Varieties of treatment in the United States, England, West Germany and France. New York, Henry Holt and Company, 1988.

PLO; **Landsoverenskomsten mellem Sygesikringens Forhandlungsudvalg og Praktiserende Laegers Organisation**. Copenhagen, Praktiserende Laegers Organisation, 1988.

Raffel, M.W. (ed.); **Comparative health systems**. Descriptive analysis of fourteen national health systems. University Park/London, Pennsylvania State University Press, 1984.

Richardson, J.; The inducement hypothesis: that doctors generate demand for their own services, in: J. van der Gaag, M. Perlman (eds.); **Health, Economics and Health Economics**. Proceedings of the World Congress on Health Economics, Leiden, The Netherlands, September 1980. Amsterdam/New York/Oxford, North Holland, 1981.

Sandier, S.; Health services utilization and physician income trends. **Health Care Financing Review**, 1989 Annual Supplement: 33-48.

Scherjon, S.A.; De organisatie van de Deense gezondheidszorg. **Tijdschrift voor Sociale Gezondheidszorg** 1983; 61: 266-270.

Schneider, K.C., J.L. Lichtenstein, J.L. Freeman, R.C. Newbold, R.B. Fetter, L. Gottlieb, P.J. Leaf, C.S. Portlock; Ambulatory visit groups: an outpatient classification system. **Journal of Ambulatory Care Management** 1988; 11(3): 1-12.

SFU; **Sygesikringsstatistik 1984-1985**. Copenhagen, Sygesikringens Forhandlings Udvalg, 1987.

Shapiro, M.F., J.E. Ware, C.D. Sherbourne; Effects of Cost Sharing on Seeking Care for Serious and Minor Symptoms. Results of a Randomized Controlled Trial. **Annuals of Internal Medicine** 1986; 104: 246-251.

Sonsbeek, J.L.A. van; (On)gezonde steden. (Un)healthy cities. **Maandbericht Gezondheidsstatistiek (CBS)** (Monthly Bulletin of Health Statistics, Netherlands Central Bureau of Statistics) 1989; 8(8): 5-12.

Tielens, V.C.L.; Standaarden: het gezicht van de huisarts. **Huisarts en Wetenschap** 1989; 32(1): 3.

Vayda, E.; A comparison of surgical rates in Canada and in England and Wales. **The New England Journal of Medicine** 1973; 289: 1224-1229.

Velden, J. van der; De rol van de huisartspraktijk in Nederland. **Medisch Contact** 1990; 45(19): 605-608.

Velden, J. van der, A. Claessens; **Clustering van diagnosen.** Interne rapportage. Utrecht, NIVEL, 1990.

Velden, J. van der, F. Schellevis, J. van der Steen; **ICPC**. International Classification of Primary Care. Tabulaire lijst. Utrecht, NIVEL, 1989.

Verdenius, W., P.J. Brands, R.H. Oudkerk; Over kwaliteitsbeleid. **De huisarts** 1990; 1(6): 24-26.

Vliet, R.C.J.A. van, W.P.M.M. van de Ven; Consumptieverschillen tussen ziekenfonds- en particulier verzekerden nader onderzocht. Overzicht van empirische bevindingen. **Gezondheid & Samenleving** 1986; 7(2): 81-93.

Wal, G. van der; Voor een differentiëring binnen het abonnementshonorarium. **Huisarts en Wetenschap** 1984; 27(suppl.8): 68-71.

Wennberg, J.E.; The paradox of appropriate care. **Journal of the American Medical Association** 1987; 258: 2568-2569.

Wolff, N.; Professional uncertainty and physician medical decision-making in a multiple treatment framework. **Social Science & Medicine**, 1989; 28: 99-107.

Wolters, O.D., M. Stellfeld; Skadestueanvendelser i en periode med patientbetaling for laegeydelser i almen praksis. **Ugeskrift for Laeger** 1986; 148: 975-978.

Zee, J. van der; Wat verdient de huisarts in het buitenland? IV: Denemarken (buiten Kopenhagen). **Praktijkmanagement** 1985; 1(7): 17-19. V: Denemarken (tweede aflevering): Kopenhagen. **Praktijkmanagement** 1985; 1(8): 14-15.

Zee, J. van der; De Nationale Studie en het overheidsbeleid ten aanzien van de huisarts. **Medisch Contact** 1990; 45(19): 611-614.

Curriculum vitae

Hendrik Antheunis Flierman werd geboren op 24 mei 1948 te Middelburg. Van 1960 tot 1965 volgde hij het onderwijs aan de Rijks Hogere Burgerschool aldaar. In 1965 behaalde hij zijn HBS-B diploma. Hij studeerde in 1965-1966 wiskunde, en van 1966 tot 1975 sociologie aan de Rijksuniversiteit te Utrecht. In 1970 slaagde hij voor zijn kandidaatsexamen, en in 1975 voor zijn doctoraalexamen sociologie met als bijvak theorie en methodologie. Nadien was hij aanvankelijk als onderwijsonderzoeker verbonden aan het Research Instituut voor de Toegepaste Psychologie RITP en zijn rechtsopvolger de Stichting Centrum voor Onderwijsonderzoek SCO, beide te Amsterdam. Sinds 1988 is hij als gezondheidszorgonderzoeker verbonden aan het Nederlands Instituut voor Onderzoek van de Eerstelijnsgezondheidszorg NIVEL te Utrecht.